



**HAL**  
open science

## Finite-size corrections in simulation of dipolar fluids

Luc Belloni, Joël Puibasset

► **To cite this version:**

Luc Belloni, Joël Puibasset. Finite-size corrections in simulation of dipolar fluids. The Journal of Chemical Physics, 2017, 147 (22), pp.224110. 10.1063/1.5005912 . hal-02382284

**HAL Id: hal-02382284**

**<https://hal.science/hal-02382284>**

Submitted on 27 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Finite-size corrections in simulation of dipolar fluids

Luc Belloni<sup>1</sup> and Joël Puibasset<sup>2</sup>

<sup>1</sup>LIONS, NIMBE, CEA, CNRS, Université Paris-Saclay, 91191-Gif-sur-Yvette, France

<sup>2</sup>ICMN, UMR 7374, CNRS et Université d'Orléans, 1B rue de la Férollerie, 45071-Orléans Cedex, France

### Abstract

Monte Carlo simulations of dipolar fluids are performed at different numbers of particles  $N=100-4000$ . For each size of the cubic cell, the non-spherically symmetric pair distribution function  $g(r, \Omega)$  is accumulated in terms of projections  $g^{ml}(r)$  onto rotational invariants. The observed  $N$  dependence is in very good agreement with the theoretical predictions for the finite-size corrections of different origins: the *explicit* corrections due to the absence of fluctuations in the number of particles within the canonical simulation and the *implicit* corrections due to the coupling between the environment around a given particle and that around its images in the neighboring cells. The latter dominate in fluids of strong dipolar coupling characterized by low compressibility and high dielectric constant. The ability to clean with great precision the simulation data from these corrections combined with the use of very powerful anisotropic integral equation techniques means that *exact* correlation functions both in real and Fourier spaces, Kirkwood-Buff integrals and bridge functions can be derived from box sizes as small as  $N \approx 100$ , even with existing long-range tails. In presence of dielectric discontinuity with the external medium surrounding the central box and its replica within the Ewald treatment of the coulombic interactions, the  $1/N$  dependence of the  $g^{ml}(r)$  is shown to disagree with the, yet well-accepted, prediction of the literature.

[luc.belloni@cea.fr](mailto:luc.belloni@cea.fr) ; [puibasset@cnrs-orleans.fr](mailto:puibasset@cnrs-orleans.fr)

## I. INTRODUCTION

The basic problem in statistical physics of liquids consists to derive the pair distribution function (*pdf*)  $g$  from the pair potential  $v$ . In principle, simulation approaches like Molecular Dynamics or Monte Carlo (MC) provide an exact solution. In practice, the situation is not so simple. Beyond the problem of statistical noise intrinsically always present in any numerical simulation, the "measured" profile  $g(r)$  is neither complete nor exact, due to the finite size  $L$  of the simulation (cubic) cell. First, information is available in a limited  $r$  range *only*, say  $r < r_{max} \approx L/2$ . For most standard simulations, that cut-off is not large compared to the correlation range, especially in dense systems and/or at high electrostatic coupling, and the non-negligible tail of  $g(r)$  is lacking. Yet, various kinds of studies require the complete knowledge of the *pdf* on the whole  $r$  domain or, equivalently, its Fourier transform, the structure factor  $S(q)$ , especially at low  $q$ : some are interested<sup>1</sup> in the so-called Kirkwood-Buff integrals of the *pdf* which are related to thermodynamic derivative properties<sup>2</sup>, the prototype

being the normalized isothermal compressibility  $\chi = \left( \frac{\partial \rho}{\partial \beta P} \right)_T = S(0) = 1 + \rho \int_0^\infty (g-1) d\bar{r}$  ( $\rho$  is

the number density,  $P$  the pressure and  $\beta=1/kT$  the inverse temperature); others need to extract the direct correlation function  $c(r)$  from *pdf* data through the Ornstein-Zernike equation<sup>3</sup> in order to feed density functional theories for describing solvation properties of molecular solutes<sup>4</sup>; last ones focus on the bridge function of central interest in integral equation approaches<sup>3</sup>. In most cases, simple truncation of the correlations beyond  $r_{max}$  is not sufficient. The brute method that consists to increase the size of the cell  $L$  with higher and higher numbers of simulated particles  $N (= \rho L^3)$  becomes rapidly prohibitive in computing time and alternative methods are desired. That opens to the general problem of extending simulation data beyond  $r_{max}$  with some extrapolation scheme, based for instance on approximate bare integral equations. Since the pioneer work of Verlet<sup>5</sup>, it has been shown for a variety of systems how coupling simulation data with standard hypernetted chain (HNC) or mean-spherical approximation (MSA) closures at large distances succeeds in describing the whole *pdf* self-consistently, from simple spherical fluids<sup>6</sup> up to more complex ones made of anisotropic particles<sup>4,7,8</sup> (note: it is the approximate closure which is assumed beyond  $r_{max}$ , *not* the long distance part of the *full* HNC or MSA solution; short and long distances are coupled via the Ornstein-Zernike equation). In particular, making use of all the technical machinery developed for the resolution of integral equations in the case of anisotropic potentials<sup>9,10,11</sup>, the authors have recently extracted the exact bridge function of dipolar fluids from such MC/HNC, MC/MSA approaches and found that HNC or MSA can be safely applied beyond cut-off distances as *low* as 2-3 diameters<sup>12</sup> (paper noted I in the following).

Does it mean that small boxes characterized by  $L/2$  values of this order of magnitude ( $N \approx 100$ ) are sufficient and can be used without precaution? Not really, owing to the second limitation of the simulation, the fact that the available information below  $r_{max}$  is corrupted by systematic errors. These so-called *finite-size* corrections have two origins. First, the *explicit* corrections are due to the absence of fluctuations in the number of particles within the canonical simulation. As a consequence, the *pdf* deviates from the bulk reference by a

function whose  $1/N$  leading term can be easily derived by linking canonical and grand-canonical ensembles and expressed in terms of pair distribution density-derivatives<sup>13</sup>. In particular, the well-known long distance asymptote of the *pdf* reads:

$$\lim_{\text{large separation}} g_N = 1 - \frac{\chi}{N} \quad (1)$$

The departure from the asymptote 1 in (1) simply means that the number of neighbors available to the far field environment around a given particle is  $N$  minus 1 (the central particle itself) and minus the excess particles locally adsorbed or desorbed around it (given by the integral of  $g-1$ ), so is  $N-\chi$ . In grand-canonical conditions, this finite-size correction exactly vanishes due to the exchange of particles with the reservoir and the equality  $\langle N^2 \rangle = \langle N \rangle^2 + \chi \langle N \rangle$ . Secondly, the *implicit* corrections are due to the coupling between the environment around a given particle and that around its periodic images in the neighboring cells. This means that the analysis of pair distances inside the central box does not provide the *pair* distribution exactly but rather some higher order  $n$ -body information. A simple superposition approximation predicts<sup>14</sup>:

$$g_N(\vec{r}) \approx g(\vec{r}) \prod_{\vec{p} \neq 0} g(\vec{r} - \vec{p}L) \quad (2)$$

where the vector  $\vec{p}$ , of integer components, describes the array of cell replica. In presence of long-range correlations, the bulk *pdf*  $g$  has not reached its asymptote 1 at the cell edge and "overflows" into the first image boxes. The actual  $g_N$  measures the sum of all these "leaks". Note that (2), which does *not* represent a  $1/N$  term in some expansion<sup>15</sup>, involves the long range behavior of the *pdf* which, by definition, cannot be extracted from the simulation itself and must be derived from a complementary model.

The measured  $N$ -dependence of simulation data has been successfully interpreted in terms of these two classes of finite-size corrections for a limited number of simple spherical potentials describing model argon<sup>16</sup> and krypton<sup>17 15</sup> fluids, for which the *pdf*  $g(r)$  depends on the distance  $r$  *only*. On the other hand, to the author's knowledge, nothing has been done in that respect for *non-spherical* particles. The reason is certainly due to the numerical difficulty to deal with correlation functions that depend now on separation  $r$  *and* orientations  $\Omega$ , namely the (relative) orientations of the two particles  $\Omega_1, \Omega_2$  and the vector joining them  $\hat{r} \equiv \hat{r}_{12}$ . As usual in the treatment of anisotropic particles, rather to manipulate the complete  $g(r, \Omega)$  as explicit function of the different Euler angles, it is more fruitful to project it onto a basis of carefully chosen angular functions and to play with the projections that depend on the separation  $r$  *only*. Following Blum's notation and normalization<sup>18 19</sup>:

$$g(r, \Omega) \equiv g(r, \Omega_1, \Omega_2, \hat{r}) = \sum_{m,n,l,\mu,\nu} g_{\mu\nu}^{mnl}(r) \Phi_{\mu\nu}^{mnl}(\Omega_1, \Omega_2, \hat{r}) \quad (3)$$

with

$$\Phi_{\mu\nu}^{mnl}(\Omega_1, \Omega_2, \hat{r}) = \sqrt{(2m+1)(2n+1)} \sum_{\mu', \nu', \lambda'} \begin{pmatrix} m & n & l \\ \mu' & \nu' & \lambda' \end{pmatrix} R_{\mu'\mu}^m(\Omega_1) R_{\nu'\nu}^n(\Omega_2) R_{\lambda'0}^l(\hat{r}) \quad (4)$$

The coefficients  $\begin{pmatrix} m & n & l \\ \mu' & \nu' & \lambda' \end{pmatrix}$  are the usual 3-j-symbols. The  $R_{\mu'\mu}^m(\Omega)$  are Wigner generalized spherical harmonics (definition and notation from Messiah<sup>20</sup>). The  $\Phi_{\mu\nu}^{mnl}$  rotational invariants (independent of the reference frame) form an orthogonal basis. They depend on the relative orientation of the two molecules and of the vector joining them (five Euler angles) and are characterized by five indices  $m, n, l, \mu, \nu$ . For *linear* particles (axis noted  $\hat{\mu}$ ) as in the present dipolar fluid study, three Euler angles are sufficient,  $m+n+l$  is even and  $\mu=\nu=0$ . In the following, although much of the analysis will be valid for any anisotropy, the indices  $\mu, \nu$  will be dropped to simplify the notation (in that case,  $R_{\mu 0}^m \equiv \sqrt{4\pi/(2m+1)} Y_m^{\mu*}$  where  $Y_m^{\mu}$  is the standard spherical harmonics). The first terms are  $\Phi^{000} = 1, \Phi^{110} = -\sqrt{3}\hat{\mu}_1\hat{\mu}_2, \Phi^{112} = \sqrt{3/10}(3(\hat{\mu}_1\hat{r})(\hat{\mu}_2\hat{r}) - \hat{\mu}_1\hat{\mu}_2), \dots$  (note the coefficients  $-\sqrt{3}$  and  $\sqrt{3/10}$  instead of the value 1 for the definition of the  $\Phi^{110}$  and  $\Phi^{112}$  invariants, respectively, which results from Blum's normalization; that induces the inverse coefficients in the corresponding projections). Inversely, the coefficients  $g^{mnl}(r)$  are derived from the complete function by angular projection:

$$g^{mnl}(r) = \iiint g(r, \Omega) \Phi^{mnl*}(\Omega) d\Omega / \iiint |\Phi^{mnl}|^2(\Omega) d\Omega \equiv (2l+1) \iiint g(r, \Omega) \Phi^{mnl*}(\Omega) d\Omega \quad (5)$$

In short notation, if  $\Phi_\alpha$  stands for  $\Phi^{mnl}$ , expansion and projection read, respectively:

$$g(r, \Omega) = \sum_{\alpha} g_{\alpha}(r) \Phi_{\alpha}(\Omega) \quad (6)$$

$$g_{\alpha}(r) = \left\langle g(r, \Omega) \Phi_{\alpha}^*(\Omega) \right\rangle \quad (7)$$

where the brackets in (7) means a (normalized) triple angular integral. The first projection  $g_1 \equiv g^{000}$  represents the center of mass-center of mass *pdf* (averaged over all orientations).

In principle, the basis is infinite. In practice, it is assumed and verified that a limited number  $\alpha_{max}$  of projections, characterized by  $m, n \leq n_{max}$ , is sufficient, at least for the excess potential of mean force, to quantitatively capture the angular dependence of the correlations, see Paper I<sup>12</sup>.

The purpose of the present study is to calculate the theoretical finite-size corrections of different origins and quantitatively analyze the  $N$ -dependence of MC  $g^{mnl}(r)$  *pdf* data for the case of *dipolar* fluids. Beside the interest in understanding these subtle effects, the ultimate goal is to be able to correct very easily cheap MC data (cheap means here performed at low  $N$  values) for feeding MC/HNC, MC/MSA approaches and deriving the complete exact *pdf*  $g$  and structure factor  $S$  on the whole  $r$ - $q$  domains.

