Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 212 (2006) 247–267
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

On the computation of ensemble averages for spatially non-uniform particle systems

Q. Zhang, K. Ichiki [a], A. Prosperetti [a,b,*]

[a] *Department of Mechanical Engineering, The Johns Hopkins University, 223 Latrobe Hall, Charles St. at 34th St., Baltimore MD 21218, United States*
[b] *Faculty of Applied Science and IMPACT, University of Twente, AE 7500 Enschede, The Netherlands, and Burgerscentrum, The Netherlands*

Received 13 October 2004; received in revised form 11 May 2005; accepted 4 July 2005
Available online 15 August 2005

Abstract

When there is no clear separation between micro- and macro-scales, ergodicity cannot be invoked to transform ensemble into volume averages. In such cases it is necessary to use ensemble averaging directly. A straightforward calculation of such averages converges slowly and therefore requires a large number of realizations of the system. This paper describes a much more efficient method based on the use of a Fourier expansion of the quantity to be averaged. The advantages of the Fourier approach are estimated in general terms and demonstrated explicitly with several examples for the specific problem of equal spheres in a viscous fluid. The analytical estimates suggest that similar results can be expected for other situations as well. It is shown both analytically and numerically that the variance of the Fourier coefficients is in many cases significantly smaller than that of the direct method, which leads to a much faster convergence of the former. The paper also describes a method by which the probability distribution of a uniform ensemble can be biased so as to mimic that of a non-uniform one with prescribed properties.

Keywords: Computational ensemble averaging; Non-uniform systems; Stokes flow; Sedimentation

1. Introduction

In the derivation of macroscopic properties for system consisting of a large number of microscopic constituents, the power of ensemble averaging lies in its applicability irrespective of the presence of separation of scales. In this respect, ensemble averaging is quite distinct from time or space averaging, which can be

* Corresponding author. Tel.: +1 410 516 8534; fax: +1 410 5167254.
E-mail address: prosperetti@jhu.edu (A. Prosperetti).

0021-9991/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2005.07.003

meaningfully applied only when the time or spatial scales of the macroscopic behavior of the system are much larger than those of its microscopic constituents. With this greater flexibility, however, comes a greater difficulty of implementation. For example, in a classic study of the stress system in a suspension of particles, Batchelor [1] remarks:

> Ensemble averages can neither be calculated directly nor observed conveniently, and it is necessary to consider the relation between ensemble averages and ... calculable averages ... To establish this relation we shall take for granted the usual ergodicity property of equality of the ensemble average of some quantity and an integral average of the same quantity over any position co-ordinate with respect to which the quantity is statistically stationary.

Thus, with the assumption of ergodicity, ensemble averages can be replaced by volume averages, the numerical computation of which is straightforward provided an efficient method to calculate the microscopic dynamics is available. This approach has been very fruitful and a large number of instances of its application can be found in the literature.

There are cases, however, in which one would be interested in ensemble averages of spatially non-uniform systems. Examples may be particles in a spatially non-uniform force field, molecular dynamics simulations of strong shocks, particles suspended in fluids, and others. A situation which has recently gained particular prominence is that encountered in multi-scale computing (see e.g. [2–4]), where the matching of the micro-and macro-regions requires the evaluation of averages of the quantities obtained from the micro-simulation.

A standard way to deal with these situations is to assume a state of local homogeneity, which is subsequently treated as slowly varying in space. Evidently this approach is only justified when the macroscopic scale L greatly exceeds the microscopic scales which, in addition to the particle scale a, include the characteristic inter-particle distance d and possibly others. In order to avoid this limitation, it is desirable to develop a computational method for the direct evaluation of ensemble averages for spatially non-uniform systems.

The difficulty in evaluating spatially non-uniform ensemble averages can be illustrated already with reference to the simplest average quantity, the particle number density. Consider a system consisting of N indistinguishable particles labelled by the index $\alpha = 1, 2, \ldots, N$. The mean number density $n(\mathbf{x})$ of particles centered at $\mathbf{x}$ is given by

$$n(\mathbf{x}) = \left\langle \sum_{\alpha=1}^{N} \delta(\mathbf{x} - \mathbf{y}^{\alpha}) \right\rangle, \tag{1}$$

where the angle brackets indicate ensemble averaging over some probability distribution dependent on the positions $\mathbf{y}^{\alpha}$ of the particle centers and on all the other degrees of freedom of the system; here and in the following time is a parameter and will not be indicated for brevity. In principle, the direct calculation of $n(\mathbf{x})$ would require the following steps:

(1) A discretization of the system into volumes, or boxes, with centers at $\mathbf{x}_1, \mathbf{x}_2, \ldots$;
(2) The generation of a large number of macroscopically equivalent realizations of the system;
(3) The calculation of the average number of particles in each box.

While conceptually straightforward, it is evident that the convergence of the procedure would be extremely slow, which requires an ensemble with a very large number of realizations. This number will be the larger, the finer the desired estimate of n. If interest lies in more complicated properties of the system, themselves the result of complex micro-dynamics, the required amount of computation may well become prohibitive.

It is the purpose of this paper to present an alternative approach to the problem which not only considerably increases the rate of convergence, but also permits to balance the computational effort with the

required degree of accuracy. This objective would be next to impossible with the direct Monte Carlo approach just described. To illustrate our method, we apply it to the number density $n(\mathbf{x})$ and to a system of equal hard spheres interacting hydrodynamically in a viscous fluid in the low-Reynolds-number regime. These are just examples: the method itself has a much broader applicability.

2. Description of the method

In many applications of ensemble averages, in order to simulate a system with a large spatial extent, one has recourse to the artifice of using a computational cell with periodic boundary conditions. We limit ourselves to this case, although the approach we describe can be extended to more general situations as will be mentioned later in this section.

We consider a generic quantity q^α pertaining to the αth particle, such as velocity, acceleration, and others. The ensemble average $\bar{q}(\mathbf{x})$ of this quantity for a system of N particles is given by

$$n(\mathbf{x})\bar{q}(\mathbf{x}) = \left\langle \sum_{\alpha=1}^{N} q^\alpha \delta(\mathbf{x} - \mathbf{y}^\alpha) \right\rangle. \tag{2}$$

In order to approximate this quantity by the direct method mentioned in the previous section, we would discretize the domain of interest into small boxes ΔV_j and, in each box, approximate the continuous function $n\bar{q}$ by a constant $n(\mathbf{x}_j)\bar{q}(\mathbf{x}_j)$ given by

$$n(\mathbf{x}_j)\bar{q}(\mathbf{x}_j) = \frac{1}{\Delta V_j} \left\langle \sum_{\mathbf{y}^\alpha \in \Delta V_j} q^\alpha \right\rangle. \tag{3}$$

In other words, for each realization, the sum of the q^α for all the particles in the jth box is evaluated, and then the average over all realizations is taken. This result is at the basis of the direct procedure for the evaluation of ensemble averages mentioned in the previous section. As shown later in this section, this relation gives the best (in the least-squares sense) piecewise-constant approximation of the function $n\bar{q}$.

The alternative approach we advocate here is based on the calculation of the ensemble average by expanding $n\bar{q}$ in a Fourier series:

$$n(\mathbf{x})\bar{q}(\mathbf{x}) = \sum_{\mathbf{k}} (nq)_{\mathbf{k}} \exp(-\mathrm{i}\mathbf{k} \cdot \mathbf{x}), \tag{4}$$

where the summation is extended over all the wave numbers of the reciprocal lattice. For example, if the fundamental periodic cell is a cube of side L aligned with the coordinate axes, the allowed values of $\mathbf{k} = (k_x, k_y, k_z)$ are $2\pi(i,j,k)/L$, where i, j and k are integers, and the summation extends over the entire range $-\infty < i,j,k < \infty$. It is easy to show that, according to the rules for the calculation of the coefficients of a Fourier series, since volume integration and ensemble averaging commute, we have

$$(nq)_{\mathbf{k}} = \frac{1}{V} \left\langle \sum_{\alpha=1}^{N} \exp(\mathrm{i}\mathbf{k} \cdot \mathbf{y}^\alpha) q^\alpha \right\rangle, \tag{5}$$

where V is the volume of the fundamental cell of the periodic system. In particular, for $\mathbf{k} = 0$

$$(nq)_0 = \frac{1}{V} \left\langle \sum_{\alpha=1}^{N} q^\alpha \right\rangle, \tag{6}$$

which is the way in which ensemble averages are calculated for a statistically uniform system by exploiting ergodicity. Eq. (5) extends the procedure to the non-uniform case. Once the Fourier coefficients are known, the spatial dependence of the ensemble average is reconstructed from (4).

One may expect – and the results to be shown later confirm it for most cases – that the second method requires a smaller number of realizations for an acceptably converged average than the first one. In the first place, it is well known that, for a given number of terms and sufficiently smooth functions, the Fourier expansion gives a more faithful representation of the function of interest than a piece-wise constant approximation and is, in this sense, more economical [5,6]. Secondly, at an intuitive level, by the very way they are calculated, the Fourier series coefficients include information over the entire domain. This feature is useful as the spatial distribution of interacting particles includes non-local effects. For example, the presence of a particle at a point makes the presence of a particle at a neighboring point less likely – an effect that would not be as easily captured by the discretized representation (7).

It is evident that the idea of the expansion (4) can be implemented by using different basis functions, which may be more suitable for other situations. The coefficients of any expansion would be given by scalar products, which will always involve volume integrals which commute with averaging. Thus, although the details of the calculation would be different, one would expect to encounter features similar to those found here in the case of Fourier coefficients.

In conclusion, we prove the results (3) and (5). For the former, we start by replacing $n\bar{q}$ by a piecewise constant function

$$n(\mathbf{x})\bar{q}(\mathbf{x}) \simeq \sum_j n(\mathbf{x}_j)\bar{q}(\mathbf{x}_j)\chi_j(\mathbf{x}), \tag{7}$$

where $\chi_j(\mathbf{x})$ denotes the characteristic function of the jth box centered at $\mathbf{x}_j$. A least square minimization of the resulting error shows that the constants $n(\mathbf{x}_j)\bar{q}(\mathbf{x}_j)$ should be chosen as

$$n(\mathbf{x}_j)\bar{q}(\mathbf{x}_j) = \frac{1}{\Delta V_j}\int_V \mathrm{d}^3x n(\mathbf{x})\bar{q}(\mathbf{x})\chi_j(\mathbf{x}) = \frac{1}{\Delta V_j}\int_{\Delta V_j} \mathrm{d}^3x n(\mathbf{x})\bar{q}(\mathbf{x}) \tag{8}$$

from which, upon substitution of the definition (2) of $n\bar{q}$, (3) follows immediately. For (5), we exploit the commutativity of volume integration and ensemble averaging to find

$$(nq)_{\mathbf{k}} = \frac{1}{V}\int_V \mathrm{d}^3x \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} n(\mathbf{x})\bar{q}(\mathbf{x}) = \frac{1}{V}\left\langle \int_V \mathrm{d}^3x \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \sum_{\alpha=1}^{N} q^\alpha \delta(\mathbf{x}-\mathbf{y}^\alpha)\right\rangle \tag{9}$$

from which (5) follows.

