

Measuring Structure and Interactions in Colloidal Fluids Using Test-Particle Insertion

Adam Edward Stones

Hertford College
University of Oxford

Supervisors: Prof. Dirk G. A. L. Aarts and
Prof. Roel P. A. Dullens

A thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2020

Abstract

We use the Potential Distribution Theorem to evaluate distribution functions from equilibrium configurations using test-particle insertion. We use this methodology to determine the contact value of the pair distribution function in hard-disk systems: in contrast with the conventional distance-histogram method, the insertion measurement is exact and does not require an approximate extrapolation. The resulting equations of state in both simulations and a hard-disk colloidal model system agree well with the predictions of Scaled Particle Theory. We also provide the first experimental measurement of the cavity distribution function $y^{(2)}(r)$ inside the hard core.

We next develop a model-free technique for measuring the pair potential $u^{(2)}(r)$ in pairwise-additive fluids, by matching the insertion and distance-histogram measurements of $g^{(2)}(r)$ using an iterative predictor-corrector scheme. We test the method extensively in simulation, before applying it successfully in a fluid of superparamagnetic colloidal particles, obtaining the anticipated form of $u^{(2)}(r)$ and the correct dependence on the applied magnetic field. We then extend the scheme to measure the full set of pair potentials in multicomponent fluids, demonstrating its efficacy in a three-component simulation.

We show that a given n-body distribution function $g^{(n)}$ $(n > 2)$ can be measured using n different insertion routes, which correspond to simultaneous insertion of between 1 and n test particles. We use these methods to measure $g^{(3)}$ in simulation: while the noise depends strongly on the number of simultaneous insertions, the resolution is superior to that of the distance-histogram method. Finally, we consider systems with three-body interactions. We show that matching $g^{(2)}(r)$ alone gives an effective pair potential which is unable to reproduce $g^{(3)}$ in pairwise-additive simulations. We therefore extend the predictor-corrector scheme to measure the three-body interaction $u^{(3)}$ by matching $g^{(3)}$, and test it in simulation. The scheme broadly recovers the correct $u^{(3)}$, but requires further development to reduce the noise.

Declaration

This thesis is submitted for the degree of Doctor of Philosophy in Physical and Theoretical Chemistry at the University of Oxford. No part of this thesis has been accepted or is currently being submitted for any degree, diploma, certificate or other qualification in this University or elsewhere, except where indicated at the beginning of the relevant chapter(s). This thesis is wholly my own work, except where indicated.

Contents

Abstract **i**

Declaration **ii**

1 Introduction **1**

1.1 Colloidal Systems 1

1.2 Interparticle Interactions 3

1.3 Thesis Content 4

2 Theory **6**

2.1 The Grand Ensemble 7

2.1.1 Probability Density Functions 7

2.1.2 Multicomponent Systems 9

2.2 Interactions in Fluids 10

2.2.1 The Potential Energy U_N 10

2.2.2 Pairwise Additivity 11

2.3 Particle Distribution Functions 11

2.3.1 The One-Body Density 11

2.3.2 Many-Body Density Functions 15

2.3.3 Distribution Functions and Structure 16

2.4 The Potential Distribution Theorem 18

2.5 Links with Thermodynamics 21

3 Models and Methods **23**

3.1 Colloidal Model Systems 24

3.1.1 Colloidal Interactions 24

3.1.2 Colloidal Monolayers 28

3.2 Experimental Methods 30

3.2.1 Preparing Colloidal Samples 30

3.2.2 Magnetic Field 32

3.2.3 Video-Microscopy 32

3.2.4 Image Analysis 33

3.3 Monte Carlo Simulations 36

3.3.1 Importance Sampling and the Metropolis Algorithm 37

3.3.2 Local Monte Carlo 39

3.3.3 Event-Chain Monte Carlo 42

3.3.4 Cell Schemes 43

4 Measuring the Contact Value Exactly in Hard-Disk Fluids **46**

4.1 Introduction 47

4.2 Theory 49

4.2.1 Hard Disks and the Potential Distribution Theorem 49

4.2.2 Contact Values and the Pressure Equation 49

4.2.3 Advantages of the Insertion Approach 50

4.2.4 The Cavity Distribution Function 51

4.2.5 Multicomponent Systems 54

4.3 Methods 56

4.3.1 Implementation 56

4.3.2 Protocols 57

4.4 Results and Discussion 59

4.4.1 Single-Component Hard Disks 59

4.4.2 Equation of State . . . 61
4.4.3 Binary Hard Disks . . . 61
4.5 Conclusion and Outlook . . . 62

5 Measuring the Pair Potential from Equilibrium Configurations 64

5.1 Introduction . . . 65
5.2 Theory . . . 66
5.2.1 The Predictor-Corrector Scheme . . . 66
5.2.2 Multicomponent Systems . . . 68
5.3 Methods . . . 69
5.3.1 Implementation . . . 69
5.3.2 Protocols . . . 73
5.4 Results and Discussion . . . 76
5.4.1 Simulations: Testing the Method . . . 76
5.4.2 Experiments: Superparamagnetic Colloidal Particles . . . 77
5.4.3 Multicomponent Simulation . . . 79
5.5 Uncertainty in the Measured Pair Potential . . . 80
5.5.1 Uncertainty Estimation Using the Measurement Noise . . . 80
5.5.2 Sensitivity to the Input Data . . . 83
5.6 Conclusion and Outlook . . . 84