The pair potential between spheres carrying a point-like dipole moment  $\mu$  is:

$$v(r, \Omega) = v_{short-range}(r) - \frac{\mu^2}{4\pi\epsilon_0 r^3} (3(\hat{\mu}_1 \cdot \hat{r})(\hat{\mu}_2 \cdot \hat{r}) - (\hat{\mu}_1 \cdot \hat{\mu}_2)) \quad (8)$$

The spherically symmetric  $v_{short-range}$  contributes to  $v^{000}$  while the dipole-dipole potential reads  $v^{112}(r) = -\sqrt{10/3} \frac{\mu^2}{4\pi\epsilon_0 r^3}$ . For such purely dipolar anisotropy,  $(\bar{\mu}_1, \bar{\mu}_2)$  and  $(-\bar{\mu}_1, -\bar{\mu}_2)$  configurations are equivalent and the projections  $m, n, l$  are restricted to *even*  $l$  (and so even  $m+n$ ). We have analyzed in great details the finite-size corrections for *two* systems which differ by the form of the short-range potential and the strength of the dipolar coupling. *System I* corresponds to the Lennard-Jones (LJ) form (Stockmayer fluid):

$$v_{short-range}(r) = v_{LJ}(r) = 4\epsilon \left( \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right) \quad (9)$$

$\sigma$  and  $\epsilon$  are the usual diameter and energy LJ parameters. The LJ potential is truncated, shifted at  $r_c = 2.5\sigma$ . The thermodynamical state, previously studied in paper I<sup>12</sup>, corresponds to

$$T^* \equiv kT/\epsilon = 2, \quad \rho^* \equiv \rho\sigma^3 = 0.8, \quad \mu^{*2} \equiv \frac{\mu^2}{4\pi\epsilon_0 \sigma^3 kT} = 1.5$$

and mimics a dense polar solvent of moderate dielectric constant,  $\epsilon \approx 16$ . For *system II*, we have chosen the soft-sphere (SS) dipolar fluid which has been intensively studied by Kusalik<sup>21 22</sup> using molecular dynamics at various  $N$ :

$$v_{short-range}(r) = v_{SS}(r) = 4\epsilon \left( \frac{\sigma}{r} \right)^{12} \quad (10)$$

with  $T^* = 1.35$ ,  $\rho^* = 0.8$  and  $\mu^{*2} = 4/1.35 \approx 2.963\dots$ . The SS potential is truncated, shifted at  $r_c = 2.5\sigma$  as well. This second dense polar solvent is controlled by strong short and long range correlations of coulombic origin. In particular, Kusalik has nicely shown how simulation cells as large as  $N=4000$  ( $L/2 \approx 9\sigma$ ) were necessary to be able to extract the high dielectric constant with reasonable accuracy,  $\epsilon \approx 99 \pm 4$ <sup>22</sup>.

An additional source of finite-size corrections, specific to electrostatic systems, is related to the possible dielectric discontinuity with the external medium surrounding the central box and its replica. Indeed, the Ewald treatment of the long range dipole-dipole interaction in periodic boundary geometry involves *explicitly* the dielectric constant  $\epsilon'$  of that medium. When  $\epsilon'$  differs from  $\epsilon$  (unknown at the beginning), the *pdf* exhibits finite-size polarization effects. The leading  $1/N$  contribution has been given for the first 3 projections  $g^{mnl}$  by de Leeuw, Perram and Smith in a seminal paper<sup>23</sup> (noted DPS in the following).

The paper is organized as follows. The standard MC simulation is briefly recalled in section II. Section III will show how the measured  $\epsilon'$ -dependence disagrees with the short-range part of the DPS theory. Section IV analyses the explicit corrections by calculating the density-derivatives of the *pdf* projections, again with MC. The evaluation of the implicit

corrections in Section V is the most delicate part of the whole analysis because it requires the *pdf* at large distances (provided by the mixed MC/integral equations of paper I<sup>12</sup>) and needs to project n-body distributions as in (2) onto rotational invariants. It will be shown that the superposition approximation for the 3-body distribution refined with the first elementary diagram correction which involves 4 particles leads to very good agreement between MC data and theoretical finite-size corrections, even at the  $10^{-3}$  level of precision! As a conclusion and illuminating illustration (Section VI), we will naively perform a *cheap* study of the Kusalik's system II, associating MC data restricted to  $N=108$  ( $L/2 \approx 2.5\sigma$ ) with the MC/MSA mixed integral equation. Starting from the brute MC data, the dielectric properties are strongly overestimated ( $\approx 167!$ ). On the other hand, beforehand correct cleaning from finite-size corrections calculated in a self-consistent manner provides the *complete* long-range dielectric properties with great accuracy.

## II. Monte Carlo simulation

Since the analysis will focus in detail on *differences* between simulations performed at various  $N$  or  $\varepsilon'$ , each of the brute MC data sets must require a precision much more restrictive than in a standard study. That not only means long simulations but also correct control of all possible bias.

$N \approx 100$ -4000 dipolar particles are put in a cubic cell of edge  $L \approx 5$ -17 $\sigma$  with periodic boundary conditions ( $N, V, T$  canonical ensemble). The dipole-dipole interaction is evaluated using the Ewald summation technique<sup>24</sup> and becomes:

$$v_{dd}^{Ewald}(12) = -(\bar{\mu}_1 \bar{\nabla})(\bar{\mu}_2 \bar{\nabla})\psi(\bar{r}) + \frac{4\pi}{L^3} \frac{1}{2\varepsilon'+1} \bar{\mu}_1 \bar{\mu}_2 \quad (11)$$

in which the usual charge-charge Ewald potential is split in:

$$\psi(\bar{r}) = \sum_{\bar{p}} \frac{\text{erfc}(\kappa|\bar{r} + \bar{p}L|)}{|\bar{r} + \bar{p}L|} + \frac{1}{\pi L} \sum_{\bar{m} \neq 0} \frac{1}{m^2} e^{-(\pi m/\kappa L)^2} e^{2i\pi \bar{m}\bar{r}/L} - \frac{\pi}{\kappa^2 L^3} \quad (12)$$

The dielectric constant  $\varepsilon'$  of the external medium appears explicitly through the last term of (11). The sum in the  $r$  space is spherically cut at  $r=L/2$ , so only the minimum image distance  $\mathbf{p}=\mathbf{0}$  is retained. The screening constant  $\kappa$  is chosen such that the factor  $\kappa r$  at the cut is equal to the high, constraining value  $s=4$ , so  $\kappa L=8$ . In the same way, the largest  $\mathbf{m}$  vector retained in the Fourier sum is similarly chosen such that  $\pi m/\kappa L=s=4$ , so  $m_{max}=4\kappa L/\pi=10$  (647 independent  $\mathbf{m}$  vectors). With this choice, the relative precision in the dipolar energy is better than  $10^{-6}$ , each sum contributing on equal level to the uncertainty. Note that this precision is much better than the standard ones given in the literature which correspond to lower values

( $\approx 3$ ) for the  $s$  parameter. With this strong constraint, we are sure to avoid any bias related to a limited precision in the Ewald summation<sup>25</sup>.

Each trial displacement is made of a local translation inside a cube (of size  $\approx 0.2\sigma$ ) and of a rotation of the dipole axis inside a cone (of angle  $\approx 40-60^\circ$ ) and is accepted according to the Metropolis method. After equilibration, the projections  $g^{mnl}(r)$  are constructed from the relative positions and orientations of the particles in the box. For each retained configuration, the minimum image separations  $r_{ij}$  of all pairs  $i < j$  of dipoles are sorted into a histogram of width  $\delta r$ . The bin  $k$  corresponds to the distance interval  $I_k = [r_{k-1}, r_k]$  with  $r_k = k\delta r$ , and to the volume of the spherical cap  $V_k = 4\pi/3(r_k^3 - r_{k-1}^3)$ . So, following the definition (7), the projections are constructed according to:

$$g^{mnl}(\text{bin } k) = \frac{2l+1}{\frac{1}{2}N^2V_k} \sum_{\substack{i < j \\ r_{ij} \in I_k}} \Phi^{mnl*}(\hat{\mu}_i, \hat{\mu}_j, \hat{r}_{ij}) \quad (13)$$

Note the correct normalization factor  $\frac{1}{2}N^2$  (and not  $\frac{1}{2}N(N-1)$ ). Due to the finiteness of the width  $\delta r$ , that information in the bin  $k$  measures some average of  $g$  over the interval  $I_k$  and is attributed to the mean separation  $(r_{k-1} + r_k)/2$ . A good resolution of the *pdf* details is obtained with  $\delta r \approx \sigma/80$ . Gathering successive bins to reach a somewhat higher value ( $\approx \sigma/20$ ) enables to improve the local statistics of each point and the display of the noisy *pdf*-difference curves at the price of a reasonable convolution.

It is important to remember that the simulated fluid is not perfectly isotropic due to the periodic boundary conditions and the implicit finite-size corrections. So, the correlation ( $ij$ ) depends not only on the *relative* orientations of  $\hat{\mu}_i$ ,  $\hat{\mu}_j$  and  $\hat{r}_{ij}$  but also weakly on the *absolute* orientation of the ensemble with respect to the axes of the box. The expansion (3) in rotational invariants is recovered only by spherically averaging over that absolute orientation, as implicit in (13).

In the present study, 7 or 13 projections are usually accumulated, which correspond to  $n_{max}=2$  or 3, respectively. These finite bases are sufficient to describe the strong anisotropy of all correlation functions within the mixed MC/integral equation treatment, even for the highest dipolar strength<sup>12</sup>. For each pair, the spherical harmonics  $\Phi^{110}$ ,  $\Phi^{112}$  and  $\Phi^{022}$  (or  $\Phi^{202}$ ) are calculated from the scalar products between the three unit vectors. The remaining harmonics,  $\Phi^{221}$  (and  $\Phi^{131}$ ,  $\Phi^{331}$ ) are deduced from products of the previous ones. No manipulation of trigonometric functions is required. At the end, the MC simulation provides 7 or 13 projections of the *pdf* in the limited range  $r < r_{max} = L/2$ . Only in a few cases, we explored the corners of the box,  $L/2 < r < L\sqrt{3}/2$  ( $V_k$  in (13) becomes in that case the intersection of the spherical layer with the cube), keeping in mind that this extended set of data does not result from a complete spherical averaging and is (thus) sensitive to systematic bias<sup>26</sup>.

Thermodynamical quantities are calculated as usual. The dipolar contribution to the energy is the average of the whole Ewald sum. The virial pressure makes use of the simple  $1/r^3$  or  $1/L^3$  behavior for the dipolar interaction. The compressibility  $\chi$  is derived from the fluctuations of



the virial accumulated during the simulation <sup>24</sup> (with a proper account of the truncated LJ/SS force discontinuity at the cut-off  $r_c$  in the hypervirial function). This somewhat ignored route should be recommended in any study requiring exact  $S(0)$  with good accuracy. Discretized differentiations of the virial pressures measured at neighboring densities agree with the  $\chi$  values within statistical uncertainty. The dielectric constant  $\varepsilon$  of the fluid is related to the orientational order fluctuations of the box  $\langle M^2 \rangle$  ( $M$  is the total moment, of theoretical zero average  $\langle M \rangle$ ) through an equation which depends on  $\varepsilon'$  explicitly <sup>24</sup>:

$$\frac{(\varepsilon - 1)(2\varepsilon' + 1)}{2\varepsilon' + \varepsilon} = \frac{4\pi}{3} \frac{\langle M^2 \rangle}{4\pi\varepsilon_0 L^3 kT} = 3yg \quad (14)$$

where  $y = 4\pi\rho^* \mu^{*2}/9$ . The Kirkwood  $g$ -factor  $g = \langle M^2 \rangle / N\mu^2$  may also be expressed as the integral of  $g^{110}$  over the whole box:

$$g = \frac{1}{N\mu^2} \left\langle \left| \sum_{i=1}^N \vec{\mu}_i \right|^2 \right\rangle = 1 + (N-1) \langle \hat{\mu}_1 \hat{\mu}_2 \rangle = 1 - \frac{\rho}{\sqrt{3}} \int_{\text{box}} g^{110}(r) d\vec{r} \quad (15)$$

Equation (14) proves that the moment fluctuations and the  $g$ -factor must be functions of  $\varepsilon'$ ,  $g(\varepsilon')$ , even in the bulk limit <sup>23</sup>.

Tables 1 and 2 collect all information about the MC parameters and the thermodynamics for systems I and II, respectively. Note the unusual length of the simulations (a few  $10^7$  cycles, 1 cycle =  $N$  individual trial displacements) performed on desk computers and unusual precision of the *pdf* projections ( $\approx 10^{-4}$ - $10^{-3}$ ). The systematic dependences in  $N$  and  $\varepsilon'$  observed in the various MC data illustrate the presence of the finite size corrections. The variance in the thermodynamical quantities and in the *pdf* projections at some characteristic distances  $r$  has been very carefully monitored using standard block average analysis <sup>27</sup>. As shown by Kusalik <sup>21</sup>, the instantaneous moment  $M$  of highly polar fluids is subject to long lived fluctuations which are directly associated to *collective* fluctuations in the  $g^{110}$  projection on the whole  $r$  domain. That means that the uncertainty on  $g^{110}(r)$  may be (much) larger than that naively derived from the visual *local* noise which results from the finite histogram width  $\delta r$ . The error bar on the measured dielectric constant  $\varepsilon$  is consequently quite large. For instance, for System II with  $\varepsilon' = 99$ , the instantaneous quantity  $\varepsilon$  extracted from (14) has an average around 100, a RMS deviation of 120 and a statistical inefficiency around 350 (two configurations must be separated by at least this number of consecutive cycles to be statistically fully decorrelated <sup>24</sup>). So, a variance of the order 1 on  $\varepsilon$  (one standard deviation) requires  $5 \times 10^6$  cycles, see Table 2. This evaluation is roughly independent of  $N$ .