3. Variance

For practical computations, the convergence rate of the averages is of the utmost importance. This feature may be assessed by estimating the standard deviation of the quantity being averaged as, from the Chebyshef inequality, the smaller this quantity the smaller the probability with which values far from the mean occur [7]. For this reason, we now present an approximate calculation of the variance of the direct-sum and of the Fourier coefficient methods described in Section 2. This theoretical analysis will be illustrated with numerical results in the sections that follow.

In the application of ensemble averaging to a specific physical process, one would normally deal with realizations which are the result of that process and which, therefore, reflect its statistics. The probability distribution P would then be a byproduct of the numerical simulation of many realizations of the microphysics of the process, each one of which would be started from a microscopically different, but macroscopically equivalent, realization.

For the purposes of illustrating the ideas described before, however, it is unnecessary to go through this step as it is sufficient to use a "model", or "synthetic", ensemble artificially generated according to some

convenient rule. The simplest procedure is to make use of a spatially uniform ensemble, characterized by a statistically uniform probability distribution $P_0(C^N)$, which is rendered non-uniform by assigning to each realization C^N a weight $W(C^N)$ writing

$$P(C^N) = W(C^N)P_0(C^N). \tag{10}$$

It is shown in Appendix how to generate weights $W(C^N)$ capable of producing ensembles with a prescribed non-uniform number density $n(\mathbf{x})$. In order to illustrate our method and conclusions, we focus on the specific simple case in which the prescribed n is given by

$$n(\mathbf{x}) = n_0 + n_s \sin \mathbf{k} \cdot \mathbf{x}. \tag{11}$$

According to (A.1) in Appendix A, the corresponding weights should be chosen as[1]

$$P(C^N) = P_0(C^N)\left[1 + \frac{n_s}{n_0}\Phi(C^N)\right] \quad \text{with} \quad \Phi(C^N) = \frac{1}{S_0(\mathbf{k})}\sum_{\alpha=1}^{N} \sin \mathbf{k} \cdot \mathbf{y}^\alpha \tag{12}$$

in which $S_0(\mathbf{k})$ is the structure factor of the uniform distribution defined in (A.3). Here and in the following, the index 0 denotes quantities pertaining to the uniform probability distribution $P_0(C^N)$. Since S_0 does not vanish, for n_s sufficiently small, the weights defined in (12) are positive and therefore generate a legitimate probability distribution.

The ensemble average denoted by angle brackets in (2) can be written more explicitly as

$$\left\langle \sum_{\alpha=1}^{N} q^\alpha \delta(\mathbf{x} - \mathbf{y}^\alpha) \right\rangle = \frac{1}{N!}\int dC^N P(C^N) \sum_{\alpha=1}^{N} q^\alpha \delta(\mathbf{x} - \mathbf{y}^\alpha), \tag{13}$$

where the $N!$ normalization is due to the identity of the particles and $P(C^N)$ denotes the probability density with which the specific realization C^N of the system occurs in the ensemble. We use the word "realization" in a broad sense including all the independent degrees of freedom of each particle: position, velocity, orientation, angular velocity etc. The integration is over all these degrees of freedom.

Since our interest here is in non-uniform ensembles, we will focus on the difference between the averages for the non-uniform and uniform distributions for the generic quantity q, scaled by n_0/n_s for convenience:

$$\Delta^q(\mathbf{x}) = \frac{n_0}{n_s}\left(\left\langle \sum_{\alpha=1}^{N} q^\alpha \delta(\mathbf{x} - \mathbf{y}^\alpha) \right\rangle - \left\langle \sum_{\alpha=1}^{N} q^\alpha \delta(\mathbf{x} - \mathbf{y}^\alpha) \right\rangle_0\right). \tag{14}$$

Here $\langle \cdots \rangle$ is the average (13) taken with the non-uniform probability P and $\langle \ldots \rangle_0$ the corresponding average calculated from (13) with P replaced by the uniform probability P_0. In view of definition (13), we have

$$\Delta^q(\mathbf{x}) = \left\langle \Phi(C^N) \sum_{\alpha=1}^{N} q^\alpha \delta(\mathbf{x} - \mathbf{y}^\alpha) \right\rangle_0, \tag{15}$$

in which Φ, defined in (12), is a measure of the non-uniformity of the probability distribution. For the Fourier method, by (4) and (5), $\Delta^q(\mathbf{x})$ equals

$$\Delta^q(\mathbf{x}) = \Delta^q_s(\mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{x}) + \Delta^q_c(\mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{x}) \tag{16}$$

with

[1] One may view this probability distribution as the result of displacing each particle in each realization of the uniform ensemble by an amount proportional to $\sin \mathbf{k} \cdot \mathbf{y}^\alpha$ from its initial position $\mathbf{y}^\alpha$.

$$\Delta_s^q(\mathbf{k}) \equiv \frac{2}{V}\left\langle \Phi(C^N)\sum_{\alpha=1}^{N} q^\alpha \sin \mathbf{k}\cdot\mathbf{y}^\alpha \right\rangle_0, \tag{17}$$

$$\Delta_c^q(\mathbf{k}) \equiv \frac{2}{V}\left\langle \Phi(C^N)\sum_{\alpha=1}^{N} q^\alpha \cos \mathbf{k}\cdot\mathbf{y}^\alpha \right\rangle_0. \tag{18}$$

For the direct-sum method, according to (3), we wish to estimate the variance of

$$\Delta_j^q \equiv \frac{1}{\Delta V_j}\left\langle \sum_{\mathbf{y}^\alpha\in\Delta V_j} q^\alpha \right\rangle - \frac{1}{\Delta V_j}\left\langle \sum_{\mathbf{y}^\alpha\in\Delta V_j} q^\alpha \right\rangle_0 = \left\langle \Phi\left(\frac{1}{\Delta V_j}\sum_{\mathbf{y}^\alpha\in\Delta V_j} q^\alpha\right)\right\rangle_0. \tag{19}$$

It is shown in Appendix B that the result is, approximately

$$\mathrm{var}[\Delta_j^q] \simeq \frac{n_0^2\langle q^2\rangle_1}{2S_0(\mathbf{k})}[N + N_\mathrm{b} + N\Gamma_j^q] + (\Delta_j^q)^2. \tag{20}$$

Here N is the number of particles, N_b the number of boxes in which the domain of interest has been discretized for the calculation of the ensemble average according to the direct method, and $\langle q^2\rangle_1$ is the value of q^2 for a particle averaged over the position of all the other particles; a precise definition is given in (B.13). Furthermore, Γ_j^q is given by

$$\Gamma_j^q = \frac{1}{\Delta V_j^2}\int_{\Delta V_j} \mathrm{d}^3y \int_{\Delta V_j} \mathrm{d}^3z g_0(\mathbf{z}-\mathbf{y})\frac{\langle q^y q^z\rangle_2(\mathbf{z}-\mathbf{y}) - \langle q^2\rangle_1}{\langle q^2\rangle_1}, \tag{21}$$

in which $\langle q^y q^z\rangle_2(\mathbf{y},\mathbf{z})$ is the average over the other particles of the product of the qs for two particles, one centered at $\mathbf{y}$ and the other at $\mathbf{z}$. In (21) $g_0(\mathbf{r})$ is the pair distribution function of the uniform ensemble, proportional to the probability of having two particles separated by $\mathbf{r}$. If the values of q for two particles are completely uncorrelated, $\langle q^y q^z\rangle_2 = \langle q^2\rangle_1$ and $\Gamma_j^q = 0$. This quantity therefore is sensitive to particle correlations.

The corresponding results for the variances of $\Delta_s^q(\mathbf{k})$ and $\Delta_c^q(\mathbf{k})$ according to the Fourier method are shown in Appendix B to be given by

$$\mathrm{var}[\Delta_s^q] = \frac{n_0^2\langle q^2\rangle_1}{S_0(\mathbf{k})}[S_0(\mathbf{k}) + N\Gamma_\mathbf{k}^q] + [\Delta_s^q(\mathbf{k})]^2, \tag{22}$$

$$\mathrm{var}[\Delta_c^q] = \frac{n_0^2\langle q^2\rangle_1}{S_0(\mathbf{k})}[S_0(\mathbf{k}) + N\Gamma_\mathbf{k}^q] + [\Delta_c^q(\mathbf{k})]^2 \tag{23}$$

in which

$$\Gamma_\mathbf{k}^q = \frac{1}{V}\int \mathrm{d}^3r \cos\mathbf{k}\cdot\mathbf{r} g_0(\mathbf{r})\frac{\langle q^y q^z\rangle_2(\mathbf{r}) - \langle q^2\rangle_1}{\langle q^2\rangle_1}. \tag{24}$$

Again, this quantity is sensitive to particle correlations and will become smaller with increasing $|\mathbf{k}|$, at least from some point on.

It is evident from (8) that Δ_j^q is the average over the jth box of the quantity $\Delta^q(\mathbf{x})$ defined in (16) and, therefore, Δ_j^q is at most of the order of Δ_s^q or Δ_c^q. The real difference between the two estimates (20) of the variance of the direct method and (22), (23) of the Fourier variance lies therefore in the first terms in the right-hand sides. If the distance over which the property q of two particles is strongly correlated is small, $\Gamma_j^q \simeq 0$, $\Gamma_\mathbf{k}^q \simeq 0$ and the quantities of interest are then

$$\frac{N_\mathrm{b}+N}{2S_0(\mathbf{k})}n_0^2\langle q^2\rangle_1 \quad \text{and} \quad n_0^2\langle q^2\rangle_1. \tag{25}$$

In typical applications, $N + N_b$ is a large number. Furthermore, $S_0(\mathbf{k})$ is less than 1 when the modulus of $|\mathbf{k}|$ is less than about $2\pi/d$, where d is a measure of the mean interparticle distance. It may therefore be concluded that, in this case, the Fourier-method variance is much less than that of the direct method. When the Γ terms dominate, the comparison is between

$$n_0^2\langle q^2\rangle_1 \frac{N\Gamma_j^q}{2S_0(\mathbf{k})} \quad \text{and} \quad n_0^2\langle q^2\rangle_1 \frac{N\Gamma_{\mathbf{k}}^q}{S_0(\mathbf{k})}. \tag{26}$$

For small $|\mathbf{k}|$, Γ_j^q and $\Gamma_{\mathbf{k}}^q$ may be expected to be of the same order of magnitude, from which we conclude that the two variances are comparable in this case. However, $\Gamma_{\mathbf{k}}^q$ depends on $\mathbf{k}$, while Γ_j^q does not, and [8,9]

$$\frac{1}{V}\int \mathrm{d}^3 r \cos\mathbf{k}\cdot\mathbf{r} g_0(\mathbf{r}) = \frac{1}{N}[S_0(\mathbf{k}) - 1]. \tag{27}$$

As $|\mathbf{k}|$ increases, $S_0 \to 1$, which suggests that, once again, the direct-method variance would be larger than the Fourier-method one.