6 Measuring Many-Body Distribution Functions in Simple Fluids 86

6.1 Introduction . . . 87
6.2 Theory . . . 89
6.2.1 Two-Body Distribution Function . . . 89
6.2.2 Three-Body Distribution Function . . . 90
6.2.3 Kirkwood Superposition Approximation . . . 92
6.3 Methods . . . 93
6.3.1 Implementation . . . 93

6.3.2 Protocols . . . 97
6.4 Results and Discussion . . . 98
6.4.1 Subsets Depending on a Single Variable . . . 98
6.4.2 Subsets with Two Particles at a Fixed Separation . . . 102
6.5 Conclusion and Outlook . . . 104

7 Measuring Three-Body Interactions 106
7.1 Introduction . . . 107
7.2 Theory . . . 109
7.2.1 A Simple Three-Body Interaction . . . 109
7.2.2 Effective Pair Potentials . . . 109
7.2.3 Three-Body Interactions: Predictor-Corrector Scheme . . . 111
7.3 Methods . . . 113
7.3.1 Implementation . . . 113
7.3.2 Protocols . . . 116
7.4 Results and Discussion . . . 118
7.4.1 Effective Pair Potentials . . . 118
7.4.2 Three-Body Interactions . . . 120
7.5 Conclusion and Outlook . . . 124

Conclusion and Outlook 126

Appendix 128
A Links between $g^{(2)}(r)$ and Thermodynamic Quantities . . . 128
A.1 The Energy Equation . . . 128
A.2 The Pressure Equation . . . 129
B Statistical Geometry of Hard-Disk Systems . . . 131

List of Publications 152

Acknowledgements 153

Chapter 1

Introduction

1.1 Colloidal Systems

According to Hansen and McDonald [1]:[1]

> Colloidal dispersions are highly asymmetric systems of mesoscopic particles suspended in a molecular solvent, with particle sizes typically in the range from 10 to 10^3 nm.

Such colloidal systems, often simply referred to as *colloids*, are ubiquitous: an everyday example is milk, which is composed of ∼100 nm fat particles dispersed in water [3]. In recent decades, colloids have seen widespread use as experimental model systems to address fundamental research questions in condensed matter physics, having been used to study crystals [4–6], glasses [7–10], liquids [11–15], liquid-crystals [16–18], gels [19, 20] and interfaces [21–23], for example. In contrast with molecular systems, the real-space structure and dynamics of systems composed of large colloidal particles is directly accessible, with single-particle resolution, using optical microscopy [8, 13, 24, 25]. This has enabled direct visualisation of, for example, the migration of grain boundaries [26–28], the formation and annihilation of topological defects [29–31], and cooperative motion in glasses [32, 33].

A key challenge in studying molecular systems is the small size of the particles, which is associated with small length scales (∼1 Å) and fast response times (∼1 ns) [3] which cannot

[1] The upper bound of the colloidal length scale is somewhat approximate, and some definitions use 10 µm instead [2]. In this thesis, we will work with spherical colloidal particles with a diameter ∼3 µm.

be probed by optical microscopy—most bulk experimental studies of these systems use indirect techniques such as X-ray and neutron scattering [34]. By contrast, colloidal systems can display structure at much larger length scales ($\sim$1 µm) and responses at slower timescales ($\gtrsim$1 s) [3], which are readily observed using optical microscopy techniques [8, 25, 35, 36]. Moreover, the thermal agitation of colloidal particles due to high-frequency collisions with solvent molecules, known as *Brownian Motion* [3], allows them to explore a configuration space and reach an equilibrium distribution, giving rise to a configurational statistical mechanics analogous to that of molecular systems [11]. Their effective interactions[2] are highly tuneable [38, 39]: colloidal particles can be made with various shapes [40, 41] and materials [42], and further modifications can be achieved by altering the solvent [35], adding salt [1, 38, 43] or depletants [1, 44, 45], or by grafting polymers onto the particles' surfaces [1, 46, 47]. In turn, this gives rise to diverse phase behaviour, which can include several liquid-crystalline phases [16, 18, 48, 49], liquid-gas phase separation [14, 21, 45, 50], a host of crystal structures [4, 51–55] and equilibrium cluster phases [56], providing rich opportunities for comparison with molecular systems. The inherent *softness* of colloidal systems, which respond strongly and slowly to weak external stimuli [3], also facilitates their experimental study: colloidal systems are readily manipulated by optical tweezing [57, 58], or by using electric [43, 59] and magnetic fields [60–62].

Another popular method for understanding condensed matter is molecular simulation [63, 64]. Providing the interactions within and between the molecular constituents are known or can be calculated, molecular dynamics [65, 66] and Monte Carlo [67, 68] techniques can be used to generate real-space snapshots to give insight into the real-space behaviour of materials, i.e. as a 'digital microscope' [69]. While popular, simulations have some drawbacks: generally, comparison with experiment is based only on integrated properties of the material, such as thermodynamic quantities or structure factors, and it may be possible to achieve good agreement of these properties even if the simulated configurations are physically unrealistic [70]. Limits in computational power mean that the number of particles simulated is much smaller than in a real molecular system, giving rise to finite-size effects which can influence the phase behaviour [69,71]. Some systems are challenging to simulate, for example multicomponent systems with large size-asymmetry [72,73], and the Monte Carlo approach is limited to the calculation of

[2]Strictly speaking, the colloid is a multicomponent mixture of the colloidal particles and solvent molecules, potentially with additional components such as ions and polymers. To describe the statistical mechanics in terms of the colloidal particle positions alone requires a coarse-graining procedure in which the solvent (and other) degrees of freedom are integrated out, resulting in state-dependent effective interactions [1, 37]. Further discussion is provided in Section 3.1.1.

static properties [1]. Colloidal systems, on the other hand, have the dual advantage of containing a macroscopic number of particles and providing experimentally-accessible real-space configurational data, allowing for 'test-tube simulations' [74] into phenomena such as two-dimensional melting [61, 75].