The curves of the most relevant *pdf* projections for systems I and II can be found in references <sup>12</sup> and <sup>22</sup>, respectively, and do not deserve special further comments. Plotting on the same graph projections accumulated at different  $N$  (or  $\varepsilon'$ ) is usually not the appropriate way to analyze finite-size corrections because the different curves are hardly distinguishable at the scale 1. So, in the following of the paper, and since we have no knowledge, at least in advance, of the exact *pdf* reference, free of any correction, the best we can do is to calculate

and plot *differences* between MC data performed at different  $N$  and compare to the corresponding differences for the theoretical finite-size corrections.

### III. Dielectric discontinuity with the external medium

According to the last term in (11), simulations performed at two different dielectric constants of the external medium,  $\varepsilon'$  and  $\varepsilon''$ , correspond to pair potentials which differ by the following  $1/N$  term:

$$v_{\varepsilon''}(12) - v_{\varepsilon'}(12) = \frac{4\pi}{L^3} \left( \frac{1}{2\varepsilon''+1} - \frac{1}{2\varepsilon'+1} \right) \bar{\mu}_1 \bar{\mu}_2 = -kT \frac{\lambda_{\varepsilon''\varepsilon'}}{N} \hat{\mu}_1 \hat{\mu}_2 \quad (16)$$

with  $\lambda_{\varepsilon''\varepsilon'} = 9\gamma((2\varepsilon''+1)^{-1} - (2\varepsilon'+1)^{-1})$  (DPS<sup>23</sup>, eq.1.5,1.8). This unusual extra potential is independent of the positions and depends on the orientations through a pure  $^{110}$  component. Starting from the diagrammatic expansion and using a perturbation theory for the correlation function, DPS have obtained the differences in the first three projections  $g^{mnl}$  between  $\varepsilon''$  and  $\varepsilon'$  boundary conditions to first order in  $1/N$  (note that this is richer than first order in  $\lambda_{\varepsilon''\varepsilon'}$  due to the  $r$ -independence of (16)). Their result may be written as (DPS, eq.2.23-25):

$$g_{\varepsilon''}^{mnl} - g_{\varepsilon'}^{mnl} = - \frac{\lambda_{\varepsilon''\varepsilon'} g(\varepsilon'') g(\varepsilon') / \sqrt{3}}{N} G^{mnl} \equiv - \frac{A_{\varepsilon''\varepsilon'}}{N} G^{mnl} \quad (17)$$

(which defines the factors  $A_{\varepsilon''\varepsilon'}$ ) with the DPS predictions for the functions  $G^{mnl}(r)$ :

$$\begin{aligned} G^{000} &= g^{110} \\ G^{110} &= g^{000} \\ G^{112} &= 0 \end{aligned} \quad (18)$$

In particular,  $G^{110}$  tends to 1 at large separation. When inserted in (15), this asymptote leads in the thermodynamic limit to the relation  $g(\varepsilon'') = g(\varepsilon') / (1 - \lambda_{\varepsilon''\varepsilon'} g(\varepsilon') / 3)$  between the  $g$ -factors and, via (14), to the important conclusion that the dielectric constant of the fluid  $\varepsilon$  is, fortunately, independent of the choice of  $\varepsilon'$  (DPS, eq.3.5).

Our test of the DPS theory consists to construct ratio  $\left[ g_{\varepsilon''}^{mnl}(r) - g_{\varepsilon'}^{mnl}(r) \right] / \left[ -A_{\varepsilon''\varepsilon'} / N \right]$  for different couples  $\varepsilon'', \varepsilon'$  in the range  $1-\infty$  and/or different  $N=100-800$  and to compare to the functions (18), see figures 1 and 2 for systems I and II, respectively. Note that such analysis can be performed only thanks to the sub- $10^{-3}$  precision of the *pdf*. The conclusion of our study is manifold: i) within the statistical noise quantified by the error bars in the figures, almost all cases seem to coincide on same master curves  $G^{mnl}(r)$ ; this indicates that the DPS factor  $A_{\varepsilon''\varepsilon'}$  alone fully accounts for the  $\varepsilon'$ -dependence of the present  $1/N$  finite-size corrections. As a side effect, this means that the *pdf* at a third  $\varepsilon'''$  can be easily derived from that measured at  $\varepsilon'$  and  $\varepsilon''$ . For system I, we have even performed a few Grand-Canonical MC simulations at  $N \approx 100$

with an imposed activity  $a^*=a\sigma^3=1.704$  carefully chosen to give a mean density  $\langle\rho^*\rangle$  very close to the target canonical value 0.8. As seen in figure 1, the corresponding *pdf* (curves labelled GC) match again the same master curves. For system II, the  $^{110}$  projection in figure 2 reveals a small but clear departure from that universality which increases with decreasing  $N$ : the curves  $(\varepsilon',\varepsilon)=(1,99)$ ,  $(20,99)$  and  $(\infty,99)$  are localized on top of each other in that order, the gap being larger for  $N=108$  than for  $N=300$  and roughly scaling with  $1/N$ . That reveals without doubt the existence of second order terms in the  $1/N$  expansion of  $g_{\varepsilon'}-g_{\varepsilon}$ . It is interesting to note, and quite intriguing, that the three curves are almost perfectly proportional: they may become indistinguishable by just artificially correcting the coefficients  $A_{\varepsilon'\varepsilon}$  alone. That seems to indicate that the second order correction is proportional to the same  $G^{110}$  function. An equivalent behavior may be observed for the  $^{000}$  and  $^{112}$  projections at  $N=108$ , albeit with less clarity. ii) the observed function  $G^{110}$  is very well consistent with an asymptote equal to 1, so the conclusion about the independence of  $\varepsilon$  on  $\varepsilon'$  in the bulk limit remains valid. iii) on the other hand, the short-range part of the function  $G$  clearly deviates from the DPS prediction (18) for the two systems and the three projections:  $G^{000}$  presents a low but measurable structure, which unambiguously emerges from the noisy data and is inconsistent with  $g^{110}$ ; the main peak in  $G^{110}$  is close to but different from that of  $g^{000}$ ; finally,  $G^{112}$  exhibits a rich behavior which is far from being negligible. Note that this kind of disagreement has already been noticed by Kusalik in his molecular dynamics study of System II, albeit with less quantitative precision <sup>21</sup>. We have tried to improve this gap between simulation data and theory. A careful reading of the original DPS paper shows that the *pdf* projections beyond the first three terms have been explicitly omitted in the *last* part of the analysis (see discussion after eq.2.8). It seems that the final DPS derivation of the nine diagrams (eq.2.14-2.22) can be repeated with the complete *pdf* expansion, leading to the new simple function  $G=\Phi^{110}g$  and its corresponding projections (noted DPS'):

$$\begin{aligned} G^{000} &= g^{110} \\ G^{110} &= g^{000} + \frac{2}{\sqrt{5}} g^{220} \\ G^{112} &= -2\sqrt{2/5} g^{022} + \frac{\sqrt{7}}{5} g^{222} \end{aligned} \quad (19)$$

Unfortunately, as seen in figure 1,2, the so modified theory does not reproduce the simulation data either. Something else seems to be lacking in the derivation of the DPS theory. Despite many tries exploring different directions, we were unable to understand the origin of this disagreement and the problem is still open to future theoretical works.

In the rest of the paper, in order to explore the next sources of corrections without confusion, we will minimize the polarization effect by choosing  $\varepsilon'$  as close as possible to  $\varepsilon$ . In that case, it is generally accepted that the bulk limit is attained and the present finite-size correction identically vanishes. In order to appreciate the quality of the matching, it is fruitful to note that the asymptote of  $g^{110}(r)$  in presence of mismatch is given by (14),(17),(18):

$$g_{\varepsilon'}^{110} (\text{asymptote}) \equiv -A_{\varepsilon', \varepsilon} / N = -\frac{2}{3\sqrt{3}y} \frac{(\varepsilon' - \varepsilon)(\varepsilon - 1)^2}{\varepsilon(2\varepsilon' + \varepsilon)} \frac{1}{N} \xrightarrow{\varepsilon' \approx \varepsilon \ll 1} -\frac{2}{9\sqrt{3}y} \frac{\varepsilon' - \varepsilon}{N} \quad (20)$$

If a simulation is performed at  $\varepsilon'$  different from  $\varepsilon$ , a rapid fixing consists to subtract that asymptote from  $g^{110}$ . A more complete treatment would require to subtract  $-(A_{\varepsilon', \varepsilon}/N)G$  (namely, the same asymptote times  $G$ ) from the measured  $g$ . Even if the master function  $G$  is not yet correctly known from a theory, it can be easily extracted with the help of a second simulation performed at any  $\varepsilon''$  (very) different from  $\varepsilon'$ .

Figure 3 presents the first relevant MC projection differences  $g^{mnl}_N(r) - g^{mnl}_{N=800}(r)$  of System I obtained at  $\varepsilon'=16$  for  $N=100$  and  $300$ . The very small dielectric discontinuity which could subsist ( $\varepsilon \approx 16.3 \pm 0.1$ <sup>12</sup>) leads to a negligible  $g^{110}$  asymptote (20) in the range  $2 \times 10^{-4}$  for  $N=100$ . The curves are in the  $10^{-2}$  and  $10^{-3}$  regimes for  $N=100$  and  $300$ , respectively. The statistical error bars attached to them, schematically represented at  $r=1, 2, 3$  (one standard deviation), remain below  $5 \times 10^{-4}$  and  $2 \times 10^{-4}$ , respectively (histogram width =  $\sigma/80$  and  $\sigma/20$ ). These uncertainties are negligible for the  $N=100$  case and for the first projection<sup>000</sup> of the  $N=300$  case while they become relevant for the higher order  $N=300$  projections beyond the first peak. In the same way, figure 4 presents the similar projection differences for System II obtained at  $\varepsilon'=99$  (close to the Kusalik's estimate of  $\varepsilon^{22}$ ) for  $N=108, 300$  and  $4000$  (we have preferred to choose the  $N=800$  set as reference instead of  $N=4000$  because of its better statistical uncertainty). A potential unity gap between  $\varepsilon'$  and  $\varepsilon$  would correspond in this case to a  $3.5 \times 10^{-4}$   $g^{110}$  asymptote for  $N=108$ . The observed differences lie in the regimes  $3 \times 10^{-2}, 3 \times 10^{-3}, 3 \times 10^{-4}$ , respectively. Error bars are given to appreciate the quality of future comparisons. Note that for both systems, the shape of the curves is not maintained as  $N$  varies, which means that some remaining finite-size corrections do not follow a  $1/N$  expansion.

#### IV. Explicit finite-size corrections

Lebowitz and Percus have expressed to first order in  $1/N$  the difference between the canonical *pdf*  $g_N$  and the grand-canonical reference  $g$  that results from the forbidden  $N$  fluctuations<sup>13</sup>:

$$g_N \approx g - \frac{\chi}{2N} \frac{\partial^2(\rho^2 g)}{\partial \rho^2} \quad (21)$$

The asymptote (1) represents the long-distance behavior of (21). Projection onto the rotational invariants leads to:

$$\begin{aligned}
g_N^{mnl}(r) &\approx g^{mnl}(r) + \frac{1}{N} E_{\text{expl.}}^{mnl}(r) \\
E_{\text{expl.}}^{mnl}(r) &= -\frac{\chi}{2} \frac{\partial^2 (\rho^2 g^{mnl}(r))}{\partial \rho^2}
\end{aligned} \tag{22}$$

The 2<sup>nd</sup> order density-derivatives of the *pdf* are estimated in practice by performing MC simulations at neighboring densities and using standard differentiation approximation, see figures 5, 6. For the present dense fluid at  $\rho^*=0.8$ , a good compromise between precision and accuracy is obtained by choosing  $\rho^*=0.75$  and  $0.85$  ( $\rho^*=0.78$  and  $0.82$  lead to the same result). Note that, since the function  $E_{\text{expl.}}$  appears in a  $1/N$  corrective term, its knowledge does not require a precision as demanding as for the bare *pdfs* and the error of any origin attached to its determination produces in practice a negligible effect. The  $E_{\text{expl.}}$  curves extracted at different  $N=100-800$  coincide at the scale 1. Figure 5 (System I) adds for comparison the differences between the canonical and grand-canonical *pdfs*,  $g_N-g$ , measured at  $N=100$  and three different  $\epsilon$  (1, 16 and  $\infty$ ). The very good agreement with the previous curves confirms that the explicit finite-size corrections are correctly evaluated and, as a side information, that there is no detectable 2<sup>nd</sup> order coupling between the correction coming from a dielectric discontinuity and that from an  $N$  fluctuation inhibition, at least at  $N>100$ . Figures 5-6 exhibit a rich and complex behavior at short distance which adds to the well-known asymptote (1) in the 000 projection.