4. Construction of the ensembles and calculation of the averages

We illustrate the results of the previous analysis by considering equal spherical rigid particles with radius a immersed in a viscous fluid. We construct uniform ensembles of N particles in a fundamental cubic cell of side L by the following procedure. We start by randomly arranging the particles in the cell making sure that no overlap occurs, and subject them to a random walk taking care to avoid overlaps at each step. After $100N$ steps per particle, we store the resultant realization as a member of the ensemble. A different initial random arrangement is generated for every realization. To avoid biases, particles are displaced in random order rather than in a fixed sequence. By repeating this procedure, we construct ensembles of N_r realizations.

In order to test the randomness of these ensembles, we calculate the structure factor $S_0(\mathbf{k})$ defined in (A.3) considering wave numbers $\mathbf{k} \equiv (k_x, k_y, k_z)$ of the form $2\pi(i, j, k)/L$, where i, j, k are integers and the coordinate directions are taken parallel to the sides of the fundamental cell, each one of which has length L. As an example of one of these calculations, in Fig. 1 we compare our results (circles) with the Percus–Yevick approximation to the hard-sphere structure factor in infinite space [10–12] for a sphere volume fraction of 15% (line). These results were obtained with a variable number realizations ranging from 256 for 160 particles to 1.048 for 16

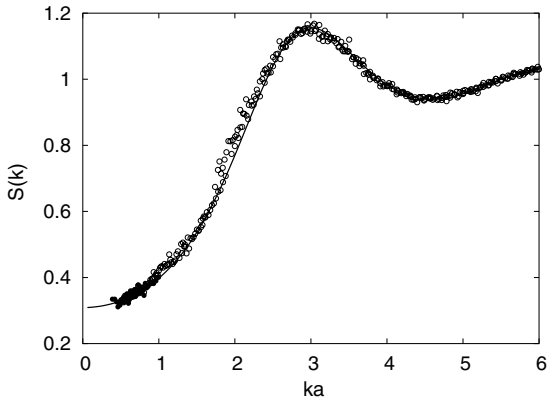

Fig. 1. Comparison between the structure factor given by the Percus–Yevick solution $S_{PY}(k)$ (solid line) and $S(k)$ numerically calculated from (A.3) from the realizations used in the present work for a volume fraction $\beta = 0.15$. The solid circles are calculated with $k = 2\pi/L$ and the open circles are with higher spatial modes.

particles. [13] Since the structure factor in this case is isotropic, $S_0(k)$ is plotted as a function of $k = |\mathbf{k}|$ and the point plotted is the average over the three different directions of the wave vector. This result shows that, although our ensembles are not strictly isotropic, they give rise to a structure factor essentially indistinguishable from the Percus–Yevick distribution in infinite space in the wave vectors range greater than $2\pi/L$. This result is reproduced from [13], where it is also shown that the relative difference between our structure factor and the Percus–Yevick solution is independent of the volume fraction and around 6–8%.

Due to the functional form of the weights (A.1), we now consider in detail the special case

$$n(\mathbf{x}) = n_0 + n_s \sin \mathbf{k}_\mathrm{m} \cdot \mathbf{x}, \tag{28}$$

where $\mathbf{k}_\mathrm{m}$ is a vector parallel to one of the sides of the fundamental cell with magnitude

$$k_\mathrm{m} = |\mathbf{k}_\mathrm{m}| = \frac{2\pi}{L} m, \tag{29}$$

where m is a positive integer. The corresponding probability distribution is

$$P(C^N) = P_0(C^N)\left(1 + \frac{n_s}{n_0}\Phi_\mathrm{m}\right), \tag{30}$$

with

$$\Phi_\mathrm{m}(C^N) = \frac{1}{S_0(\mathbf{k}_\mathrm{m})} \sum_{\alpha=1}^{N} \sin(\mathbf{k}_\mathrm{m} \cdot \mathbf{y}^\alpha). \tag{31}$$

Then, with (13), (5) becomes

$$\begin{aligned}(nq)_\mathbf{k} &= \frac{1}{N!}\int \mathrm{d}C^N P_0\left(1 + \frac{n_s}{n_0}\Phi_\mathrm{m}\right)\frac{1}{V}\sum_{\alpha=1}^{N}\exp(\mathrm{i}\mathbf{k}_\mathrm{m}\cdot\mathbf{y}^\alpha)q^\alpha \\ &= \frac{n_s}{n_0}\frac{1}{N!}\int \mathrm{d}C^N P_0 \Phi_\mathrm{m}\frac{1}{V}\sum_{\alpha=1}^{N}\exp(\mathrm{i}\mathbf{k}_\mathrm{m}\cdot\mathbf{y}^\alpha)q^\alpha,\end{aligned} \tag{32}$$

where the omitted term is readily seen to vanish in the limit of an infinite ensemble. In the numerical calculation, the factor $N!$ can be ignored, as it merely amounts to a renumbering of the particles, and the integral over the degrees of freedom of the system is approximated by the average over the N_r realizations of the ensemble. In this way we have

$$\frac{n_0}{n_s}(nq)_{\mathbf{k}_\mathrm{m}} \simeq \frac{1}{N_r}\sum_{\ell=1}^{N_r}\frac{1}{V}[\Phi_\mathrm{m}]_\ell\left[\sum_{\alpha=1}^{N}\exp(\mathrm{i}\mathbf{k}_\mathrm{m}\cdot\mathbf{y}^\alpha)q^\alpha\right]_\ell, \tag{33}$$

where $[\ldots]_\ell$ denotes the value of the bracketed quantity calculated for the ℓth realization.

In a similar way we write (3) as

$$n(\mathbf{x}_j)\bar{q}(\mathbf{x}_j) = (n\bar{q})_j^0 + (n\bar{q})_j^m, \tag{34}$$

where

$$(n\bar{q})_j^0 = \frac{1}{\Delta V_j}\frac{1}{N_r}\sum_{\ell=1}^{N_r}\left[\sum_{\mathbf{y}^\alpha\in\Delta V_j} q^\alpha\right]_\ell, \tag{35}$$

$$\frac{n_0}{n_s}(n\bar{q})_j^m = \frac{1}{\Delta V_j}\frac{1}{N_r}\sum_{\ell=1}^{N_r}\left[\Phi_\mathrm{m}\sum_{\mathbf{y}^\alpha\in\Delta V_j} q^\alpha\right]_\ell. \tag{36}$$

Since here we are interested in the effects of non-uniformity, we will focus on the quantity defined by (36). The uniform part $(n\bar{q})_j^0$ is obviously independent of j and equal to the uniform part of the Fourier method, $(nq)_0$ defined in (6).

5. Application to the number density

To illustrate the analysis of Section 3, we start by considering the simple case of the number density for which $q^\alpha = 1$ and $\Gamma_j^q = 0$. In this case the result (20) for the direct-method variance simply becomes

$$\mathrm{var}[\Delta_j^n] \simeq \frac{n_0^2\langle q^2\rangle_1}{2S_0(k_\mathrm{m})}(N_\mathrm{b}+N) + (\Delta_j^n)^2. \tag{37}$$

But, from (14) and (3)

$$\Delta_j^n = \frac{n_0}{\Delta V_j}\int_{\Delta V_j} \mathrm{d}^3x \sin\mathbf{k}_\mathrm{m}\cdot\mathbf{x} = n_0 \sin\mathbf{k}_\mathrm{m}\cdot\boldsymbol{\xi}_j, \tag{38}$$

where use has been made of the mean value theorem, so that

$$\mathrm{var}[\Delta_j^n] \simeq \left[\frac{N+N_\mathrm{b}}{2S_0(k_\mathrm{m})} + \sin^2\mathbf{k}_\mathrm{m}\cdot\boldsymbol{\xi}_j\right] n_0^2 \simeq \frac{N+N_\mathrm{b}}{2S_0(k_\mathrm{m})} n_0^2. \tag{39}$$

For the Fourier method on the other hand, from (B.8), $\Delta_s^n = n_0$ while $\Delta_c^n = 0$, so that

$$\mathrm{var}[\Delta_s^n] \simeq 2n_0^2, \quad \mathrm{var}[\Delta_c^n] \simeq n_0^2. \tag{40}$$

The structure factor is essentially of order 1 or smaller while $N + N_\mathrm{b}$ is typically large. It is therefore evident that the right-hand side of (40) is smaller than that of (39).

These expectations are supported by the numerical evidence. Fig. 2 show the variance (normalized by multiplication by a^6), calculated according to the direct method for volume fractions of 15% and 45% with $N = 150$ particles. Here the fundamental cell has been divided into $N_\mathrm{b} = 16$ boxes by planes perpendicular to the direction of $\mathbf{k}_\mathrm{m}$. The points, showing the numerically calculated variances for $k_\mathrm{m} = 2m\pi/L$ with $m = 1, 2, 4, 6, 8$ are placed at the center of the corresponding box. A total of 1536 realizations were used to generate these results. In agreement with (39), the variance is nearly constant, only exhibiting a slight sinusoidal variation corresponding to the term $2S_0(k_\mathrm{m})\sin^2\mathbf{k}_\mathrm{m}\cdot\boldsymbol{\xi}_j$ in the equation. The Fourier-method variances are essentially independent of the wave number (or, equivalently, of m) and to be very closely predicted by (40) which gives values of $2.565\times10^{-3}a^6$ and $2.850\times10^{-2}a^6$ for volume fractions of 15% and 45%, respectively; these values are too small to be shown on the scale of these figures.

The direct-method variance averaged over all the boxes and the Fourier variance are shown as functions of k on a logarithmic scale in Fig. 3. Here the lines are the theoretical predictions (39) and (40) and the symbols the computed numerical results. These results fully support the previous analysis.