In addition to their use as experimental models of molecular systems, colloidal systems have many applications in their own right. They are widespread in industry, with applications in, e.g. food, cosmetics and paints [1], and are also widely used in technology, such as e-readers [76], with potential applications including photonic crystals [77], drug delivery [78] and artificial joints [79]. Finally, many biological systems are colloidal, ranging from proteins through to organelles and small cells [80].

1.2 Interparticle Interactions

Interparticle interactions are essential to the properties and functions exhibited by materials and are the driving force for phase transitions. For example, while repulsive interactions alone can drive symmetry-breaking transitions [34] such as the spontaneous crystallisation of hard spheres [65, 81] and nematisation of hard rods [82, 83] with increasing density, liquid-gas phase separation is additionally driven by attractive interactions[3] [12, 34, 50, 84]. Other phenomena, such as dynamic clustering, may be driven by competition between attractive and repulsive interactions at different length scales [56]. Furthermore, the responses of materials to external stimuli, such as their thermal expansion, are ultimately governed by the interactions between the constituent particles of the material, and the structure which results [85].

As discussed above, the widespread use of colloidal materials in science and industry is largely a result of their tuneable interparticle interactions; it is therefore surprising that our knowledge of these interactions is often only approximate, relying on theoretical assumptions [39]. Existing measurements of colloidal interactions are technically-demanding, requiring specialist equipment such as optical tweezers [86, 87], and usually measure interactions away from the context of the equilibrium colloidal fluid [86, 88, 89]. Straightforward Boltzmann Inversion [90] of the pair distribution function $g^{(2)}(r)$ neglects many-body contributions and is only

[3] In some cases, these attractions may be effective interactions induced by the presence of additional components in the system, known as *depletion interactions* [1, 45].

strictly valid in the limit of infinite dilution. Approaches based on comparing with $g^{(2)}(r)$ from simulation [91–94] are numerically-demanding and do not reliably recover the input interactions when tested on simulation data [95, 96]. The most important outcome of this thesis is the development of a fast, cheap and versatile method for measuring pairwise interactions in colloidal systems using particle-coordinate data from optical microscopy [97], which will provide for more *quantitative* comparison of experimental results with those from theory and simulation. The ability to measure interactions between colloidal particles is expected to be very useful when designing novel colloidal materials, with the role of interactions in, e.g. colloidal stability [1] having direct consequences for industrial formulations, for instance on their shelf-life [98]. The techniques also have clear applications in measuring effective interactions for use in coarse-grained simulations of biological systems [99].

1.3 Thesis Content

This thesis hinges on the development of an alternative method for measuring the particle distribution functions in classical fluids, of which the pair distribution function $g^{(2)}(r)$ is most common. This alternative *test-particle insertion* method is based on the Potential Distribution Theorem (PDT) [100, 101]. While this theorem dates back to 1963, its most popular application remains measuring the chemical potential in molecular simulation [1, 102]. Although the application to measure $g^{(2)}(r)$ is discussed even in the original articles [100, 101], it has been employed only sparingly in simulations, where the focus has been on measuring the cavity distribution function $y^{(2)}(r)$ in the core region [103–114]. This is most likely because $g^{(2)}(r)$ can be measured simply by counting particles using the conventional distance-histogram method [8,64]. The test-particle insertion approach, however, offers considerable advantages: unlike the distance-histogram method, it is formally exact and allows for arbitrary resolution in the interparticle separation [115]; it may also offer superior statistics at low densities.

We begin by introducing a statistical mechanical framework for handling classical fluids, and then use the PDT to express the particle distribution functions in terms of ensemble averages which can be measured by test-particle insertion. In Chapter 3, we introduce the relevant physics of colloidal model systems and the experimental and simulation methods used throughout the rest of the thesis. In Chapter 4, we use the test-particle insertion approach to provide an exact

solution to the long-standing problem of measuring the contact value of $g^{(2)}(r)$ in hard-disk systems [64], which gives direct access to the thermodynamics of the fluid [1,116]. We show its application in simulations of single-component and multicomponent systems, and also apply it to a previously-characterised hard-disk colloidal model system, with a known effective hard-disk diameter [15].

At first sight, the lack of knowledge of colloidal interactions seems to limit the application of the test-particle insertion approach in colloidal systems, where $g^{(2)}(r)$ has long been measured using the distance-histogram approach [8,90,117]. In Chapter 5, however, we show that enforcing consistency between $g^{(2)}(r)$ measured separately by the two methods provides a route to measuring colloidal interactions, in particular the effective pair potential. The experimental requirements are minimal and the measurements can be performed quickly and routinely. In addressing the inverse problem of obtaining interactions from equilibrium configurations (particle coordinates), the method also contributes to liquid state theory. We demonstrate the method's versatility in simulation, including its extension to multicomponent systems.

In Chapter 6, we show how the test-particle insertion method can be used to measure many-body distribution functions, and explicitly measure the three-body distribution function $g^{(3)}$ in simulation. The advantages of the insertion approach discussed in Chapter 4 are still more pronounced when measuring many-body distribution functions, which not only offer additional structural information [118], but also play an important role in closure approximations in liquid state theory [119–126].