Multiplying the  $E_{\text{expl.}}(r)$  projections of figures 5, 6 by the scaling factor  $(1/N-1/800)$  and comparing to the projection differences  $g^{mnl}_N(r)-g^{mnl}_{N=800}(r)$  of figure 3, 4 clearly reveals that the present explicit contribution is insufficient to fully account for the measured finite-size corrections, both in shape and in height. While for System I (figures 3 vs 5), it appears as being an important part of the total correction, it becomes almost negligible for System II (figures 4 vs 6). This latter behavior is specific to strongly coupled, dense systems characterized by low compressibility (so small explicit corrections) and long range correlations (so large implicit corrections).

For the following, we subtract the explicit contribution from the *pdf* differences and will compare the resulting  $g^{mnl}_N-g^{mnl}_{N=800}-(1/N-1/800)E_{\text{expl.}}^{mnl}$ , plotted in figures 7, 8, to the last potential source of finite-size corrections, the *implicit* ones.

## V. Implicit finite-size corrections

We consider a particle noted 1 at the center of the central box with absolute orientation  $\Omega_1$ . Within the periodic boundary conditions, an infinity of images 1' are located at the center of the replica, at positions  $\mathbf{r}_{11'}=\mathbf{p}L$  (vector  $\mathbf{p}$  of integer components and norm  $p$ ), with the *very same* orientation  $\Omega_{1'}=\Omega_1$ , see figure 9. The problem is to express the local density of dipoles 2

at  $\mathbf{r}_2$  with orientation  $\Omega_2$  inside the central box in terms of the pair correlation with central dipole 1 and the perturbations coming from the 1' images. In a first step, we will assume that the somewhat intuitive superposition approximation (2) holds. For short-range potentials, it has been shown how this approximation could result from the neglect of a class of diagrams linking the different particles and images<sup>14</sup>. Without any further theoretical justification, we will generalize its use to the present case of long-range dipolar interaction. The *pdf*  $g_N$  measured during a simulation with  $N$  particles (box size  $L$ ) is thus related to the bulk reference  $g$  through:

$$\begin{aligned} g_N^{\text{superp.}}(12) &= g(12) \prod_{1' \text{ images of } 1} g(1'2) \\ &= g(12) \prod_{1' \text{ images of } 1} (1 + h(1'2)) \\ &\approx g(12) \left[ 1 + \sum_{1'} h(1'2) \right] \end{aligned} \quad (23)$$

In the third line of (23), we have benefited from the fact that the perturbations  $h=g-1$  due to the different images 1' are small and we have kept the first order in  $h$  only (addition of 3-body superpositions)<sup>15</sup>.

According to (3), the function  $h(1'2)$  which appears in the *rhs* of (23) is known as an expansion onto rotational invariants specific to the pair 1'2. The difficulty is to convert it in a similar expansion, this time specific to the central pair 12. As mentioned earlier, this is totally meaningful only by spherically averaging over the absolute orientation of the ensemble  $\mu_1, \mu_2, \mathbf{r}_{12}$  with respect to the axes of the box as done during the construction (13) of the *pdf* or, equivalently, by averaging over the orientation of the vector  $\mathbf{p}$  pointing to the image 1' (analysis valid only at  $r < L/2$ ). That operation, noted  $\langle \rangle$  in the following equations, will greatly simplify the mathematics. Working in the laboratory frame, the expansion (3)(4) gives:

$$\langle h(1'2) \rangle = \sum_{mnl} \sqrt{(2m+1)(2n+1)} \sum_{\mu', \nu', \lambda'} \begin{pmatrix} m & n & l \\ \mu' & \nu' & \lambda' \end{pmatrix} R_{\mu'0}^m(\Omega_1) R_{\nu'0}^n(\Omega_2) \langle h^{mnl}(r_{1'2}) R_{\lambda'0}^l(\hat{r}_{1'2}) \rangle \quad (24)$$

The average over the orientation of  $\mathbf{p}$  is better done in the molecular frame attached to the 1, 2 particles (principal axis along  $\mathbf{r}_{12}$ ). For that purpose, we use the standard rotation formula between spherical harmonics<sup>20</sup>:

$$R_{\lambda'0}^l(\hat{r}_{1'2}) = \sum_{\lambda''} R_{\lambda''0}^l(\hat{r}_{1'2} / \hat{r}_{12}) R_{\lambda'\lambda''}^l(\hat{r}_{12}) \quad (25)$$

If  $\mathbf{p}$  is defined by the polar angles  $(\theta, \varphi)$  in this local frame (see figure 9), averaging over the longitude  $\varphi$  gives 0 except for the first term  $\lambda''=0$  in the sum (25) where it gives 1. Finally, averaging over the colatitude  $\theta$  leads to:

$$\langle h^{mnl}(r_{1'2})R_{\lambda'0}^l(\hat{r}_{1'2}) \rangle = \langle h^{mnl}(r')P_l(\cos\theta') \rangle_{\theta} R_{\lambda'0}^l(\hat{r}_{1'2}) \quad (26)$$

where the distance  $r' \equiv r_{1'2}$  and the angle  $\theta'$  between  $\mathbf{r}_{12}$  and  $\mathbf{r}'_{12}$  are obvious functions of  $r$ ,  $pL$  and  $z = \cos\theta$  in the triangle 11'2 (figure 9):

$$\begin{aligned} \langle h^{mnl}(r')P_l(\cos\theta') \rangle_{\theta} &= \frac{1}{2} \int_{-1}^1 h^{mnl} \left( \sqrt{r^2 - 2rpLz + (pL)^2} \right) P_l \left( \frac{r - pLz}{\sqrt{r^2 - 2rpLz + (pL)^2}} \right) dz \\ &\equiv I^{mnl}(r, pL) \end{aligned} \quad (27)$$

$P_l = R_{00}^l$  is the Legendre polynomial of order  $l$ . When (26) is inserted in (24), a simple expansion of  $\langle h(1'2) \rangle$  onto 12 invariants is recovered as desired:

$$\langle h(1'2) \rangle = \sum_{mnl} I^{mnl}(r_{12}, pL) \Phi^{mnl}(12) \quad (28)$$

One notes that the  $^{mnl}$  projection of the perturbation  $\langle h(1'2) \rangle$  depends on the projection  $h^{mnl}$  *only*. That simple, unexpected behavior is again due to the orientation spherical averaging operation. For the present dipolar case, it is also interesting and somewhat reassuring to verify that the well-known  $1/r^3$  asymptote of the  $h^{112}(r)$  projection

$$h^{112}(r) \underset{r \text{ large}}{\approx} \frac{\sqrt{10/3} (\varepsilon - 1)^2}{4\pi\rho y} \frac{1}{\varepsilon r^3} \quad (29)$$

leads to a perfectly zero contribution to  $I^{112}$  in (27). That guarantees the convergence of the sum over the infinity of images. To be complete, one could also mention that the data extracted beyond  $r=L/2$ , in the corners of the box, would not result from a complete orientation averaging and would exhibit a spurious behavior due to this asymptote, as nicely shown by Caillol using a more rigorous route<sup>26</sup>.

Finally, the last step consists to derive the projections of the product  $g(12)\langle h(1'2) \rangle$  in (23). Starting from the known reduction relation between rotational invariants,  $\Phi_{\alpha_1} \Phi_{\alpha_2} = \sum_{\alpha} t_{\alpha_1\alpha_2\alpha} \Phi_{\alpha}$ , with well-documented coefficients  $t^{20}$ , the desired expression reads:

$$\mathcal{g}_N^{\text{superp.}}{}_{\alpha}(r) - \mathcal{g}_{\alpha}(r) = \sum_{\alpha_1\alpha_2} t_{\alpha_1\alpha_2\alpha} \mathcal{g}_{\alpha_1}(r) \sum_{\bar{p}} I_{\alpha_2}(r, pL) \quad (30)$$

In conclusion, provided that the *pdf* is known at long distances (the minimum distance for  $r'$  is  $L/2$ ), a simple one-dimensional integral (27) of its projections, performed at distances  $r < L/2$  for each neighboring image  $\mathbf{p}$ , is sufficient to calculate the implicit finite-size corrections within the superposition approximation. By definition, this long-range behavior of  $g$  cannot be given by the simulation under study itself. It should not be derived from the bare HNC integral equation either because the HNC solution, certainly of reasonable quality at short distances, becomes incorrect, even qualitatively, at large distances. On the other hand, it can advantageously be borrowed from the powerful MC/HNC or MC/MSA approach of Paper I. It is sufficient to know here that this consists to solve the mixed integral equation:

$$\begin{aligned}
g^{mnl}(r) &= g^{mnl}(r)(\text{MC}) && ; r \leq r_{\max} \\
b^{mnl}(r) &= 0(\text{HNC}) \quad \text{or} \quad c^{mnl}(r) = -\beta v^{mnl}(r)(\text{MSA}) && ; r > r_{\max}
\end{aligned} \tag{31}$$

$c$  is the direct correlation function (related to  $g$  via the Ornstein-Zernike equation),  $b$  is the bridge function<sup>3</sup>. The iterative resolution makes intensive use of powerful techniques developed for the treatment of anisotropic interactions. As mentioned in the Introduction, it has been shown how HNC and MSA approximate closures can safely be applied beyond  $r_{\max}$  cut-offs as low as two or three diameters<sup>12</sup> (with a clear advantage to MSA in the former case). It is important to realize that equations (30), (31) are in principle and in practice *coupled*: the evaluation of the implicit finite-size corrections requires the knowledge of the MC/HNC,MSA solution at large separation which, itself, depends on short-range MC data corrupted by these corrections! That coupling simply illustrates how, when the range of the correlations is larger than the size of the simulation box, the missing long range part of the *pdf* beyond the cell edge is somewhat "folded" and recovered inside the box via the far-field environment around the images! For a first clear analysis, we have fed (31) with MC data obtained with a large enough  $N$  number ( $N=800$  or  $4000$ ,  $r_{\max}=3-5\sigma$ ) chosen such that the (very) small implicit error attached to them has a negligible effect on the evaluation (30) of the corrections for lower  $N$ . If, on the other hand, only small  $N$  ( $\approx 100$ ) MC data are available, an iterative procedure would be required which improves step by step the cleaning of the *pdf* from the implicit corrections, see Section VI as an illuminating exercise.

In order to test the present analysis, we compare the differences  $g_N - g_{800} - (1/N - 1/800)E_{\text{expl}}$  between MC *pdfs* (which are cleaned from explicit corrections, see Section IV) to the theoretical predictions for the differences in implicit corrections between  $N$  and  $N=800$ , see the solid curves in figure 7, 8 (for system I, the corrections for  $N=800$  are negligible at the scale of figure 7). The sum over the images  $\mathbf{p}$  in (30) involves mainly the first shell (6 neighbors with  $p=1$ ) with a small but detectable contribution of the next neighbors (12 and 8 at  $p=\sqrt{2}$  and  $\sqrt{3}$ , respectively) for the smallest  $N$  values. The integrals  $I^{mnl}$  in (27) are calculated using Gauss-Legendre quadrature with 50 to 100  $\theta$  values. The numerical extra-work required in the evaluation of (27), (30) at the end of the integral equation resolution is marginal. For the two dipolar fluids and for the various investigated  $N$  values, the quantitative agreement within the statistical uncertainty is very good for all projections and at all distances  $r < L/2$ , *except* near contact. As expected, the corrections increase as the cell edge and thus the first neighbors are approached. The <sup>000</sup> term is very similar to that observed for pure spherical interactions and presents oscillations at large distances which are the signature of the packing inside the dense fluid; the <sup>110</sup> contribution comes from orientation correlations  $\mu_1 \mu_2$  between dipole 2 and dipole images 1' which again replace (in a perfect manner within the superposition approximation) the missing correlations between the central dipole 1 and the dipoles located beyond the frontiers of the box. That interesting effect explains in particular why the  $g$ -factor (15) which integrates the function  $g^{110}$  inside the cell *only* leads to correct values of the dielectric constant  $\varepsilon$  when inserted in the Kirkwood equation (14) *even* when the size  $L$  of the box is (much) smaller than the range of the orientation fluctuations! Lastly, the <sup>112</sup> correction is mainly due to the weakly damped tail which adds to the  $g^{112}(r) 1/r^3$



asymptote. Note that the long-range behaviors of  $g^{110}$  and of this corrective  $g^{112}$  tail are directly related through the Ornstein-Zernike equation in the low  $q$  limit.

It remains to understand the disagreement observed at the first peak, near contact. The same kind of deviation is observed for the pure LJ fluid. One may question the validity of the superposition approximation which is known to be insufficient at short 1-2 separation even when the image 1' is far from the pair. Following the same phenomenological approach and without any theoretical basis, we will consider that the simulation measures in fact  $n$ -body distribution functions which, as in (23), can be expressed as the addition of 3-body terms. The following development focuses on the first correction to the superposition approximation which involves a well-known 4-particle diagram  $\tau$ .