To illustrate the effect of the larger variance on the convergence rate of the averaging process, we show in Fig. 4 graphs of

$$a^3|[\Delta_s^n(k_1)]_{N_r} - [(\Delta_s^n(k_1)]_\mathrm{exact}|, \quad \text{and} \quad a^3|[\Delta_j^n]_{N_r} - [\Delta_j^n]_\mathrm{exact}|, \tag{41}$$

the latter for $j = 4$ and 8. Here $\Delta_s^q(k_1)$ is the scaled Fourier amplitude of the particle number density fluctuation Δ_s^n defined in (17) evaluated for $k_1 = 2\pi/L$ and Δ_j^n is the corresponding quantity for the direct method defined in (19). The subscript *exact* denotes the exact value and the subscript N_r denotes the running average over N_r realizations of the ensemble, which is the variable plotted on the horizontal axis. Here the fundamental cell has been divided into 16 boxes in the same way as previously described for Fig. 2. Box 4 contains a maximum of $n(\mathbf{x})$, while box 8 is close to the point where $n(\mathbf{x}) = n_0$. In this case the

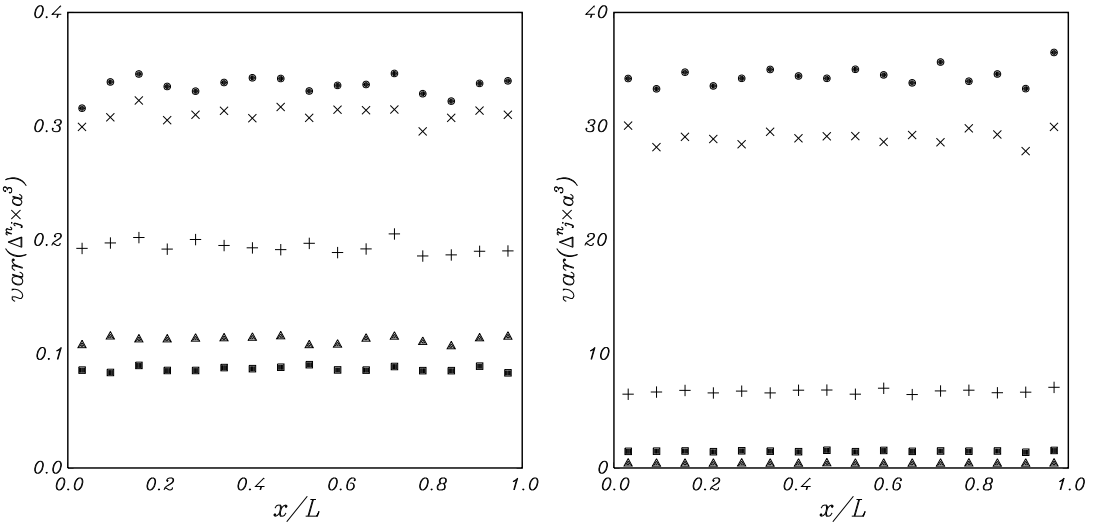

Fig. 2. Direct-method variance (normalized by multiplication by a^6, with a the particle radius) for the number density. The volume fractions are 15% (left) and 45%; the volume is divided into 16 boxes with $N = 150$ particles and 1536 realizations. The points, showing the numerically calculated variances for $k = 2m\pi/L$ with $m = 1, 2, 4, 6, 8$, are placed at the center of the corresponding box: $m = 1$ (●), $m = 2$ (×), $m = 4$ (+), $m = 6$ (▲), $m = 8$ (■). The normalized Fourier variances calculated from (40) are 2.565×10^{-3} and 2.850×10^{-2} for 15% and 45%, respectively. They are too small to show clearly in these figures.

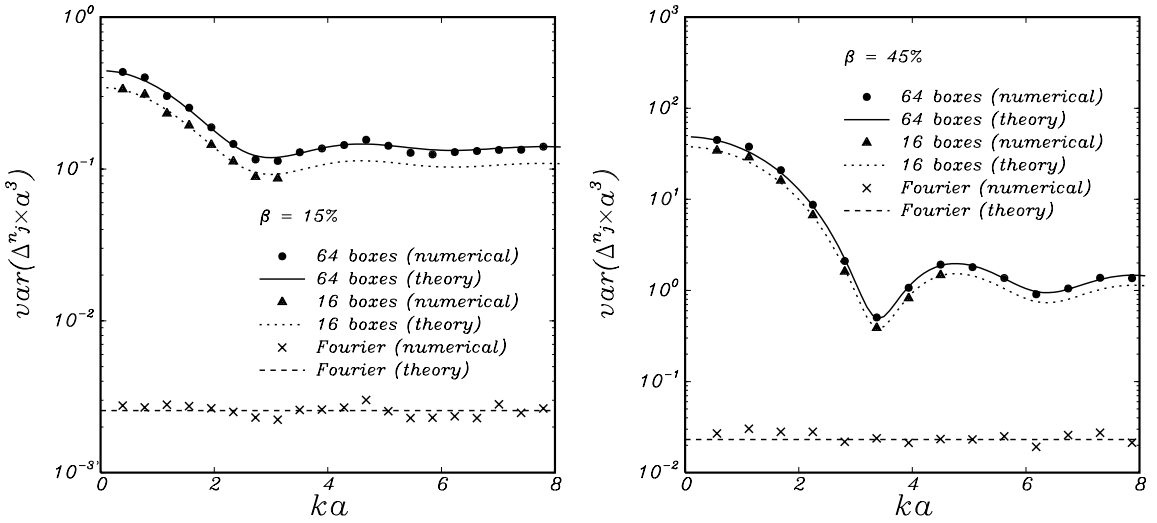

Fig. 3. Variances for the Fourier and the direct method normalized by multiplication by a^6 as functions of ka with the parameters of the previous figure for volume fractions of 15% (left) and 45%; a is the particle radius. The direct-method variance is averaged over all the boxes. The lines are the theoretical predictions (39) and (40) and the symbols the computed results.

fundamental cell contains 150 particles, the volume fraction is 15% and 1536 realizations were used. The straight line with long dashes has a slope of $-1/2$. The figure shows that the error decreases according to $N_r^{-1/2}$ for both methods but, for the same number of realizations, the error affecting the direct method is consistently at least one order of magnitude greater than that affecting the Fourier method.

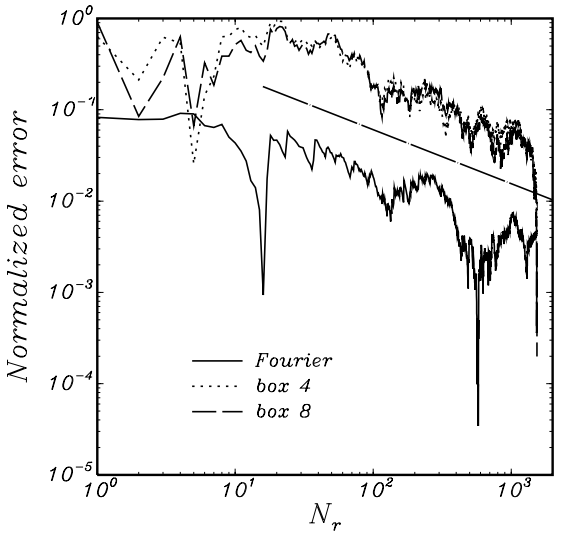

Fig. 4. Running average of the normalized error (41) affecting the density fluctuation Δ_s^n defined in (17) and of the similar quantity for the direct method, Δ_j^n defined in (19). The broken lines are for $j = 4$ and 8 for a subdivision of the interval into 16 boxes with $k_1 = 2\pi/L$. The long-dashed line has a slope of $-1/2$.

As a final example we consider a situation in which the number density has a polynomial non-uniformity in the x-direction

$$\frac{n(x) - n_0}{An_0} = \left(\frac{x}{L}\right)^2 \left(1 - \frac{x}{L}\right)^2 \left(\frac{x}{L} - \frac{1}{4}\right) \left(\frac{x}{L} - \frac{9}{14}\right). \tag{42}$$

where A is an arbitrary amplitude. The x-dependent part has mean zero, so that the volume average of n is n_0 as in the previous cases. A Fourier approximation to $n(x)$ as defined by this function including the first four sines and five cosines has a maximum error of 2.5% relative to the maximum. The solid line in Fig. 5 shows $[n(x) - n_0]/(An_0)$ as given by (42), the dotted line is the value reconstructed from the ensemble average of the first four cosine and five sine coefficients, while the broken line is the result of direct ensemble averaging (as in (3) with $q^\alpha = 1$), all calculated with 1536 realizations.

6. Application to Stokes flow

The number density considered in the previous section only embodies the simple inter-particle interaction due to impenetrability. It is interesting to consider also a more complex interaction dependent on an actual physical process. For this purpose, we turn to the case of equal spherical particles suspended in a slow fluid flow for which inertia effects are unimportant, the so-called Stokes regime. It may be recalled that, in Stokes flow, for a given driving agent, velocities and angular velocities are uniquely determined by the particle position. Thus, in order to examine the velocity statistics, it is sufficient to focus on an ensemble of particle configurations at a given instant; the subsequent time evolution of the system is immaterial. For the numerical simulations we used the multipole method of [14]; details can be found in [13].

We consider first the process of sedimentation, in which an equal force, parallel to the sides of the fundamental cell, is applied to each particle. For translating particles, the decay rate of the fluid velocity disturbance is very slow (for a single particle, inversely proportional to the distance), so that even fairly distant

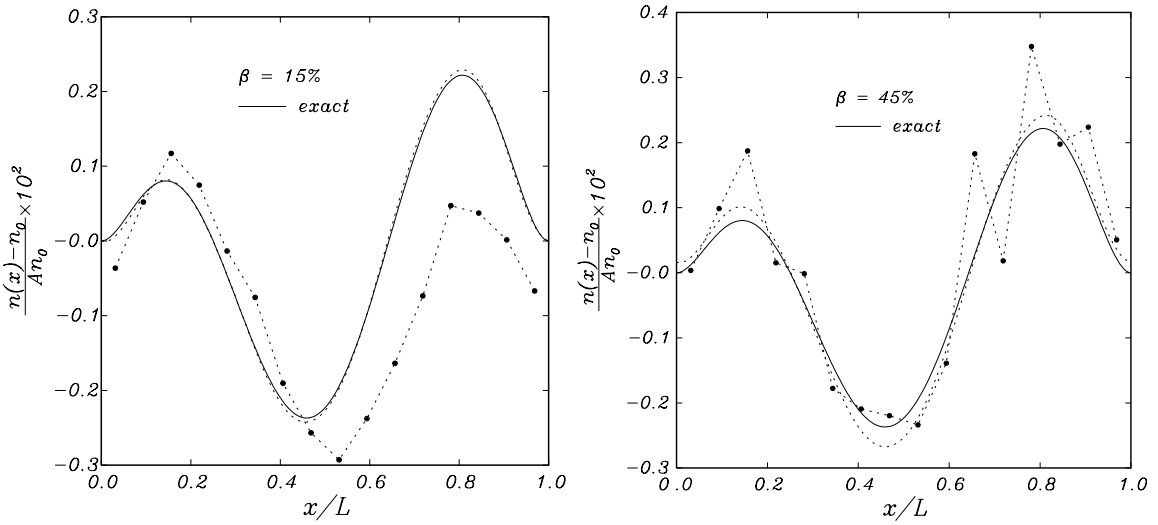

Fig. 5. Comparison of the exact number density non-uniformity (42) (solid line), the Fourier ensemble average (dotted line) and the direct ensemble average (broken line with dots) for volume fractions of 15% (left) and 45% calculated with 1536 realizations of 150 particles.

particles strongly influence each other (see e.g. [15–17]). Consideration of this situation represents therefore a case of rather extreme long-range particle–particle interaction, and lends itself well to an illustration of the effect of the terms Γ in (20), (22) and (23).