Finally, in Chapter 7 we move beyond the pairwise-additive approximation by considering systems with three-body interactions. We perform simulations with a model three-body interaction and consider two approaches to characterise the interactions from equilibrium configurations. In the first instance, we employ the method of Chapter 5 to measure effective pair potentials which partially capture the three-body effects. Through pairwise-additive simulations, we show that while these effective interactions correctly reproduce $g^{(2)}(r)$, they are unable to reproduce $g^{(3)}$ and so do not fully capture the behaviour of the system [37,127,128]. We therefore extend the method of Chapter 5, enforcing consistency of $g^{(3)}$ and measuring explicit three-body interactions $u^{(3)}$. The implementation shows encouraging results in recovering the correct input $u^{(3)}$ from simulation data, and merits further optimisation to reduce the noise.

Chapter 2

Theory

In this chapter, we introduce the statistical mechanical framework used in this thesis and derive the key theoretical result underpinning the measurement of particle distribution functions using test-particle insertion. We begin by introducing the conventional description of fluids in the grand ensemble, showing how the probability density function (PDF) of fluid configurations depends on the potential energy, which can be written as a sum of terms involving increasing numbers of particles. We describe how this PDF can be reduced to a series of density functions which depend on the positions of only a few particles, and can be used to find expressions for the distribution functions which are related to the fluid's structure. We then use the Potential Distribution Theorem to express the distribution functions in terms of ensemble averages which can be measured by test-particle insertion, providing an alternative approach for measuring distribution functions from equilibrium fluid configurations. Finally, we discuss how the pair distribution function can be used to calculate thermodynamic quantities of the fluid.

2.1 The Grand Ensemble

We adopt the conventional statistical mechanical description of a fluid as an open system in which the total energy E and number of particles N fluctuate about their average values $\langle E \rangle$ and $\langle N \rangle$ [1,116,129,130]. The relevant statistical ensemble is the grand ensemble, where a microstate is specified by N and the particles' momentum and position coordinates,[1] $\mathbf{p}^N \equiv \mathbf{p}_1, \ldots, \mathbf{p}_N$ and $\mathbf{r}^N \equiv \mathbf{r}_1, \ldots, \mathbf{r}_N$ [1]. This ensemble is useful in the theoretical description of liquids because 'we can envisage arbitrary changes of the numbers of molecules in a system, or of the equilibrium density at any point' [130], without the need for compensating changes elsewhere as for a system with fixed N. It is also particularly relevant to the experiments on colloidal fluids in this thesis, since the number of colloidal particles observed fluctuates as particles enter and leave the microscope's field of view [131]. The majority of this chapter (excluding Section 2.4) draws on References [1, 116, 130], which can be consulted for further details.

2.1.1 Probability Density Functions

The probability density function (PDF) of a microstate in this ensemble is given by [1, 116]

$$P(\mathbf{p}^N, \mathbf{r}^N; N) = \frac{1}{\Xi h^{sN}} \exp\left[-\beta\left(H_N(\mathbf{p}^N, \mathbf{r}^N) - N\mu\right)\right], \tag{2.1}$$

where h is the Planck constant, s is the system dimensionality, H_N is the Hamiltonian, the inverse temperature $\beta = 1/k_\mathrm{B}T$, with $k_\mathrm{B}T$ the thermal energy, and μ is the chemical potential. The grand partition function Ξ is a normalisation factor obtained by summing over all microstates [116]

$$\Xi = \sum_{N=0}^{\infty} \frac{1}{N! h^{sN}} \int \mathrm{d}\mathbf{p}^N \int \mathrm{d}\mathbf{r}^N \exp\left[-\beta\left(H_N(\mathbf{p}^N, \mathbf{r}^N) - N\mu\right)\right], \tag{2.2}$$

with the factor $N!$ included to prevent counting configurations arising from the interchange of *'undistinguished'* particles as separate microstates [132].[2] The classical Hamiltonian H_N

[1] We largely follow the notational conventions of Reference [1]. In general, the position coordinates $\mathbf{r}^N$ specify the location of the particles' centres of mass and their orientations. In this thesis, we consider radially-symmetric particles and so we make no reference to the particles' orientations going forward.

[2] Following Sethna [132], 'undistinguished classical particles' include 'particles which in principle are not identical, but for which our Hamiltonian and measurement instruments treat identically (pollen grains and colloidal particles, for example).'

is composed of a kinetic energy part K_N, depending only on the particles' momenta, and a potential energy part U_N, which depends on their positions:

$$H_N(\mathbf{p}^N, \mathbf{r}^N) = K_N(\mathbf{p}^N) + U_N(\mathbf{r}^N) \tag{2.3}$$

$$= \sum_{i=1}^{N} \frac{p_i^2}{2m} + U_N(\mathbf{r}^N), \tag{2.4}$$

where on the second line K_N is written explicitly, with $p_i = |\mathbf{p}_i|$. The potential energy U_N comprises the interactions of the particles with each other and with external fields, and will be discussed in Section 2.2.