The distribution of dipoles at location 2 due to the central particle 1 and a single image 1' is now written as  $g(12)g(1'2)\exp[\tau(11'2)] \approx g(12)(1+h(1'2)+\tau(11'2))$ . The implicit correction is thus supplemented with:

$$g_N - g_N^{\text{superp.}} \approx g \sum_{1'} \tau(11'2) \quad (32)$$

The diagram  $\tau$  involves integration over the position and orientation of a fourth dipole<sup>3</sup>:

$$\tau(11'2) = \rho \int h(13)h(23)h(1'3)d3 \quad (33)$$

Once again, the operation of orientation averaging over the vector  $\mathbf{p}$  joining the images 1, 1' will greatly simplify the calculation. That concerns only the term  $h(1'3)$  in (33) and we just have shown how  $\langle h(1'3) \rangle$  can be explicitly expanded on the basis  $\Phi(13)$  with projections  $I^{mnl}(r_{13}, pL)$ , see (28). The diagram  $\tau$  thus becomes:

$$\tau(11'2) = \rho \int [hI](13)h(32)d3 \quad (34)$$

That is nothing but a simple convolution product, both in position and orientation. Such kind of product is routinely calculated during the resolution of anisotropic Ornstein-Zernike equation and it is sufficient to use the very same procedure which goes through the Fourier space: the projections of  $h$  and  $I$  (summation over the relevant images  $\mathbf{p}$  is performed at this early stage) are combined as in (30) to give the expansion of  $[hI]$ . The Hankel transform (of order  $l$ ) is performed for  $[hI]^{mnl}$ . The algebraic product with the Hankel transforms of  $h$  is then better expressed in the molecular frame ( $\hat{q}$  taken as the reference axis) by using the so-called  $\chi$ -transforms of Blum<sup>19</sup>. Lastly, inverse Hankel transforms lead back to the desired  $\tau^{mnl}$  projections and, using again the reduction relation, to the first correction to the superposition approximation (32).

The new predictions for the implicit finite-size corrections, sum of superposition (23)-(30) and first correction (32)-(34), are added as dashed lines to figures 7, 8. The improvement appears clearly for both systems. For System I, the agreement with the MC data is now visually perfect at the sub  $10^{-3}$  level for  $N=100$  even at the first peak (figure 7 top) and stays within the error bars for  $N=300$  (figure 7 bottom). For the more coupled System II, the new

extra term brings detectable improvement for  $N=100$  and  $300$  at the first peak as well at intermediate separations  $r$ . Nevertheless, a careful analysis of figure 8 reveals that a few discrepancies between theoretical predictions and MC data differences still exceed the statistical uncertainty (110 projection for  $N=108$ , top figure, 112 projection for  $N=300$ , middle figure). That reveals an incomplete evaluation of the implicit corrections at the  $10^{-3}$ - $10^{-4}$  level. For such highly coupled system, the existence of long range correlations pushes the present theory up to its limit of validity. For instance, without needing to invoke the existence of next order corrections to the superposition approximation, one may first simply question the linearization operation  $\prod_{l' \text{ images of } l} (1+h(l'2)) \approx 1 + \sum_{l'} h(l'2)$  in (23) which could result in non-negligible effects.

That observation closes the whole analysis. Since all sources of finite-size corrections have been identified and precisely quantified, we are now able in practice to correct any set of MC data from such corrections, at least at reasonable  $N$  values for which  $1/N$  expansions cut to first order and superposition approximations with first  $\tau$  diagram apply.

## VI. Illuminating application and Conclusion

In practice, what is the interest of understanding and evaluating with great precision finite-size corrections in numerical simulation? To be able to correct cheap simulation data, accumulated at small  $N$  with limited CPU resources, and derive anyway spatial correlations on the whole  $r$  and  $q$  ranges. In the present example, we deal with the highly coupled dipolar System II and assume that MC data are available for  $N=108$  *only*. Despite the smallness of the simulation box, the existence of long range correlations beyond the cutoff  $r_{max}=L/2 \approx 2.5\sigma$  and the approximated character of the MC data below it, the goal is to extract all correlations at all distances, free of corrections, in particular those concerning the dipole alignment characterized by the 110 projection. Rather to follow the function  $g_{00}^{110}(r)$  or the dielectric constant  $\varepsilon$ , it is more illuminating to plot the well-known running Kirkwood g-factor  $g_K(R)$  which accounts for the orientation correlation between dipoles separated by *less* than the distance  $R$ :

$$g_K(R) = \frac{1}{N\mu^2} \left\langle \sum_{\substack{i,j=1 \\ r_{ij} < R}}^N \bar{\mu}_i \bar{\mu}_j \right\rangle \equiv 1 - \frac{\rho}{\sqrt{3}} \int_0^R g^{110}(r) d\bar{r} \quad (35)$$

At large distances, when  $R$  reaches the corner of the cell in MC simulation, or when  $R \rightarrow \infty$  in MC/MSA mixed integral equation, one recovers the full factor (15) from which is derived the dielectric constant through (14). It is known that the shape and fluctuations of the  $g_K(R)$  curve are very sensitive to the choice of the Ewald boundary condition, to the large statistical

uncertainties which affect the long range behavior of  $g_{00}^{110}(r)$  and to the quality of the MC data. It is thus a quantity of choice to test the present theory. We start from MC simulation data performed with  $N=108$  dipoles ( $L \approx 5.13 \text{ \AA}$ ) and  $\varepsilon'=99$  (the strict reader may consider that choosing  $\varepsilon'$  close to the final value of  $\varepsilon$  before this one has been determined with precision is somewhat cheating; it is possible to start with a guessed, non-optimal value for  $\varepsilon'$ , provided that one corrects afterwards the effect of dielectric discontinuity with the external medium according to the procedure described in Section III; in the following, we will forget about this effect and focus on the most difficult corrections to deal with, the implicit ones). Figure 10 plots the resulting  $g_K(R)$  curve up to the corner  $R=L\sqrt{3}/2$  (black symbols). We then clean these bare data from explicit finite-size corrections by subtracting  $E_{\text{expl}}/108$  of Section IV where the density derivatives in  $E_{\text{expl}}$  are estimated from MC simulations performed again with  $N=108$ . On the other hand, we cannot for the moment clean from implicit corrections since no estimation is available yet. The MC/MSA integral equation is then solved from these MC data kept up to  $r_{\text{max}}=2.5\sigma \approx L/2$ . The resulting  $g_K(R)$  curve is added in figure 10 (dotted curve). While one observes an almost perfect agreement with the original bare MC curve up to  $r_{\text{max}}$ , which confirms that the explicit corrections are somewhat negligible in the present incompressible fluid, the new MC/MSA solution exhibits a large, clearly too large, increase beyond that cutoff and reaches the asymptote  $g_K(\infty)=11.1$  or the dielectric constant  $\varepsilon=167!$  This incorrect behavior originates from the distortion of the MC brute data by the implicit corrections: the degree of alignment of the dipoles is overestimated ( $g_{00}^{110}(r)$  too negative) at  $r < L/2$  due to the contribution coming from the neighboring images (such contribution is thus counted *twice!*); worse, by continuity and extrapolation through the Ornstein-Zernike and mixed integral equation (31), this trend is amplified in the long-distance regime  $r > r_{\text{max}}$ . So, there is a clear need for beforehand cleaning from these implicit contributions. What happens if one evaluates them from the first MC/MSA solution using the analysis of Section V (superposition+ $\tau$  correction)? One solves again the MC/MSA theory, fed with the simulation data corrected this time from explicit *plus* implicit terms. The new  $g_K(R)$  in figure 10 (dashed curve) now presents too low values and too low  $\varepsilon=33!$  The explanation is that the implicit term, estimated from a too strong dipole alignment, has been overestimated and the data have been overcorrected. The conclusion is clear: in order to get a self-consistent analysis of the MC/MSA theory and the implicit corrections, one must iterate the procedure with some damping for a rapid convergence. In practice, at each step, rather to use a new estimation of the correction, one mixes it with the previous version in equal parts. After half a dozen iterations, self-consistency is obtained: the final MC/MSA solution produces finite-size implicit corrections (superposition+ $\tau$  correction of Section V) which equal exactly those used to first correct the MC raw data. The corresponding  $g_K(R)$  is given in figure 10 (solid curve). The asymptote produces the dielectric constant  $\varepsilon=98.2$ , very close to the "correct" value  $99 \pm 1$  of Table II. The brute MC data obtained with the largest cell,  $N=4000$ , which serve as reference, are added for comparison. The agreement is spectacular, both in the whole  $R$ -behavior and in the asymptote or  $\varepsilon$  value. Note that the errors associated with both curves are somewhat different: For the  $N=4000$  MC data, the uncertainty is statistic, especially in the long range behavior of  $g_{00}^{110}(r)$ . On the other hand, the MC/MSA starts with simulation data

measured at short distances,  $r < 2.5\sigma$ , less sensitive to noise. But they are based at the same time on systematic errors related to the use of the MSA approximated closure beyond  $r_{max}$  and to the less than perfect evaluation of the implicit corrections (see top figure 8). That being said, one is able to conclude from the overall agreement shown in figure 8 and figure 10 that long-range correlations, low- $q$  behavior, dielectric properties, Kirkwood-Buff integrals, etc., can be derived in highly coupled systems of non-spherical particles from simulation boxes as small as  $L=5\sigma$  or  $N=100$  thanks to the correct understanding and quantitative evaluation of all possible sources of finite-size corrections.

## References

1. R. Wedberg, J. P. O'Connell, G. H. Peters, and J. Abildskov, *Fluid Phase Equilib.* **302**, 32 (2011)
2. J. G. Kirkwood and F. P. Buff, *J. Chem. Phys.* **19**, 774 (1951)
3. J. P. Hansen and I. R. McDonald. *Theory of Simple Liquids*; Academic: London, 1986.
4. R. Ramirez, M. Mareschal, and D. Borgis, *Chem. Phys.* **319**, 261 (2005)
5. L. Verlet, *Phys. Rev.* **163**, 201 (1968)
6. S. M. Foiles and N. W. Ashcroft, *J. Chem. Phys.* **81**, 6140 (1984)
7. D. L. Cheung, L. Anton, M. P. Allen, and A. J. Masters, *Phys. Rev. E* **73**, 61204 (2006)
8. D. L. Cheung, L. Anton, M. P. Allen, A. J.P.J. Masters, and M. Schmidt, *Phys. Rev. E* **78**, 41201 (2008)
9. F. Lado, *Mol. Phys.* **47**, 283 (1982)
10. P. H. Fries and G. N. Patey, *J. Chem. Phys.* **82**, 429 (1985)
11. L. Belloni and I. Chikina, *Mol. Phys.* **112**, 1246 (2014)
12. J. Puibasset and L. Belloni, *J. Chem. Phys.* **136**, 154503 (2012)
13. J. L. Lebowitz and J. K. Percus, *Phys. Rev.* **124**, 1673 (1961)
14. L. R. Pratt and S. W. Haan, *J. Chem. Phys.* **74**, 1864 (1981)
15. A. R. Denton and P. A. Egelstaff, *Z. Phys. B* **103**, 343 (1997)
16. L. R. Pratt and S. W. Haan, *J. Chem. Phys.* **74**, 1873 (1981)
17. J. J. Salacuse, A. R. Denton, P. A. Egelstaff, M. Tau, and L. Reatto, *Phys. Rev. E* **53**, 2390 (1996)
18. L. Blum and A. J. Torruella, *J. Chem. Phys.* **56**, 303 (1972)
19. L. Blum, *J. Chem. phys.* **57**, 1862 (1972)

20. A. Messiah. *Quantum Mechanics*; Wiley: New York, 1962; Vol. II.
21. P. G. Kusalik, *J. Chem. Phys.* **93**, 3520 (1990)
22. P. G. Kusalik, *Mol. Phys.* **73**, 1349 (1991)
23. S. W. de Leeuw, J. W. Perram, and E. R. Smith, *Proc. R. Soc. Lond. A* **373**, 57 (1980)
24. M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
25. M. Neumann and O. Steinhauser, *Chem. Phys. Lett.* **95**, 417 (1983)
26. J. M. Caillol, *J. Chem. Phys.* **96**, 7039 (1992)
27. H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989)

**Table 1. Monte Carlo data for the System I (Stockmayer fluid)**

$N$	$\epsilon'$	$N_{MC}(10^6)$	$-U/NkT$	$P/\rho kT$	$\chi=\partial\rho/\partial\beta P$	$\epsilon$
100	1	13	3.7790 <sub>2</sub>	2.513 <sub>1</sub>		16.75 <sub>7</sub>
	16	100	3.7946 <sub>1</sub>	2.4956 <sub>3</sub>		16.27 <sub>2</sub>
	$\infty$	45	3.8040 <sub>2</sub>	2.4846 <sub>5</sub>		16.07 <sub>2</sub>
300	1	16	3.7876 <sub>1</sub>	2.5352 <sub>5</sub>		16.44 <sub>7</sub>
	16	45	3.7926 <sub>1</sub>	2.5299 <sub>3</sub>		16.33 <sub>2</sub>
	$\infty$	16	3.7959 <sub>1</sub>	2.5263 <sub>5</sub>		16.26 <sub>3</sub>
800	1	2.4	3.7907 <sub>2</sub>	2.5374 <sub>7</sub>		16.3 <sub>2</sub>
	16	21	3.7928 <sub>1</sub>	2.5347 <sub>5</sub>	0.0791 <sub>5</sub>	16.36 <sub>3</sub>
	$\infty$	6	3.7939 <sub>1</sub>	2.5323 <sub>5</sub>		16.34 <sub>5</sub>
GC $L=5\sigma$	1	4	3.7781 <sub>4</sub>	2.5300 <sub>2</sub>	0.0786 <sub>5</sub>	16.6 <sub>1</sub>
$a^*=1.704$	16	10	3.7957 <sub>3</sub>	2.521 <sub>1</sub>	0.0782 <sub>3</sub>	16.17 <sub>7</sub>
$\langle N \rangle \approx 100$	$\infty$	4	3.8068 <sub>6</sub>	2.517 <sub>2</sub>	0.0770 <sub>5</sub>	16.14 <sub>7</sub>

Stockmayer fluid eq. (8), (9):  $\rho^*=0.8$ ,  $T^*=2$ ,  $\mu^2=1.5$ . Number of particles  $N$ , number of MC cycles  $N_{MC}$ , dielectric constants of the external medium  $\epsilon'$  and of the dipolar fluid  $\epsilon$ , energy  $U$ , pressure  $P$ , compressibility  $\chi$ . The last three lines correspond to Grand Canonical simulations with imposed  $a^*$  activity. For each accumulated average, the end subscript represents one standard deviation on the last digit.