We calculate the velocity w in the direction of the applied force in a reference frame in which the center of mass of the particle-fluid mixture is at rest as would happen, for example, in a stationary container. For the direct method the boxes are generated by planes parallel to the cell faces and to the direction of gravity. To generate the results that follow we used a procedure which is equivalent to using 3072 pieces of data.[2]

Before considering the variances, it is instructive to look at the behavior of Γ_j^w defined in (21) and $\Gamma_{\mathbf{k}}^w$ defined in (24). The horizontal lines in the left panel of Fig. 6 show Γ_j^w for volume fractions of 15% (upper line) and 45%. As noted before, this quantity is actually independent of j. As the particle separation increases, $\langle w^y w^z\rangle \to (\langle w\rangle)^2$, so that $\langle w^y\, w^z\rangle - \langle w^2\rangle \to -\mathrm{var}[w] < 0$. This circumstance justifies the fact that the result is negative. The upper lines in the same figure show $\Gamma_{\mathbf{k}}^w$. For $k = k_1$ the result is positive, because the negative contribution of $\langle w^y w^z\rangle - \langle w^2\rangle$ is balanced by the negative sign of the cosine.[3] For larger values of k_m, however, the oscillating nature of the cosine quickly makes $\Gamma_{\mathbf{k}}^w$ very small.

Figs. 7 and 8 show on a log scale the variance for the vertical sedimentation velocity for volume fractions of 15% and 45%, respectively. The quantity plotted is normalized by multiplication by $(a^3/w_0)^2$, where w_0 is the settling velocity of a single particle in an unbounded fluid; explicitly, $w_0 = (2/9)(a^2 g/\nu)(\rho_\mathrm{p} - \rho_\mathrm{F})/\rho_\mathrm{F}$, with g the acceleration of gravity, ν the fluid kinematic viscosity and ρ_p, ρ_F the particle and fluid densities, respectively,

In each figure, the left panel is for $k_1 = 2\pi/L$ and the right one for $k_4 = 4 \times 2\pi/L$. The open symbols are the theoretical estimate (20) and the closed ones and the line are the numerical results. In the first case,

[2] We generated 512 distinct particle arrangements. For each one of these, we calculated the settling velocity of each particle with the applied force directed, in turn, along the three sides of the fundamental cell, which effectively results in $3 \times 516 = 1536$ realizations. For each one of these, we took $\mathbf{k}_\mathrm{m}$ in the two directions perpendicular to the applied force. Each point shown is therefore the average of $2 \times 1{,}536 = 3{,}072$ pieces of data.

[3] Note that, due to the cutoff imposed by g_0 when $r \leqslant 2a$, the average of the cosine over the integration domain is negative.

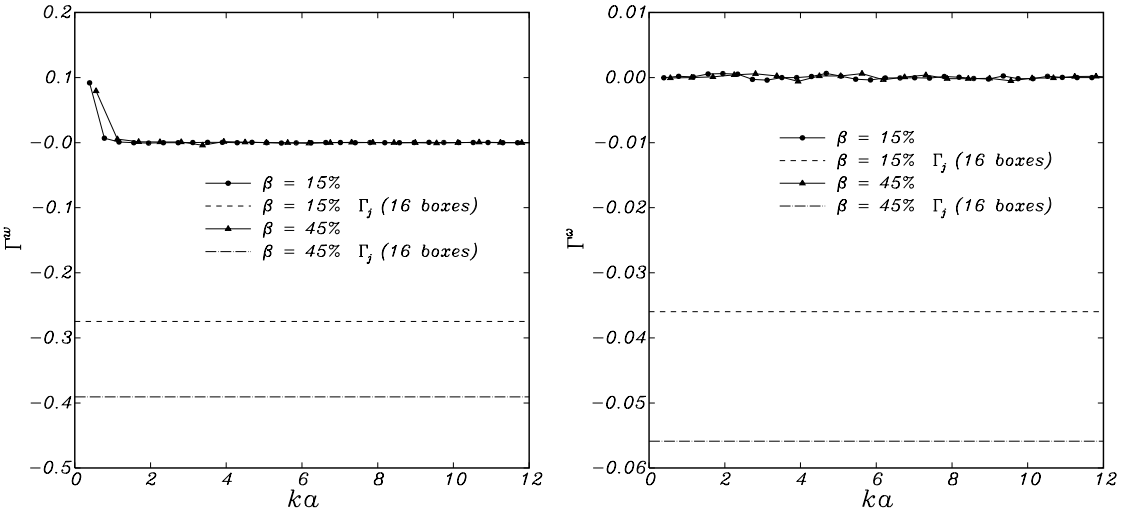

Fig. 6. The quantities Γ_j (lower horizontal lines) and $\Gamma_{\mathbf{k}}$ defined in (21) and (24) for volume fractions of 15% and 45%. The left figure is for the sedimentation velocity and the right one for the applied couple.

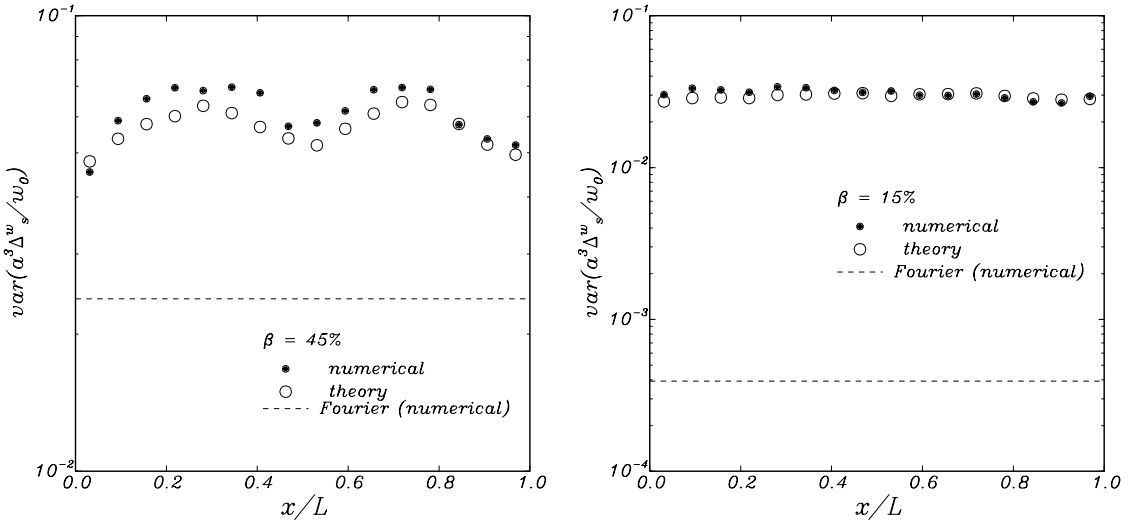

Fig. 7. Direct- (symbols) and Fourier-method (dashed line) variance of the normalized vertical velocity for sedimenting particles in the Stokes flow regime for $k_1 = 2\pi/L$ (left) and $k_4 = 4 \times 2\pi/L$. The open symbols are the theoretical estimate (20) and the closed ones and the line are the numerical results. The volume fraction is 15% and the number of boxes 16, with 150 particles and 3072 realizations.

$\frac{1}{2}(1 + \Gamma_j^w)$ is numerically not too different from $\Gamma_{\mathbf{k}}^w$ and these terms dominate in the respective expressions for the variances. Hence, the two variances are relatively close. However, for larger k, $\Gamma_{\mathbf{k}}^w$ is very close to zero and the variance for the Fourier method accordingly much smaller than that for the direct method.

We turn now to the case of particles rotating under the action of an equal couple $\mathbf{T}$ applied to each one. The Stokes velocity field induced by an isolated rotating particle decays proportionally to the inverse square of the distance (see e.g. [15,17]) and, therefore, this case is intermediate between the number density and the sedimentation velocity from the point of view of the range of the particle–particle interaction. As before, we

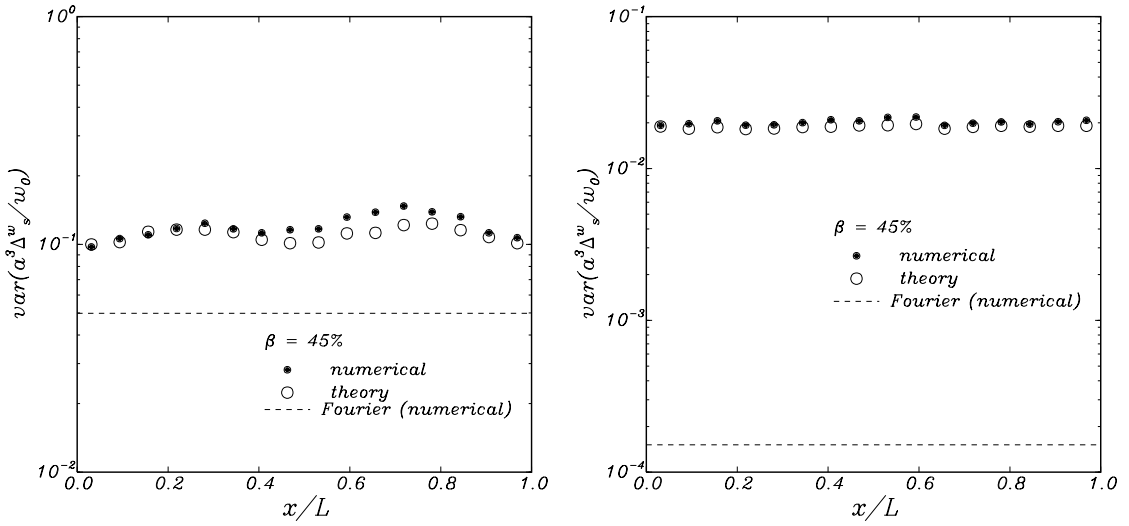

Fig. 8. Direct- (symbols) and Fourier-method (dashed line) variance of the normalized vertical velocity for sedimenting particles in the Stokes flow regime for $k = 2\pi/L$ (left) and $k = 4 \times 2\pi/L$. The open symbols are the theoretical estimate (20) and the closed ones and the line are the numerical results. The volume fraction is 45% and the number of boxes 16, with 150 particles and 3072 realizations.

use 512 distinct particle arrangements, for each one of which we calculate the angular velocity of each particle, with the couple **T** directed, in turn, along the three sides of the fundamental cell. This procedure again results in $3 \times 516 = 1536$ different realizations. Again as before, for each one of these, we take $\mathbf{k}_m$ in the two directions perpendicular to the applied torque. The results for the angular velocity variance are shown in Figs. 9 and 10 normalized by $(a^3/\omega_0)^2$, where $\omega_0 = T/(8\pi\mu a^3)$, with μ the fluid dynamic viscosity, is the angular velocity of an isolated particle subject to the couple T.