The integrals over $\mathbf{p}^N$ in (2.2) may be performed analytically, yielding

$$\Xi = \sum_{N=0}^{\infty} \frac{1}{N!\lambda^{sN}} \int \mathrm{d}\mathbf{r}^N \exp\left[-\beta\left(U_N(\mathbf{r}^N) - N\mu\right)\right] \tag{2.5}$$

$$= \sum_{N=0}^{\infty} \frac{z^N}{N!} \int \mathrm{d}\mathbf{r}^N \exp\left[-\beta U_N(\mathbf{r}^N)\right], \tag{2.6}$$

where

$$\lambda = \sqrt{\frac{h^2\beta}{2\pi m}}, \tag{2.7}$$

is the thermal de Broglie wavelength, and the reduced activity $z = \exp[\beta\mu]/\lambda^s$. This is the form of Ξ we use in this thesis, since we are concerned with the distribution of particle configurations and will ignore their momenta. To this end, we focus on the *configurational PDF*

$$P_\mathrm{c}(\mathbf{r}^N; N) = \frac{z^N}{\Xi} \exp\left[-\beta U_N(\mathbf{r}^N)\right], \tag{2.8}$$

which is the marginal PDF obtained from (2.1) by integrating out the momenta. Using (2.6) and (2.8),

$$\Xi = \sum_{N=0}^{\infty} \frac{1}{N!} \int \mathrm{d}\mathbf{r}^N P_\mathrm{c}(\mathbf{r}^N; N) \tag{2.9}$$

$$= \mathrm{Tr}_\mathrm{c}\left[P_\mathrm{c}\right], \tag{2.10}$$

where on the final line we have used classical trace notation [129],

$$\mathrm{Tr_c}\left[\dots\right] = \sum_{N=0}^{\infty} \frac{1}{N!} \int \mathrm{d}\mathbf{r}^N \cdots, \tag{2.11}$$

to denote a sum over the configurational degrees of freedom. The *configurational PDF* (2.8) can be used to express ensemble averages of a microscopic quantity $\mathcal{A}(\mathbf{r}^N; N)$, which depends on the instantaneous configuration of the system, as

$$\langle \mathcal{A} \rangle = \mathrm{Tr_c}\left[P_\mathrm{c}\mathcal{A}\right]. \tag{2.12}$$

Finally, we note that the relevant free energy for the grand ensemble is the grand free energy [132]

$$\Omega = -pV = -k_\mathrm{B}T \ln \Xi, \tag{2.13}$$

where p is the pressure and V is the volume. Its differential is given by [1]

$$\mathrm{d}\Omega = -p\mathrm{d}V - S\mathrm{d}T - \langle N \rangle \mathrm{d}\mu, \tag{2.14}$$

where S is the entropy, and implies that $\Omega = \Omega(V, T, \mu)$, i.e. a state in the grand ensemble is fully specified by the volume V, temperature T and chemical potential μ.

2.1.2 Multicomponent Systems

In this chapter, to keep the notation and derivations simple, we will only consider single-component systems, i.e. those composed of one species of particle. The results generalise straightforwardly to the multicomponent case, however, and we introduce the corresponding results for multicomponent systems in later chapters when required.

2.2 Interactions in Fluids

2.2.1 The Potential Energy U_N

The configurational PDF $P_c(\mathbf{r}^N; N)$ is independent of the particles' momenta and depends on their positions via the potential energy $U_N(\mathbf{r}^N)$ which includes the interactions of the particles with external fields and each other. In many cases[3], the potential energy may be written as a sum of terms depending on the positions of individual particles, pairs of particles, triplets of particles, etc. [133]

$$U_N(\mathbf{r}^N) = \sum_i u^{(1)}(\mathbf{r}_i) + \sum_i \sum_{j>i} u^{(2)}(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j} u^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \cdots, \tag{2.15}$$

where sums over multiple particles are written such that each pair, triplet, etc. is counted only once.

The terms in the first sum depend only on the position of one particle, accounting for interactions of particles with one-body external potentials (external fields) such as gravity, magnetic fields or optical landscapes due to laser tweezers. For a uniform system, $u^{(1)}(\mathbf{r}_i)$ is constant and taken to be 0, with the chemical potential adjusted to give the desired particle number density $\rho = \langle N \rangle / V$. In the second sum, the terms depend on the position of two particles and correspond to interactions between pairs of particles. The pair potential $u^{(2)}(\mathbf{r}_i, \mathbf{r}_j)$ would be the potential energy of particles i and j in isolation and in the absence of external fields. These interactions can have many different physical origins, such as electrostatics, dispersion forces or interactions between magnetic dipoles.

The terms in third sum depend on the simultaneous positions of three particles. Consider three particles in isolation and in the absence of external fields. The three-body potential $u^{(3)}$ is the total potential energy of this assembly, after subtracting the three pairwise interactions given by $u^{(2)}$ [corresponding to pairs (1,2),(1,3) and (2,3)]. An example is the Axilrod-Teller interaction, which is the leading three-body term in the calculation of the dispersion forces [116, 134]. Terms corresponding to the additional interactions of four or more particles are defined in a similar manner.

[3] See note 1 in Chapter 7 (page 107) for more details.

2.2.2 Pairwise Additivity

When studying simple interacting fluids, terms corresponding to interactions between three or more particles are often neglected, and the interparticle interactions are considered to be due to the pairwise interactions alone [1, 130]. In the absence of external fields, the total potential energy is then

$$U_N(\mathbf{r}^N) = \sum_i \sum_{j>i} u^{(2)}(\mathbf{r}_i, \mathbf{r}_j), \tag{2.16}$$

and the interactions are described as pairwise additive.

As written, the pair potential $u^{(2)}(\mathbf{r}_i, \mathbf{r}_j)$ depends on the absolute positions $\mathbf{r}_i$ and $\mathbf{r}_j$ of the particles, but may instead be considered to depend on the particles' centre of mass $\mathbf{R}_{ij} = (\mathbf{r}_i + \mathbf{r}_j)/2$ and their relative position $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. The interaction between the particles is usually a function of the relative position alone, and for radially-symmetric particles depends only on their separation $r_{ij} = |\mathbf{r}_{ij}|$, i.e. it is isotropic. In this case, the total potential energy is

$$U_N(\mathbf{r}^N) = \sum_i \sum_{j>i} u^{(2)}(r_{ij}). \tag{2.17}$$

This is the simplest continuum model of an interacting fluid, and is used in Chapters 4, 5 and 6. In Chapter 7, fluids with three-body interactions are considered.