**Table 2. Monte Carlo data for the System II (Soft sphere dipolar fluid)**

$N$	$\varepsilon'$	$N_{MC}(10^6)$	$-U/NkT$	$P/\rho kT$	$\chi=\partial\rho/\partial\beta P$	$\varepsilon$
108	1	26	0.4745 <sub>2</sub>	12.4376 <sub>9</sub>	0.02136 <sub>2</sub>	119 <sub>4</sub>
	20	24	0.4915 <sub>2</sub>	12.4130 <sub>9</sub>	0.02137 <sub>2</sub>	105.6 <sub>8</sub>
	99	75	0.5157 <sub>1</sub>	12.3790 <sub>5</sub>	0.02149 <sub>1</sub>	95.1 <sub>2</sub>
	$\infty$	26	0.5330 <sub>2</sub>	12.3556 <sub>9</sub>	0.02163 <sub>2</sub>	88.7 <sub>3</sub>
300	1	19	0.4858 <sub>2</sub>	12.4808 <sub>7</sub>	0.02112 <sub>1</sub>	105 <sub>4</sub>
	20	22	0.4917 <sub>2</sub>	12.4723 <sub>7</sub>	0.02117 <sub>1</sub>	103.8 <sub>8</sub>
	99	14	0.5001 <sub>2</sub>	12.4605 <sub>8</sub>	0.02123 <sub>1</sub>	98.1 <sub>6</sub>
	$\infty$	42	0.5075 <sub>1</sub>	12.4500 <sub>5</sub>	0.02124 <sub>1</sub>	95.6 <sub>3</sub>
800	1	12	0.4940 <sub>1</sub>	12.4727 <sub>5</sub>	0.02116 <sub>1</sub>	107 <sub>5</sub>
	20	6.7	0.4961 <sub>2</sub>	12.4689 <sub>7</sub>	0.02114 <sub>2</sub>	99.4 <sub>12</sub>
	99	20	0.4995 <sub>1</sub>	12.4648 <sub>4</sub>	0.02116 <sub>1</sub>	99.7 <sub>5</sub>
	$\infty$	18	0.5020 <sub>1</sub>	12.4617 <sub>4</sub>	0.02123 <sub>1</sub>	98.8 <sub>4</sub>
4000	99	6.6	0.4990 <sub>1</sub>	12.4671 <sub>3</sub>	0.02118 <sub>4</sub>	99.0 <sub>10</sub>

Soft-sphere dipolar fluid eq. (8), (10):  $\rho^*=0.8$ ,  $T^*=1.35$ ,  $\mu^{*2}=2.963$ . Same legend as Table 1.



## Figure caption

**Figure 1:** System I. Dependence on the external dielectric constant  $\varepsilon'$ . First three ratio  $G^{mnl}(r) = [g_{\varepsilon''}^{mnl}(r) - g_{\varepsilon'}^{mnl}(r)] / [-A_{\varepsilon''\varepsilon'} / N]$  (see eq.(17)) for different couples  $\varepsilon''/\varepsilon' = 1/16$  or  $\infty/16$  and different  $N=100, 300, 800$ , measured by MC. The curves labelled "GC" correspond to Grand-Canonical simulations with  $\langle N \rangle \approx 100$ . The horizontal axis expresses  $r$  in diameter  $\sigma$  units. The arrows locate  $L/2$  for the three simulation boxes. Vertical error bars around  $r=1, 2, 3, 5$  represent one standard deviation for each  $N$ . The curves labelled DPS and DPS' correspond to the original theoretical prediction (18)<sup>23</sup> and to the extended one (19), respectively.

**Figure 2:** System II. Dependence on the external dielectric constant. Same legend as figure 1 with  $\varepsilon''/\varepsilon' = 1/99, 20/99, \infty/99$  and  $N=108, 300, 800$ . The  $G^{110}(r)$  curves have been extended up to the corner of the boxes,  $r=L\sqrt{3}/2$ .

**Figure 3:** System I. Differences  $g^{mnl}_N(r) - g^{mnl}_{800}(r)$  between MC data for  $N=100$  (top) and  $N=300$  (bottom). External dielectric constant  $\varepsilon'=16$ , close to the bulk value. Statistical error bars are indicated at  $r=1, 2, 3$ . These differences should be attributed to the explicit + implicit finite-size corrections.

**Figure 4:** System II. Differences  $g^{mnl}_N(r) - g^{mnl}_{800}(r)$  between MC data for  $N=108$  (top),  $N=300$  (middle) and  $N=4000$  (bottom). External dielectric constant  $\varepsilon'=99$ , close to the bulk value. Statistical error bars are indicated at  $r=1, 2, 3, 4, 5$ .

**Figure 5:** System I. Explicit finite-size correction projections  $E_{\text{expl}}^{mnl}(r)$  (divided by  $N=100$  for comparison with grand-canonical simulations). The curves have been derived through (22) from 2<sup>nd</sup> order *pdf* derivatives evaluated with finite differentiation of MC data. The symbols represent the differences between canonical and grand-canonical MC simulations at  $N=100$  and three external dielectric constants  $\varepsilon' \equiv \varepsilon' = 16$  (full),  $\infty$  (open) and 1 (hatched).

**Figure 6:** System II. Explicit finite-size correction projections  $E_{\text{expl}}^{mnl}(r)$ . Same legend as figure 5.

**Figure 7:** System I. Symbols: Differences  $g^{mnl}_N - g^{mnl}_{800} - (1/N - 1/800)E_{\text{expl}}^{mnl}$  cleaned from the explicit corrections (22), for  $N=100$  (top) and  $N=300$  (bottom). Statistical error bars shown at  $r=1, 2, 3$ . Curves: Theoretical implicit finite-size corrections derived from the superposition approximation (23)-(30) (solid lines) or improved with the first correction in  $\tau$  (32)-(34).

**Figure 8:** System II. Symbols: Differences  $g^{mnl}_N - g^{mnl}_{800} - (1/N - 1/800)E_{\text{expl}}^{mnl}$  cleaned from the explicit corrections (22), for  $N=108$  (top),  $N=300$  (middle) and  $N=4000$  (bottom). Same legend as figure 7.

Figure 9: Schematic view of the implicit finite-size corrections. The pair correlations between particles 1 and 2 inside the central cell are perturbed by the environment around the periodic image 1' of 1, with the very same orientation, localized at the distance  $L$ .

Figure 10: Running Kirkwood factor  $g_k(R)$  (35) for System II. Symbols: MC brute data for  $N=108$  and  $N=4000$  with  $\epsilon=99$ . Lines: MC/MSA results fed with  $N=108$  MC data up to  $r_{max}=2.5\sigma$  beforehand cleaned without or with different choices for implicit corrections, see text in Section VI.