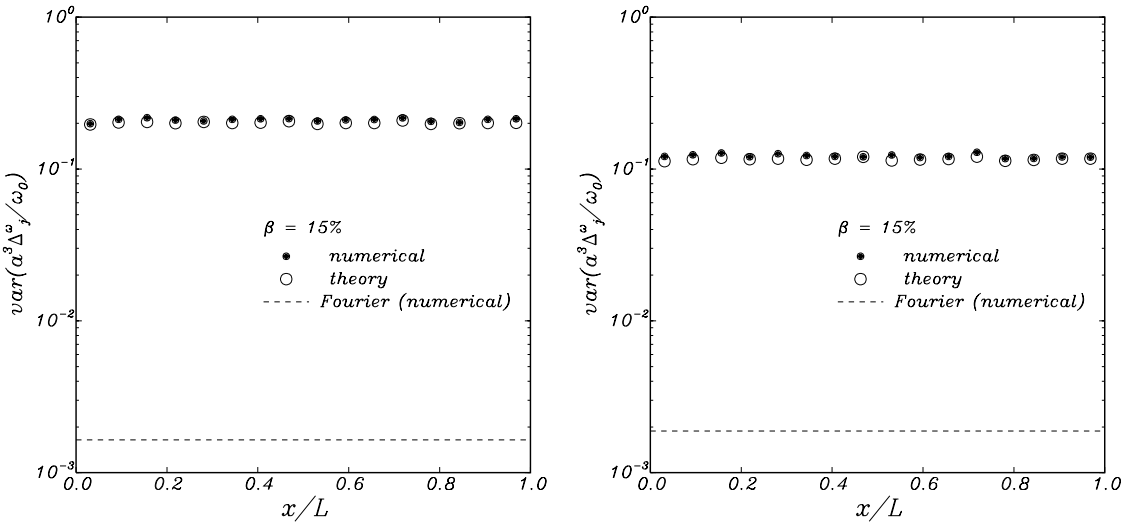

Fig. 9. Direct- (symbols) and Fourier-method (dashed line) variance of the normalized angular velocity for particles subject to a couple in the Stokes flow regime for $k = 2\pi/L$ (left) and $k = 4 \times 2\pi/L$. The open symbols are the theoretical estimate (20) and the closed ones and the line are the numerical results. The volume fraction is 15% and the number of boxes 16, with 150 particles and 3072 realizations.

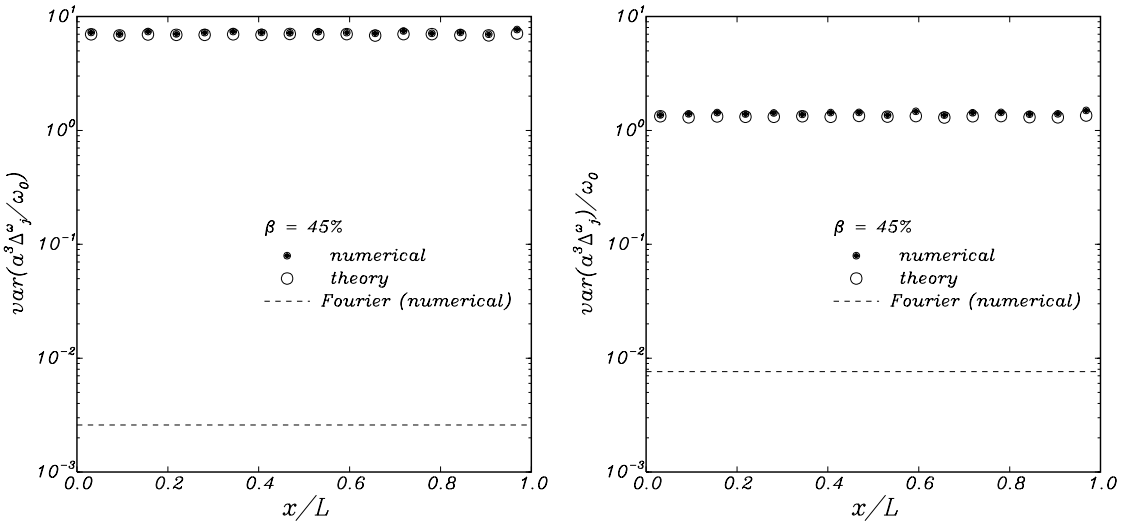

Fig. 10. Direct- (symbols) and Fourier-method (dashed line) variance of the normalized angular velocity for particles subject to a couple in the Stokes flow regime for $k = 2\pi/L$ (left) and $k = 4 \times 2\pi/L$. The open symbols are the theoretical estimate (20) and the closed ones and the line are the numerical results. The volume fraction is 45% and the number of boxes 16, with 3072 realizations of 150 particles.

The right panel in Fig. 6 shows $\Gamma^{\omega}_{\mathbf{k}}$ and Γ^{ω}_{j} for this case, also for volume fractions of 15% and 45%. The former is very close to zero for all k_{m}s, while the latter is negative as before, although about one order of magnitude smaller than for sedimentation. Such very small values of the Γ's make this case very similar to the number density studied in the previous section. The direct-method variance is very nearly constant and the approximate estimate of Section 3 (open circles) reproduces very well the numerical results (black circles). The variance for the Fourier method is about two orders of magnitude smaller. For a volume fraction

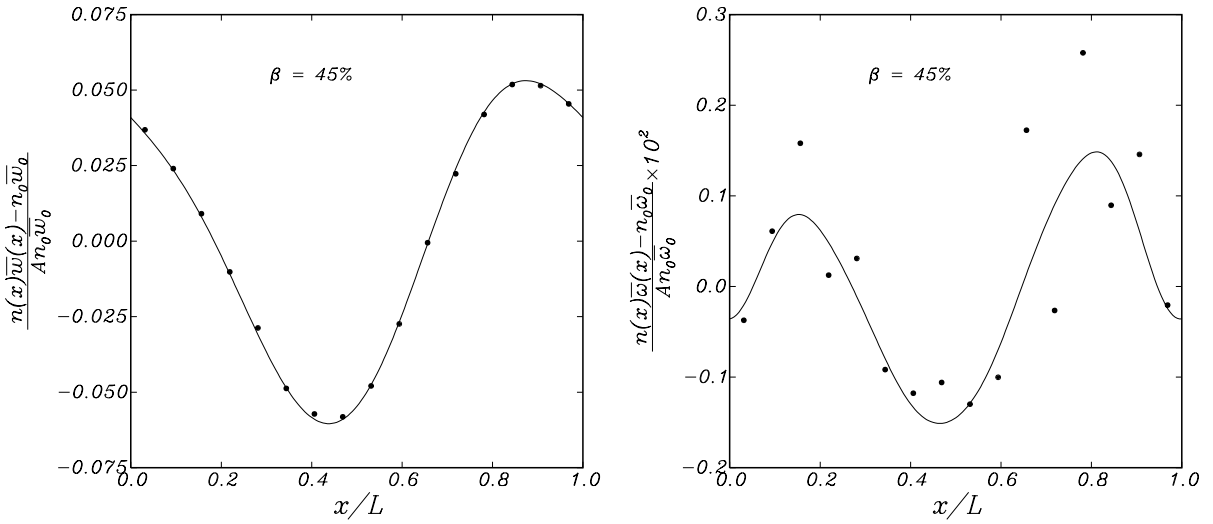

Fig. 11. Comparison of the normalized vertical (left) and angular velocity for the number density (42) as calculated by the Fourier method (line) and directly (points) for a volume fraction of 45% with 3072 realizations of 150 particles.

of 45% (Fig. 10) the situation is very similar with an even larger difference between the two variances. Unlike the vertical velocity case, the difference between the variances calculated in the two ways is smaller for the larger wave number case (right panels in Figs. 9 and 10). This result is a consequence of the fact that, for the velocity, $\Gamma^w_{\mathbf{k}}$ is positive for small $k = 2\pi/L$ and nearly zero for larger k, which decreases the variance with increasing k. For the angular velocity, instead, the $\Gamma^\omega_{\mathbf{k}}$ is small for all k, although its effect is magnified by the multiplication by the particle number (see Eqs. (22) and (23)); the variance of the direct method, however, decreases as S_0 in Eq. (20) increases as k is increased from $2\pi/L$ to $4 \times 2\pi/L$. In any case, however, it is obvious that the converging trend of the two variances with increasing k does not continue as, for large k, $S_0 \simeq 1$ and $\Gamma^\omega_{\mathbf{k}} \to 0$. Thus, the direct-method variance will always be larger than that for the Fourier method.

For the case of the polynomial number density (42), one would expect the Fourier method to perform similarly to the direct one for the velocity, but to be superior for the angular velocity for which the correlation between neighboring particles is relatively weak. This expectation is borne out by the numerical results shown in Fig. 11 calculated with 3072 realizations for a volume fraction of 45%; the results for 15% are very similar. The quantities shown in these figures are normalized by w_0/a^3 and ω_0/a^3 as before.

7. Conclusions

The frequently applied reduction of ensemble averages to volume averages rests on an assumption of ergodicity which fails in the absence of a strong separation between micro-and macro-scales. This situation is encountered e.g. in the case of spatially non-uniform suspensions. In cases such as these, ensemble averages must be dealt with as such and no simplification is possible.

In practice, the problem with the direct calculation of ensemble averages according to their definition lies in the relatively slow convergence – inversely proportional to the square root of the number of realizations – which therefore requires the use of large ensembles. In this paper, we have presented a more efficient method based on the use of a Fourier expansion of the averages. We have presented an approximate analytical calculation which shows that the variance of the Fourier averages is much smaller than that of the direct averages when the particle–particle interaction decays rapidly. This smaller variance results in a much faster convergence of the averages. The analytical estimates have been substantiated numerically with examples taken from the theory of Stokes flow. We find that, for a quantity like the sedimentation velocity (which, in the case of an isolated sphere, produces a disturbance decaying inversely with the distance), the variance for the Fourier method is only about a factor of 2 smaller than that of the direct method (Figs. 7 and 8). In this case, there is little difference between the Fourier and direct methods (Fig. 11, left panel). However, for a quantity such as the rotational velocity, which produces a more rapidly decaying disturbance in the fluid (proportional to the square of the inverse distance for an isolated sphere), the two variances differ by about two orders of magnitude (Figs. 9 and 10) and the Fourier methods outperforms the direct approach (see e.g. Fig. 11, right panel). Even in the former case, however, when the length scale of the spatial non-uniformity is reduced, the Fourier variance rapidly falls orders of magnitude below that of the direct method.

We have also demonstrated a method by which the probability distribution of a uniform ensemble can be biased so as to mimic that of a non-uniform one. This technique permits the study of the effect of non-homogeneity under controlled conditions.

Acknowledgement

Support by DOE under grant DE-FG02-99ER14966 is gratefully acknowledged.