2.3 Particle Distribution Functions

2.3.1 The One-Body Density

In the grand ensemble, the configurational PDF P_c depends on the number of particles N and their position coordinates $\mathbf{r}^N$. The domain of this function, or full configuration space, is a union of configuration spaces, each corresponding to a specific number of particles N [1]. For a given N, the configuration is specified by the coordinates $\mathbf{r}^N$, and for radially-symmetric particles the configuration space is sN-dimensional.[4] Hence, while all properties of the equilibrium fluid can in principle be calculated from P_c, direct calculations are usually too complex due to the structure

[4]For systems of particles without full radial symmetry, additional coordinates would be required to specify the orientation of each particle.

and dimensionality of its domain, and alternative approaches such as molecular simulation are used (see Section 3.3).

To calculate many properties of the fluid, it is not necessary to know the full PDF P_c, but only its dependence on much smaller set of particle positions, having integrated out the remainder to give a marginal (or *reduced* [135]) PDF. For instance, the average interaction energy $\langle U_\mathrm{ext}\rangle$ with an external field $u^{(1)}(\mathbf{r})$ can be written as

$$\langle U_\mathrm{ext}\rangle = \int u^{(1)}(\mathbf{r})\rho^{(1)}(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{2.18}$$

where $\rho^{(1)}(\mathbf{r})$ is the one-body density. Integrals of this kind are central to the density functional theory (DFT) approach to liquid state calculations [1, 129].

To determine $\rho^{(1)}(\mathbf{r})$ we begin by integrating $P_\mathrm{c}(\mathbf{r}^N; N)$ for a specific N with respect to all coordinates except $\mathbf{r}_1$:

$$\int \mathrm{d}\mathbf{r}^{(N-1)} P_\mathrm{c}(\mathbf{r}^N; N), \tag{2.19}$$

where we have introduced the notation $\mathrm{d}\mathbf{r}^{(N-n)} = \mathrm{d}\mathbf{r}_{n+1}\cdots\mathrm{d}\mathbf{r}_N$. Since we are interested in microstates where *any* particle (not just particle 1) is at $\mathbf{r}_1$, we need to multiply this by N. We then sum over the possible values of N, carefully avoiding over-counting of microstates (as discussed in Section 2.1), to yield

$$\rho^{(1)}(\mathbf{r}_1) = \sum_{N=0}^{\infty} \frac{N}{N!} \int \mathrm{d}\mathbf{r}^{(N-1)} P_\mathrm{c}(\mathbf{r}^N; N). \tag{2.20}$$

The $N = 0$ term of this sum is zero, since to have a particle at $\mathbf{r}_1$ there must be at least one particle in the system, and

$$\rho^{(1)}(\mathbf{r}_1) = \sum_{N=1}^{\infty} \frac{1}{(N-1)!} \int \mathrm{d}\mathbf{r}^{(N-1)} P_\mathrm{c}(\mathbf{r}^N; N) \tag{2.21}$$

$$= \frac{1}{\Xi} \sum_{N=1}^{\infty} \frac{z^N}{(N-1)!} \int \mathrm{d}\mathbf{r}^{(N-1)} \exp[-\beta U_N(\mathbf{r}^N)], \tag{2.22}$$

where on the second line we have substituted P_c from (2.8). We can verify this expression by

integrating (2.20) over all space [1]

$$\int \rho^{(1)}(\mathbf{r}_1)\mathrm{d}\mathbf{r}_1 = \int \mathrm{d}\mathbf{r}_1 \left(\sum_{N=0}^{\infty} \frac{N}{N!} \int \mathrm{d}\mathbf{r}^{(N-1)} P_\mathrm{c}(\mathbf{r}^N; N) \right) \tag{2.23}$$

$$= \sum_{N=0}^{\infty} \frac{1}{N!} \int \mathrm{d}\mathbf{r}^N N P_\mathrm{c}(\mathbf{r}^N; N) \tag{2.24}$$

$$= \mathrm{Tr}_\mathrm{c}\left[P_\mathrm{c} N\right] = \langle N \rangle , \tag{2.25}$$

arriving at the average number of particles, as expected.

The one-body density can be interpreted in several ways which will be useful in this thesis. Firstly (following Reference [1]), we can rewrite (2.20) as a sum over the full configuration space by using a Dirac delta to restore the integral over $\mathbf{r}_1$:

$$\rho^{(1)}(\mathbf{r}) = \sum_{N=0}^{\infty} \frac{1}{N!} N \int \mathrm{d}\mathbf{r}^N P_\mathrm{c}(N; \mathbf{r}^N) \delta(\mathbf{r}_1 - \mathbf{r}). \tag{2.26}$$

Since P_c is symmetric with respect to particle interchange, the factor N can be replaced by a sum over all particles i

$$\rho^{(1)}(\mathbf{r}) = \sum_{N=0}^{\infty} \frac{1}{N!} \sum_{i=1}^{N} \int \mathrm{d}\mathbf{r}^N P_\mathrm{c}(N; \mathbf{r}^N) \delta(\mathbf{r}_i - \mathbf{r}) \tag{2.27}$$

$$= \sum_{N=0}^{\infty} \frac{1}{N!} \int \mathrm{d}\mathbf{r}^N P_\mathrm{c}(N; \mathbf{r}^N) \sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}) \tag{2.28}$$

$$= \mathrm{Tr}_\mathrm{c}\left[P_\mathrm{c} \sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}) \right] \tag{2.29}$$

$$= \left\langle \sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}) \right\rangle , \tag{2.30}$$

where on the final line we have used equation (2.12). Thus $\rho^{(1)}(\mathbf{r})$ is the ensemble average of a microscopic density function which is a product of Dirac deltas centred at the instantaneous particle positions [1]. In this sense, $\rho^{(1)}(\mathbf{r})$ is seen as the ensemble-averaged PDF associated with finding a particle (any particle) at $\mathbf{r}$.