Figure 1

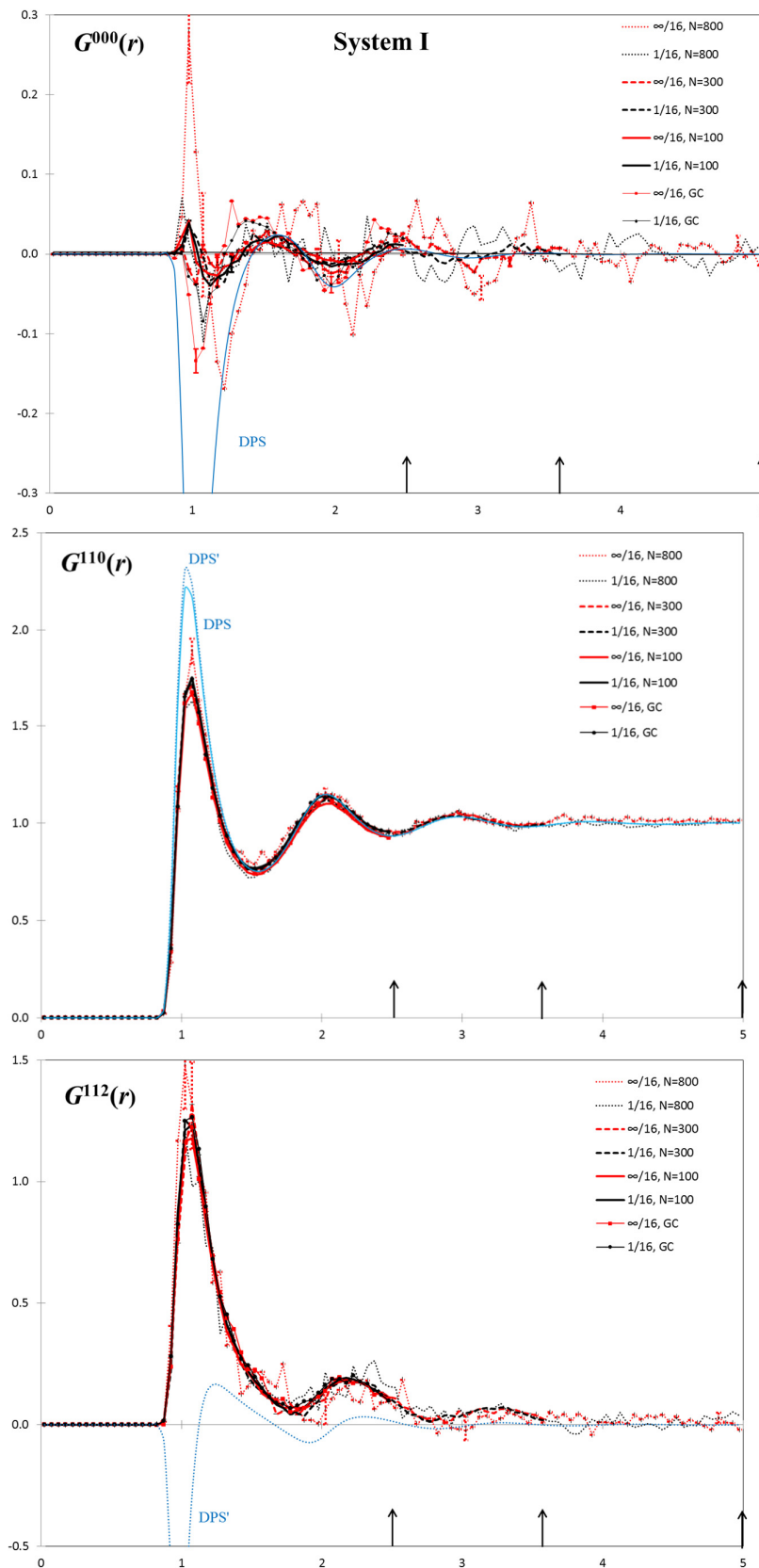


Figure 2

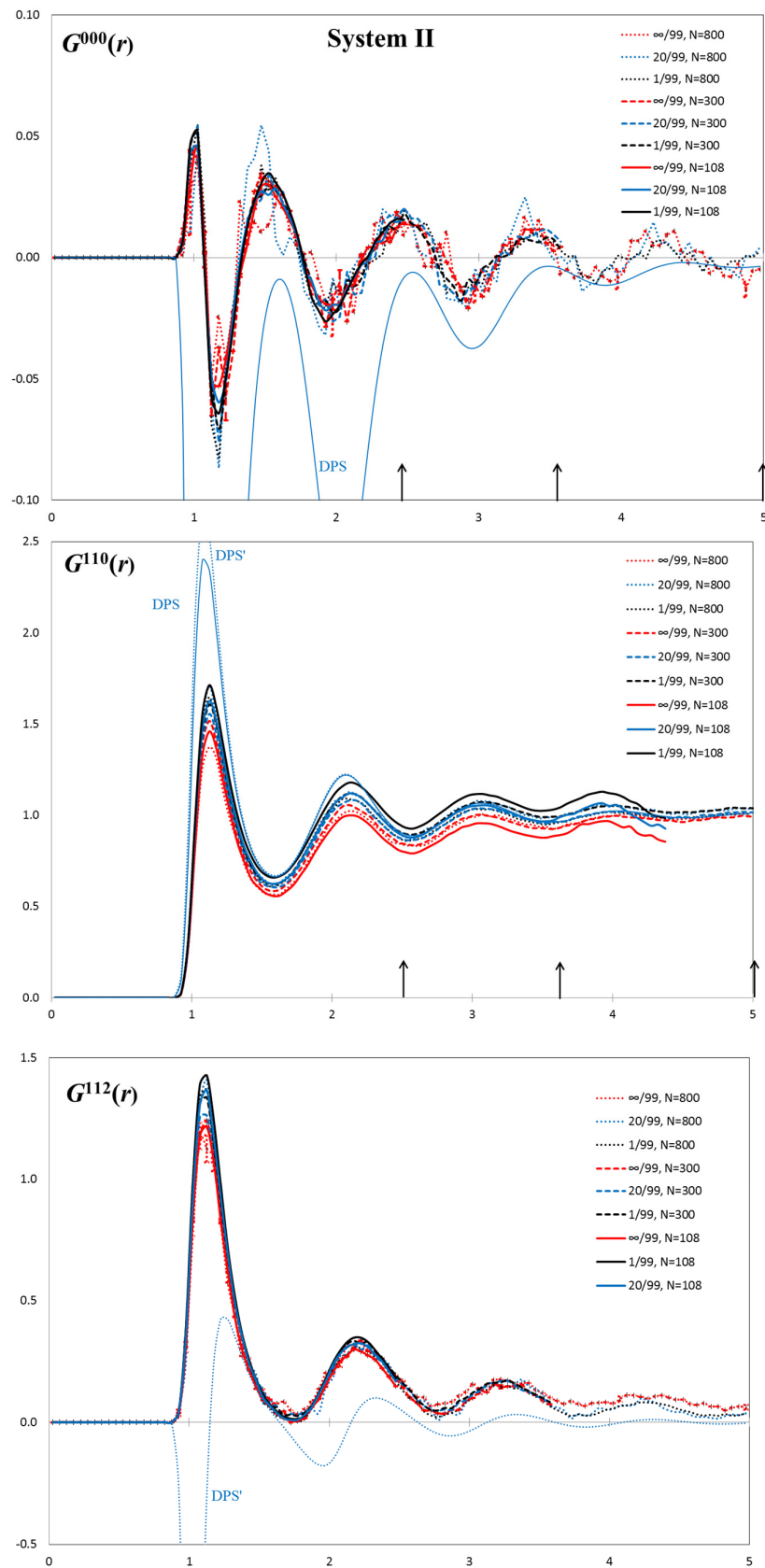


Figure 3

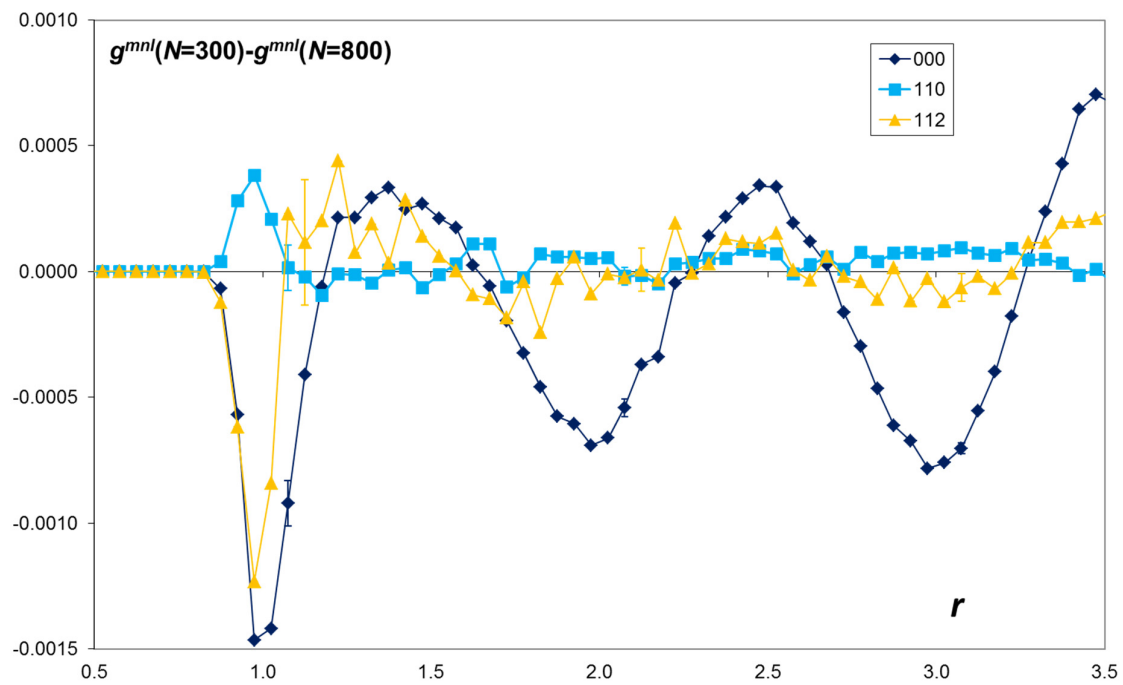
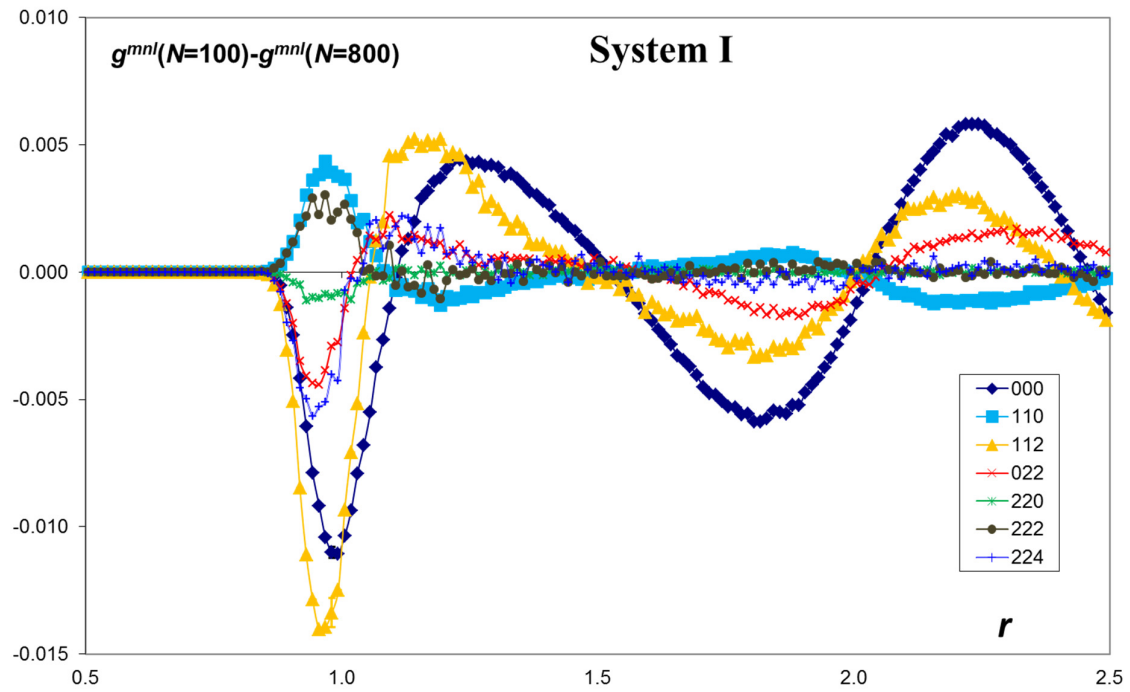