Appendix A. The non-uniform ensemble

It was mentioned in the text that it is possible to choose the weights in (10) in such a way that the number density $n(\mathbf{x})$ defined in (1) has a prescribed form. It is readily verified that this objective is realized by the rule

$$W(C^N) = 1 + \frac{1}{n_0} \sum_{\mathbf{k}\neq 0} \frac{\tilde{n}(\mathbf{k})}{S_0(\mathbf{k})} \sum_{\alpha=1}^{N} \exp(-\mathrm{i}\mathbf{k}\cdot\mathbf{y}^\alpha), \tag{A.1}$$

where

$$n_0 = \frac{N}{V} \tag{A.2}$$

is the average number density of the particles in the fundamental cell, $\tilde{n}(\mathbf{k})$ the Fourier coefficient of the prescribed number density $n(\mathbf{x})$,[4] and $S_0(\mathbf{k})$ is the structure factor of the uniform distribution defined by [8,9]

$$S_0(\mathbf{k}) = \frac{1}{N}\left\langle \sum_{\alpha,\beta=1}^{N} \exp i\mathbf{k}\cdot(\mathbf{y}^\alpha - \mathbf{y}^\beta) \right\rangle_0 = \frac{1}{N}\left\langle \sum_{\alpha,\beta=1}^{N} \cos\mathbf{k}\cdot(\mathbf{y}^\alpha - \mathbf{y}^\beta) \right\rangle_0, \tag{A.3}$$

where, as before, the index 0 appended to the averaging symbol denotes that the average is carried out according to the uniform probability distribution $P_0(C^N)$. The proof of (A.1) for $\mathbf{k}=0$ is trivial. For $\mathbf{k}\neq 0$, it is sufficient to note that, from (5),

$$\tilde{n}(\mathbf{k}) = \frac{1}{N!}\int \mathrm{d}C^N P(C^N)\left[\frac{1}{V}\sum_{\beta=1}^{N}\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{y}^\beta)\right]. \tag{A.4}$$

Upon substituting here (10) and (A.1), we find[5]

$$\begin{aligned}\tilde{n}(\mathbf{k}) &= \frac{1}{N!}\int \mathrm{d}C^N P_0(C^N)\left[1+\frac{1}{n_0}\sum_{\mathbf{k}'}\frac{\tilde{n}(\mathbf{k}')}{S_0(\mathbf{k}')}\sum_{\alpha=1}^{N}\exp(-\mathrm{i}\mathbf{k}'\cdot\mathbf{y}^\alpha)\right]\left[\frac{1}{V}\sum_{\beta=1}^{N}\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{y}^\beta)\right] \\ &= \frac{1}{N}\sum_{\mathbf{k}'}\frac{\tilde{n}(\mathbf{k}')}{S_0(\mathbf{k}')}\left\langle\sum_{\alpha,\beta=1}^{N}\exp[i(\mathbf{k}\cdot\mathbf{y}^\beta - \mathbf{k}'\cdot\mathbf{y}^\alpha)]\right\rangle_0.\end{aligned} \tag{A.5}$$

But it is easy to show that

$$\left\langle\sum_{\alpha,\beta=1}^{N}\exp[\mathrm{i}(\mathbf{k}\cdot\mathbf{y}^\beta - \mathbf{k}'\cdot\mathbf{y}^\alpha)]\right\rangle_0 = N\delta_{\mathbf{k},\mathbf{k}'}S_0(\mathbf{k}') \tag{A.6}$$

with which (A.5) reduces to the identity $\tilde{n}(\mathbf{k}) = \tilde{n}(\mathbf{k})$. Indeed, upon setting in (A.5) $\mathbf{y}^\alpha = \tilde{\mathbf{y}}^\alpha + \mathbf{Y}$, where $\mathbf{Y}$ is an arbitrary vector equal for all the particles, we find

[4] It may be noted that the summation over α is proportional to the Fourier coefficient of the microscopic number density $v(\mathbf{x})$ of the realization C_i defined by

$$v_i(\mathbf{x}) = \sum_{\alpha=1}^{N}\delta(\mathbf{x}-\mathbf{y}_i^\alpha).$$

[5] Note that, for $\mathbf{k}\neq 0$, the 1 in the square brackets contributes nothing.

$$\left\langle \sum_{\alpha,\beta=1}^{N} \exp[\mathrm{i}(\mathbf{k}\cdot\mathbf{y}^{\beta} - \mathbf{k}'\cdot\mathbf{y}^{\alpha})] \right\rangle_0 = \exp[\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{Y}] \left\langle \sum_{\alpha,\beta=1}^{N} \exp[\mathrm{i}(\mathbf{k}\cdot\tilde{\mathbf{y}}^{\beta} - \mathbf{k}'\cdot\tilde{\mathbf{y}}^{\alpha})] \right\rangle_0 . \tag{A.7}$$

But, since the ensemble is uniform and periodic, for each realization, it must contain all the realizations obtained by a rigid translation of the particles. Thus the averages on the two sides of (A.7) must be identical which, by the definition (A.3) and the arbitrariness of $\mathbf{Y}$, implies (A.6).[6]

It may be noted that the choice (A.1) for the weights does not insure a positive probability for any set of Fourier coefficients $\tilde{n}(\mathbf{k})$. However for sufficiently small $\tilde{n}$'s, W will be positive and, therefore, (10) will be an acceptable probability distribution. As an example, Eq. (12) clearly shows that $P > 0$ provided $n_s/n_0 < S_0/N$. Although in this study we use (10) only to bias the probability distribution so as to reproduce a spatially non-uniform particle number density, a similar device can be developed to bias the statistical distribution of other independent variables such as the particle velocity.

When P_0 only depends on particle positions–which is the case we consider explicitly in this paper – since all the realizations that are translates of the same realization must have the same probability, it follows that $P_0(C^N)$ can only depend upon the differences $\mathbf{y}^2 - \mathbf{y}^1$, $\mathbf{y}^3 - \mathbf{y}^1$, etc., rather than separately on $\mathbf{y}^1, \mathbf{y}^2, \ldots$. If P_0 depends on other arguments, the same conclusion holds after integrating over them.

Appendix B. Estimate of the variances

Here we prove the estimate (20) for the variance of the direct method and the estimates (22) and (23) for the variances of the Fourier coefficients.

B.1. Direct method

For the direct-sum method we wish to estimate the variance of (19). Let us define

$$Q_j = \frac{1}{\Delta V_j} \sum_{\mathbf{y}^{\alpha}\in\Delta V_j} q^{\alpha}. \tag{B.1}$$

From (19), the variance of interest, $\mathrm{var}[\Delta_j^q]$, can then be written as

$$\mathrm{var}[\Delta_j^q] = \langle \Phi^2 Q_j^2 \rangle_0 - (\langle \Phi Q_j \rangle_0)^2. \tag{B.2}$$

To estimate this quantity we observe that both Φ and Q_j are given by sums of random variables. If these variables were independent, according to the central limit theorem, the probability distribution of these sums would tend to the normal form as the number of terms – equal to the number of particles – tends to infinity. In our case, the random variables in the sums are not independent but one may disregard this fact as a first approximation, the accuracy of which will be checked numerically a posteriori. The first term in (B.2) can then be approximated by assuming a multivariate normal distribution of Φ and Q_j. According to a well-known result (see e.g. [18, p. 37]), if two random variables X_1, X_2 are jointly normally distributed, the expectation value of $\langle X_1^2 X_2^2 \rangle$ is given by

$$\langle X_1^2 X_2^2 \rangle = \left[\frac{\partial^2}{\partial^2 a} \frac{\partial^2}{\partial^2 b} \phi(a,b) \right]_{a,b=0}, \tag{B.3}$$

[6] If, as a consequence of the translation, a particle were to end up outside the fundamental cell, it would be replaced by its periodic image in the cell. But also this case one would have $\exp[\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{y}^{\alpha})] \to \exp[\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{Y})]\exp[\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{y}^{\alpha})]$ and, since all realization appear with equal probability, the same conclusion holds.

where ϕ is the characteristic function given by

$$\phi(a,b) = \exp\left[\mathrm{i}[\langle X_1\rangle a + \langle X_2\rangle b] - \frac{1}{2}\left(\sigma_{11}a^2 + \sigma_{22}b^2 + 2\sigma_{12}ab\right)\right], \tag{B.4}$$

where σ_{jj} is the variance of X_j and σ_{12} the covariance of X_1X_2. In the present application, $\langle \Phi\rangle_0 = 0$ and the previous formula gives then

$$\langle \Phi^2 Q_j^2\rangle_0 = 2\left[\langle \Phi(Q_j - \langle Q_j\rangle_0)\rangle_0\right]^2 + \sigma_{\Phi\Phi}\left[\sigma_{Q_jQ_j} + \left(\langle Q_j\rangle_0\right)^2\right] = 2\left(\langle \Phi Q_j\rangle_0\right)^2 + \langle \Phi^2\rangle_0\langle Q_j^2\rangle_0 \tag{B.5}$$

so that

$$\mathrm{var}[\Delta_j^q] = \langle \Phi^2\rangle_0\langle Q_j^2\rangle_0 + \left(\langle \Phi Q_j\rangle_0\right)^2 \tag{B.6}$$

or

$$\mathrm{var}[\Delta_j^q] = \langle \Phi^2\rangle_0\left\langle\left(\frac{1}{\Delta V_j}\sum_{\mathbf{y}^\alpha\in\Delta V_j} q^\alpha\right)^2\right\rangle_0 + (\Delta_j^q)^2. \tag{B.7}$$

The first term in the right-hand side of this expression may be estimated by noting that

$$\langle \Phi^2\rangle_0 = \frac{1}{S_0^2(\mathbf{k})}\left\langle \sum_{\alpha,\beta=1}^{N} \sin\mathbf{k}\cdot\mathbf{y}^\alpha \sin\mathbf{k}\cdot\mathbf{y}^\beta\right\rangle_0 = \frac{1}{2}\frac{N}{S_0(\mathbf{k})}. \tag{B.8}$$

For the other factor in the first term of (B.7), let

$$\delta_j^\alpha = \begin{cases} 1 & \text{if particle } \alpha \in \text{box } j, \\ 0 & \text{if particle } \alpha \notin \text{box } j. \end{cases} \tag{B.9}$$

Then we may write

$$\left\langle\left(\sum_{\mathbf{y}^\alpha\in\Delta V_j} q^\alpha\right)^2\right\rangle_0 = \left\langle\left(\sum_{\alpha=1}^{N}\delta_j^\alpha q^\alpha\right)^2\right\rangle_0. \tag{B.10}$$