Secondly, the form of the sum in (2.22) resembles that of the grand partition function Ξ

in equation (2.6). Substituting (2.6), we obtain

$$\rho^{(1)}(\mathbf{r}_1) = \frac{\sum_{N=1}^{\infty} \frac{z^N}{(N-1)!} \int \mathrm{d}\mathbf{r}^{(N-1)} \exp\left[-\beta U_N(\mathbf{r}^N)\right]}{\sum_{N=0}^{\infty} \frac{z^N}{N!} \int \mathrm{d}\mathbf{r}^N \exp\left[-\beta U_N(\mathbf{r}^N)\right]}. \tag{2.31}$$

In some sense, Ξ represents the number of thermally-accessible microstates of the system [136] and, using (2.31), $\rho^{(1)}(\mathbf{r}_1)$ can be interpreted as the proportion of thermally-accessible microstates of the system which have any particle at $\mathbf{r}_1$. This is consistent with our original assertion that $\rho^{(1)}(\mathbf{r}_1)$ is the marginal probability density of having a particle at $\mathbf{r}_1$ irrespective of the positions of the other particles.[5] The density will thus be high at a given position if, when a particle is located there, the number of microstates accessible to the other particles is large.

Finally, the comparison with Ξ can be made more formal by substituting $M = N - 1$ in (2.22). Shifting all the particle indices down by 1 and replacing $\mathbf{r}_1$ with $\mathbf{r}$, we obtain

$$\rho^{(1)}(\mathbf{r}) = \frac{z}{\Xi} \sum_{M=0}^{\infty} \frac{z^M}{M!} \int \mathrm{d}\mathbf{r}^M \exp\left[-\beta U_{M+1}(\mathbf{r}, \mathbf{r}^M)\right]. \tag{2.32}$$

Noting that M is a dummy variable, the sum now resembles that of (2.6) except that the potential energy also includes interactions of the M particles with an additional particle at $\mathbf{r}$, i.e. it is the grand partition function $\Xi^{(1)}(\mathbf{r})$ of an *inhomogeneous* system of M particles, with the external potential arising due to a particle fixed at r. From (2.13), $\Xi = \exp[-\beta\Omega]$, and we can rewrite (2.32) as

$$\rho^{(1)}(\mathbf{r}) = z \exp\left[-\beta\left(\Omega^{(1)}(\mathbf{r}) - \Omega\right)\right] \tag{2.33}$$

$$= z \exp\left[-\beta \Delta\Omega^{(1)}(\mathbf{r})\right], \tag{2.34}$$

where $\Omega^{(1)}(\mathbf{r})$ is the grand free energy of the inhomogeneous system. In this interpretation [137], $\rho^{(1)}(\mathbf{r})$ is related to the grand free energy difference $\Delta\Omega^{(1)}(\mathbf{r})$ associated with inserting a fixed particle at $\mathbf{r}$.[6]

[5] Note that $\rho^{(1)}(\mathbf{r}_1)$ differs from a typical marginal PDF in that instead of being normalised to 1, it is normalised to give the average number of particles as shown in (2.25). This is because $\rho^{(1)}(\mathbf{r}_1)$ is the probability density of *any* particle being at $\mathbf{r}_1$, rather than a specific particle.

[6] In a uniform system, $\Delta\Omega^{(1)}(\mathbf{r})$ is constant and equal to the excess chemical potential of the fluid μ_{ex} when compared with that of a perfect gas at the same density.

2.3.2 Many-Body Density Functions

Many properties depend not only on $\rho^{(1)}(\mathbf{r})$, but must also account for correlations in particle positions. These are not captured by $\rho^{(1)}(\mathbf{r})$, which only describes the likelihood of finding a particle at $\mathbf{r}$ and does not provide any information about the likelihood of simultaneously finding two particles at separate positions $\mathbf{r}_1$ and $\mathbf{r}_2$. In a uniform fluid $\rho^{(1)} = \rho$ at both positions, but if the separation $|\mathbf{r}_2 - \mathbf{r}_1|$ is small (much less than the particles' size), it is very unlikely that particles will simultaneously be found at $\mathbf{r}_1$ and $\mathbf{r}_2$ due to the high energy cost of particle overlap. Similar considerations exist when considering three or more positions.

The relevant information is summarised by a hierarchy of many-body density functions $\rho^{(n)}(\mathbf{r}^n)$, which are derived analogously to the one-body case in (2.19) to (2.22). Integrating over the positions of n particles yields the results [1]

$$\rho^{(n)}(\mathbf{r}^n) = \sum_{N=0}^{\infty} \frac{1}{N!} \left(\frac{N!}{(N-n)!} \right) \int \mathrm{d}\mathbf{r}^{(N-n)} P_{\mathrm{c}}(\mathbf{r}^N; N) \tag{2.35}$$

$$= \sum_{N=n}^{\infty} \frac{1}{(N-n)!} \int \mathrm{d}\mathbf{r}^{(N-n)} P_{\mathrm{c}}(\mathbf{r}^N; N) \tag{2.36}$$

$$= \frac{1}{\Xi} \sum_{N=n}^{\infty} \frac{z^N}{(N-n)!} \int \mathrm{d}\mathbf{r}^{(N-n)} \exp\left[-\beta U_N(\mathbf{r}^N)\right], \tag{2.37}$$

where instead of multiplying by N on the first line, we have multiplied by the combinatorial factor associated with the number of ways of selecting a subset of n particles from a set of N particles.[7]