Figure 4

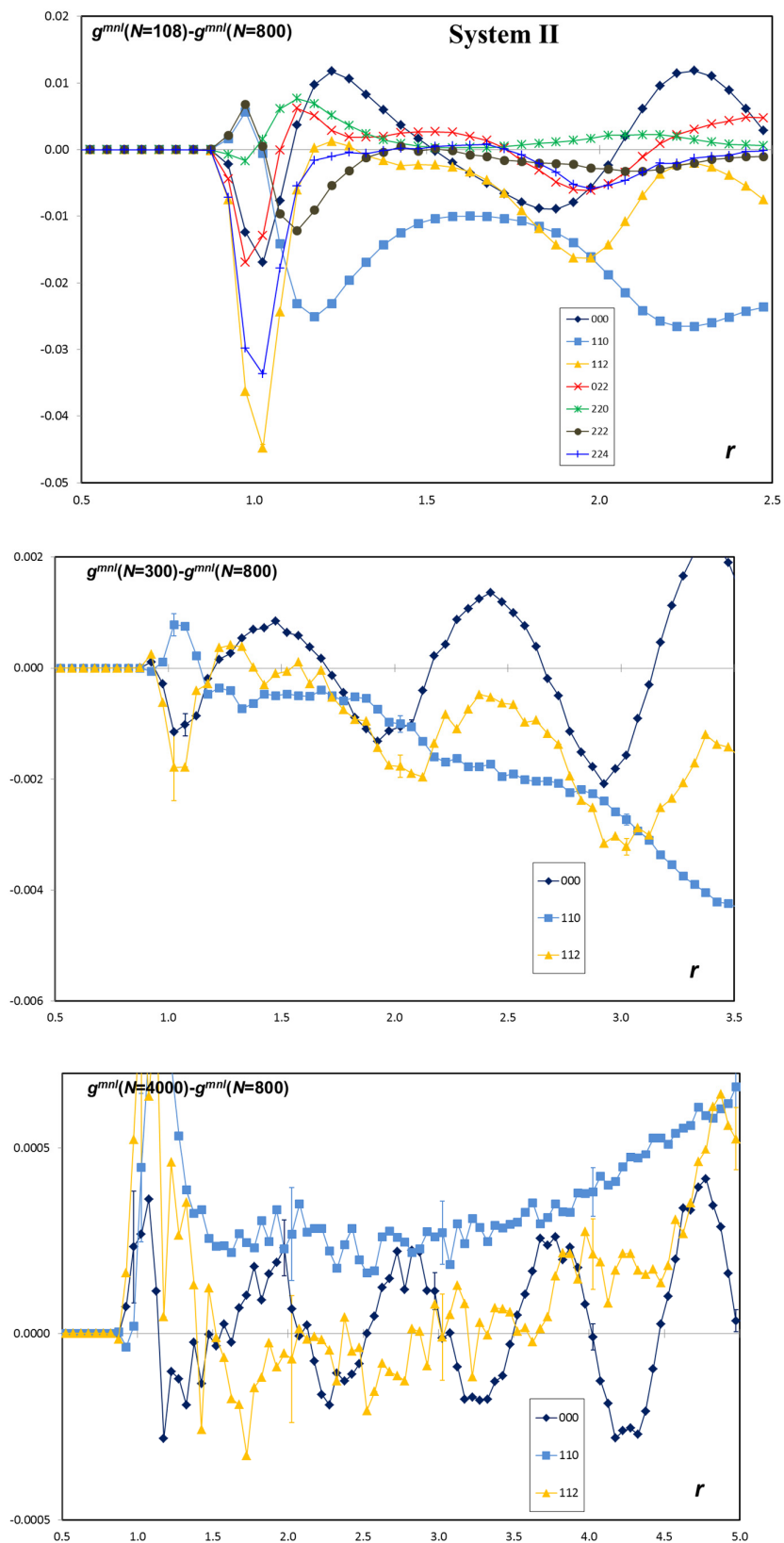


Figure 5

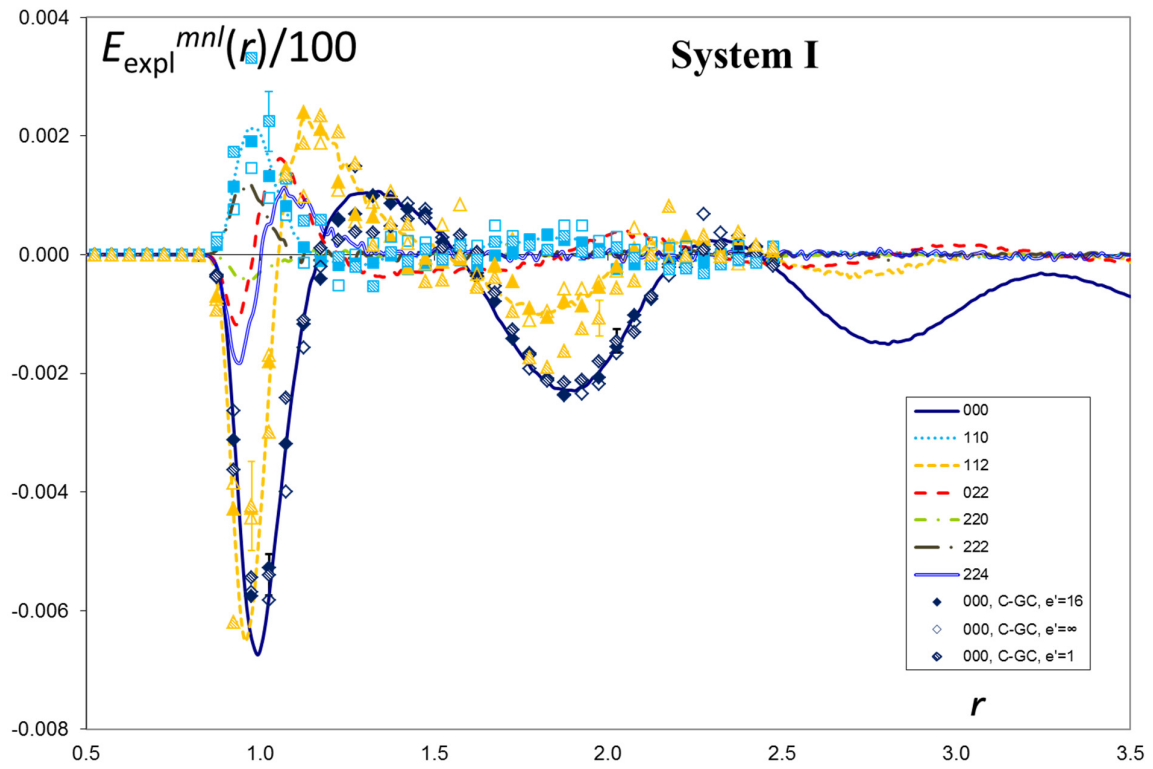


Figure 6

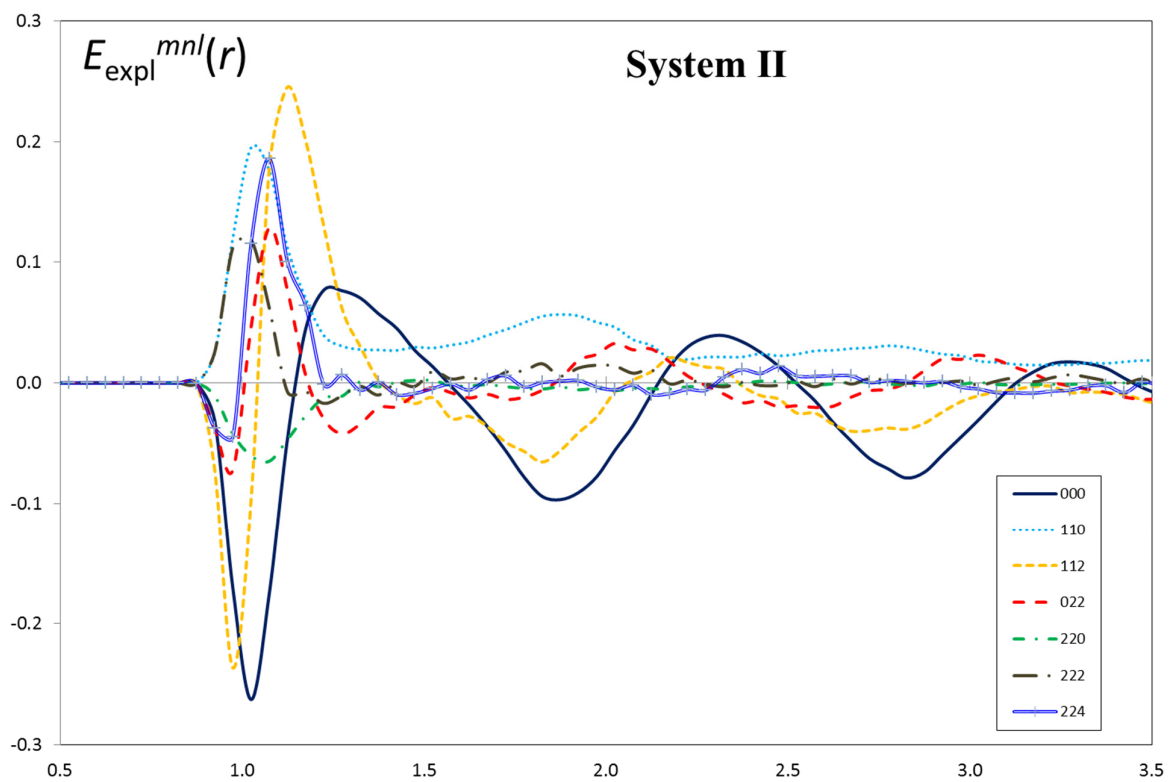




Figure 7

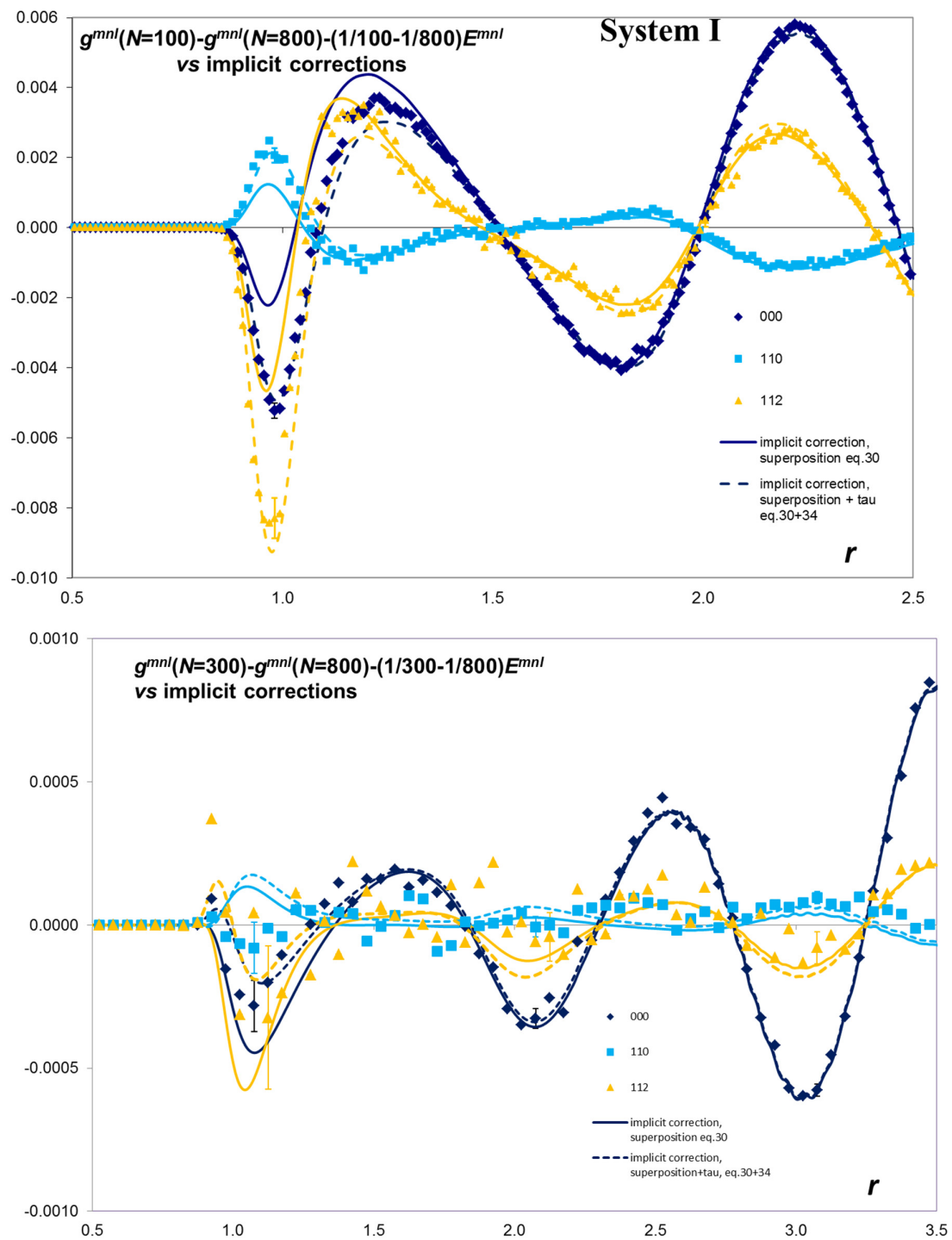


Figure 8

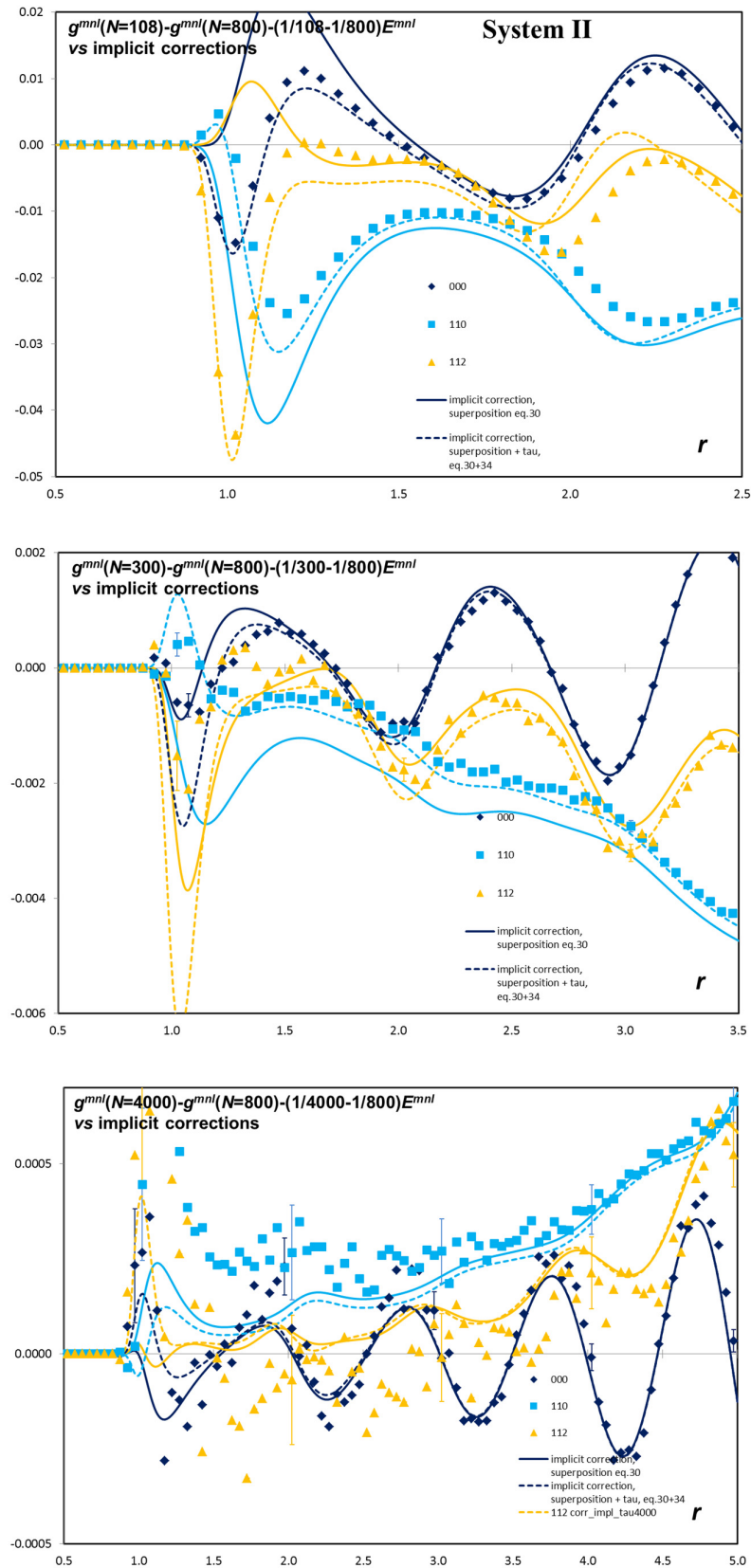


Figure 9

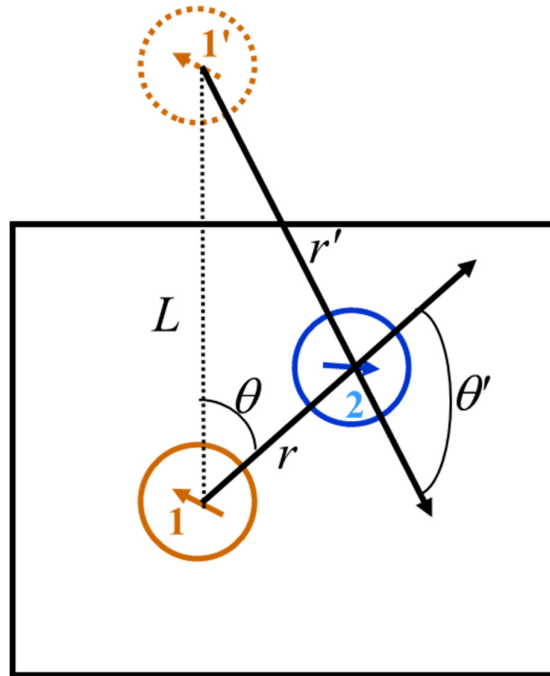


Figure 10