But, according to the definition (13) of average

$$\begin{aligned}\left\langle\left(\sum_{\alpha=1}^{N} q^\alpha\delta_j^\alpha\right)^2\right\rangle_0 &= \frac{1}{N!}\int \mathrm{d}C^N P_0(C^N)\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\delta_j^\alpha\delta_j^\beta q^\alpha q^\beta \\ &= \frac{1}{N!}\int dC^N P_0(C^N)\left[N(q^y)^2\delta_j^y + N(N-1)q^y q^z\delta_j^y\delta_j^z\right],\end{aligned} \tag{B.11}$$

where we call $\mathbf{y}$ the position of particle 1, $\mathbf{z}$ the position of particle 2 and have used the identity of the particles and the fact that $(\delta_j)^2 = \delta_j$. Therefore, we find

$$\left\langle\left(\sum_{\alpha=1}^{N}\delta_j^\alpha q^\alpha\right)^2\right\rangle_0 = \int \mathrm{d}^3y P_0(\mathbf{y})\delta_j^1\langle q^2\rangle_1(\mathbf{y}) + \int \mathrm{d}^3y\int \mathrm{d}^3z P_0(\mathbf{y},\mathbf{z})\delta_j^y\delta_j^z\langle q^y q^z\rangle_2(\mathbf{y},\mathbf{z}), \tag{B.12}$$

where $\langle q^2\rangle_1(\mathbf{y})$ is the conditional average of q^2 with one particle fixed, i.e.

$$\langle q^2\rangle_1(\mathbf{y}) = \frac{1}{(N-1)!}\int \mathrm{d}C^{N-1}P_0(C^{N-1}|\mathbf{y})(q^y)^2 \tag{B.13}$$

in which $P_0(C^{N-1}|\mathbf{y})$ is the one-particle conditional probability density, i.e. the probability density for the configuration of the remaining $N-1$ particles given that one particle is centered at $\mathbf{y}$; $\langle q^y q^z\rangle_2$ is the

conditional average with 2 particles fixed defined similarly. From the translational invariance of the ensemble noted at the end of Appendix A, we deduce that $\langle q^2\rangle_1(\mathbf{y})$ is actually independent of $\mathbf{y}$ while $\langle q^y q^z\rangle_2(\mathbf{y},\mathbf{z}) = \langle q^y q^z\rangle_2(\mathbf{z}-\mathbf{y})$ and $P_0(\mathbf{y},\mathbf{z}) = n_0^2 g_0(\mathbf{z}-\mathbf{y})$, where g_0 is the pair distribution function of the uniform ensemble. For simplicity, from now on we assume that the boxes are all equal, N_b in number, so that $\Delta V_j = V/N_\mathrm{b}$. The δ symbols have the effect of limiting the integrations over box j and, therefore

$$\left\langle \left(\frac{1}{\Delta V_j} \sum_{\mathbf{y}^\alpha \in \Delta V_j} q^\alpha \right)^2 \right\rangle_0 = \frac{N_\mathrm{b} n_0^2}{N} \langle q^2\rangle_1 + \frac{n_0^2}{\Delta V_j^2} \int_{\Delta V_j} \mathrm{d}^3 y \int_{\Delta V_j} \mathrm{d}^3 z g_0(\mathbf{z}-\mathbf{y}) \langle q^y q^z\rangle_2(\mathbf{z}-\mathbf{y}). \tag{B.14}$$

Since $g_0(\mathbf{z}-\mathbf{y})$ and $\langle q^y q^z\rangle_2$ are both bounded and independent of ΔV_j, we have

$$\begin{aligned} \frac{1}{\Delta V_j^2} \left| \int_{V_j} \mathrm{d}^3 y \int_{V_j} \mathrm{d}^3 z g_0(\mathbf{z}-\mathbf{y}) \langle q^y q^z\rangle_2(\mathbf{y},\mathbf{z}) \right| &\leqslant \max_{\Delta V_j}[g_0(\mathbf{z}-\mathbf{y})] \max_{\Delta V_j}[|\langle q^y q^z\rangle_2(\mathbf{z}-\mathbf{y})|] \\ &\leqslant \max_{V}[g_0(\mathbf{z}-\mathbf{y})] \max_{V}[|\langle q^y q^z\rangle_2(\mathbf{z}-\mathbf{y})|]. \end{aligned} \tag{B.15}$$

The second term in (B.14), therefore, is bounded as the number of boxes N_b increases and we therefore conclude that, at least for N_b sufficiently large, the quantity in the left-hand side of (B.14) is a linear function of N_b.

In order to make further progress, it is convenient to introduce the quantity Γ_j^q defined in (21) with which

$$\left\langle \left(\frac{1}{\Delta V_j} \sum_{\mathbf{y}^\alpha \in \Delta V_j} q^\alpha \right)^2 \right\rangle_0 = n_0^2 \langle q^2\rangle_1 \left(\frac{N_\mathrm{b}}{N} + \Gamma_j^q + \frac{1}{\Delta V_j^2} \int_{\Delta V_j} \mathrm{d}^3 y \int_{\Delta V_j} \mathrm{d}^3 z g_0(\mathbf{z}-\mathbf{y}) \right). \tag{B.16}$$

To estimate the integral, we consider the case $q^\alpha = 1$, for which $\Gamma_j^q = 0$ and the left-hand side is simply $\langle (N_j/\Delta V_j)^2\rangle$. In this case this equation becomes

$$\left\langle \left(\frac{N_j}{\Delta V_j} \right)^2 \right\rangle_0 = n_0^2 \left(\frac{N_\mathrm{b}}{N} + \frac{1}{\Delta V_j^2} \int_{\Delta V_j} \mathrm{d}^3 y \int_{\Delta V_j} \mathrm{d}^3 z g_0(\mathbf{z}-\mathbf{y}) \right) \tag{B.17}$$

but also, by definition,

$$\left\langle \left(\frac{N_j}{\Delta V_j} \right)^2 \right\rangle_0 = \left(\frac{\langle N_j\rangle_0}{\Delta V_j} \right)^2 + \frac{\mathrm{var}[N_j]}{\Delta V_j^2} = \left(\frac{N}{N_\mathrm{b} \Delta V_j} \right)^2 + \frac{\mathrm{var}[N_j]}{\Delta V_j^2}. \tag{B.18}$$

For non-interacting particles, we have the well-known estimate [19]

$$\mathrm{var}[N_j] \simeq \frac{N}{N_\mathrm{b}} \left(1 - \frac{1}{N_\mathrm{b}} \right) \tag{B.19}$$

with which, upon comparing (B.17) and (B.18),

$$\frac{1}{\Delta V_j^2} \int_{\Delta V_j} \mathrm{d}^3 y \int_{\Delta V_j} \mathrm{d}^3 z g_0(\mathbf{z}-\mathbf{y}) \simeq \frac{N-1}{N} \simeq 1. \tag{B.20}$$

Note that, for equal boxes, Γ_j^q is actually independent of j but we have retained the index nevertheless to remind the reader that this quantity refers to the box-counting method. With these approximations and (B.8), (B.7) becomes the result (20) given in the text.

B.2. Fourier method

The same procedure leading to (B.6) can be applied to the calculation of the variance according to the Fourier method with the result

$$\mathrm{var}[\Delta_s^q(\mathbf{k})] = \langle \Phi^2 \rangle_0 \left\langle \left(\frac{2}{V} \sum_{\alpha=1}^{N} q^\alpha \sin \mathbf{k} \cdot \mathbf{y}^\alpha \right)^2 \right\rangle_0 + [\Delta_s^q(\mathbf{k})]^2. \tag{B.21}$$

The first term in this expression can be further developed in the same way as before to find

$$\begin{aligned} \left\langle \left(\frac{2}{V} \sum_{\alpha=1}^{N} q^\alpha \sin \mathbf{k} \cdot \mathbf{y}^\alpha \right)^2 \right\rangle_0 &= \frac{2n_0^2}{N} \langle q^2 \rangle_1 + \frac{4n_0^2}{V^2} \int \mathrm{d}^3 y \int \mathrm{d}^3 z \sin \mathbf{k} \cdot \mathbf{y} \sin \mathbf{k} \cdot \mathbf{z} g_0(\mathbf{z} - \mathbf{y}) \langle q^y q^z \rangle_2 (\mathbf{z} - \mathbf{y}) \\ &= \frac{2n_0^2}{N} \langle q^2 \rangle_1 + \frac{2n_0^2}{V} \int \mathrm{d}^3 r \cos \mathbf{k} \cdot \mathbf{r} g_0(\mathbf{r}) \langle q^y q^z \rangle_2 (\mathbf{r}) \\ &= \frac{2n_0^2}{N} \langle q^2 \rangle_1 [S_0(\mathbf{k}) + N \Gamma_{\mathbf{k}}^q], \end{aligned} \tag{B.22}$$

where $\Gamma_{\mathbf{k}}^q$ is defined in (24). With these results, (22) and (23) follow.

References

[1] G. Batchelor, The stress system in a suspension of force-free particles, J. Fluid Mech. 41 (1970) 545–570.
[2] N.G. Hadjiconstantinou, Hybrid atomistic-continuum formulations and the moving contact line problem, J. Comput. Phys. 154 (1999) 245–265.
[3] E.G. Flekkøy, G. Wagner, J. Feder, Hybrid model for combined particle and continuum dynamics, Europhys. Lett. 52 (2000) 271–276.
[4] W. Ren, W.E. Heterogeneous, multiscale method for the modeling of complex fluids and micro-fluidics, J. Comput. Phys. 204 (2005) 1–26.
[5] D. Gottlieb, S. Orszag, Numerical Analysis of Spectral Methods, SIAM, Philadelphia, 1977.
[6] B. Fornberg, A Practical Guide to Pseudospectral Methods, Cambridge University Press, Cambridge, UK., 1996.
[7] A. Papoulis, S.U. Pillai, Probability, Random Variables, and Stochastic Processes, foruth ed., McGraw-Hill, New York, 2004.
[8] U. Balucani, M. Zoppi, Dynamics of the Liquid State, Clarendon Press, Oxford, 1994.
[9] P. Chaikin, T. Lubensky, Principles of Condensed Matter Physics, Cambridge U.P., Cambridge U.K., 1995.
[10] J. Percus, G. Yevick, Analysis of classical statistical mechanics by means of collective coordinates, Phys. Rev. 110 (1958) 1–13.
[11] M.S. Wertheim, Exact solution of the Percus–Yevick integral equation for hard spheres, Phys. Rev. Lett. 10 (1963) 321–323.
[12] E. Thiele, Equation of state for hard spheres, J. Chem. Phys. 39 (1963) 474–479.
[13] K. Ichiki, A. Prosperetti, Faxén-like relations for a non-uniform suspension, Phys. Fluids 16 (2004) 2483–2496.
[14] G. Mo, A. Sangani, A method for computing Stokes flow interactions among spherical objects and its application to suspensions of drops and porous particles, Phys. Fluids 6 (1994) 1637–1652.
[15] H. Lamb, Hydrodynamics, 6th ed., Cambridge University Press, Cambridge, 1932.
[16] W. Russel, D. Saville, W. Schowalter, Colloidal Dispersions, Cambridge University Press, Cambridge, UK, 1989.
[17] S. Kim, S. Karrila, Microhydrodynamics, Butterworth-Heinemann, Boston, 1991.
[18] C.W. Gardiner, Handbook of Stochastic Methods, second ed., Springer, Berlin, 1985.
[19] R. Becker, Theory of Heat, Springer, Berlin, 1967.