As with $\rho^{(1)}(\mathbf{r})$, examining (2.35) to (2.37) yields different interpretations of $\rho^{(n)}$. Following the reasoning of (2.26) and (2.30) [1], we find that

$$\rho^{(n)}(\mathbf{s}^n) = \left\langle \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \cdots \sum_{\substack{z=1 \\ z \neq i,\ldots,y}}^{N} \delta(\mathbf{r}_i - \mathbf{s}_1)\delta(\mathbf{r}_j - \mathbf{s}_2) \cdots \delta(\mathbf{r}_z - \mathbf{s}_n) \right\rangle, \tag{2.38}$$

where we have used $\mathbf{s}_i$ to indicate a fixed position in space and $\mathbf{r}_i$ to denote a coordinate which varies under integration. Similarly to $\rho^{(1)}(\mathbf{r})$, $\rho^{(n)}$ is here interpreted as the average of a microscopic n-body density function where each ordered set of n particles is represented by a

[7] Note that in this factor, all permutations of the n selected particles are included.

product of Dirac deltas.[8] Additionally, $\rho^{(n)}$ can be written similarly to (2.31) as

$$\rho^{(n)}(\mathbf{s}^n) = \frac{\sum_{N=n}^{\infty} \frac{z^N}{(N-n)!} \int \mathrm{d}\mathbf{r}^{(N-n)} \exp\left[-\beta U_N(\mathbf{s}^n, \mathbf{r}^{(N-n)})\right]}{\sum_{N=0}^{\infty} \frac{z^N}{N!} \int \mathrm{d}\mathbf{r}^N \exp\left[-\beta U_N(\mathbf{r}^N)\right]}, \tag{2.39}$$

and interpreted as the proportion of thermally-accessible microstates with particles at $\mathbf{s}^n$, i.e. as the marginal PDF of having n particles at $\mathbf{s}^n$. Lastly, we can follow the reasoning of (2.32) and substitute $M = N - n$ into (2.37) to obtain

$$\rho^{(n)}(\mathbf{s}^n) = \frac{z^n}{\Xi} \sum_{M=0}^{\infty} \frac{z^M}{M!} \int \mathrm{d}\mathbf{r}^M \exp\left[-\beta U_{M+n}(\mathbf{s}^n, \mathbf{r}^M)\right] \tag{2.40}$$

$$= z^n \exp\left[-\beta \Delta\Omega^{(n)}(\mathbf{s}^n)\right], \tag{2.41}$$

where on the final line we have recognised the sum as the inhomogeneous partition function $\Xi^{(n)}(\mathbf{s}^n)$ and thus written $\rho^{(n)}(\mathbf{s}^n)$ in terms of the grand free energy difference $\Delta\Omega^{(n)}(\mathbf{s}^n)$ of inserting fixed particles at $\mathbf{s}^n$ [137].

2.3.3 Distribution Functions and Structure

In a perfect gas there are no interparticle interactions and therefore no correlations between particles' positions: the existence of a particle at $\mathbf{r}$ has no implications for the likely positions of other particles and $\rho^{(n)}(\mathbf{r}^n)$ reduces to a product of one-body density functions

$$\rho_{\mathrm{pg}}^{(n)}(\mathbf{r}^n) = \prod_{i=1}^{n} \rho^{(1)}(\mathbf{r}_i). \tag{2.42}$$

It is useful to define a hierarchy of *distribution functions* $g^{(n)}(\mathbf{r}^n)$ quantifying the deviation from this random perfect-gas behaviour [1]:

$$g^{(n)}(\mathbf{r}^n) = \frac{\rho^{(n)}(\mathbf{r}^n)}{\prod_{i=1}^{n} \rho^{(1)}(\mathbf{r}_i)}. \tag{2.43}$$

These functions are important because they characterise the n-body structure of the fluid, and are central to the methods developed and used in this thesis.

To understand the link between distribution functions and structure, we introduce the

[8] Corresponding to Dirac deltas in the configuration space of n particles, which is $\mathbb{R}^n$.

conditional density function [138–140]

$$\rho^{(n-m)}(\mathbf{r}^{(n-m)}|\mathbf{r}^m) = \frac{\rho^{(n)}(\mathbf{r}^n)}{\rho^{(m)}(\mathbf{r}^m)} \tag{2.44}$$

$$= \frac{\sum_{N=n}^{\infty} \frac{z^N}{(N-n)!} \int \mathrm{d}\mathbf{r}^{(N-n)} \exp\left[-\beta U_N(\mathbf{r}^N)\right]}{\sum_{N=m}^{\infty} \frac{z^N}{(N-m)!} \int \mathrm{d}\mathbf{r}^{(N-m)} \exp\left[-\beta U_N(\mathbf{r}^N)\right]}, \tag{2.45}$$

where $n \geq m$. Following the interpretations of (2.31) and (2.39), we see that $\rho^{(n-m)}(\mathbf{r}^{(n-m)}|\mathbf{r}^m)$ is the ratio of the number of thermally-accessible microstates with n particles at $\mathbf{r}^n$ to the number when only m particles are at $\mathbf{r}^m$, and thus represents the conditional PDF of having $(n-m)$ particles at $\mathbf{r}^{(n-m)}$ *given* that there are already m particles at $\mathbf{r}^m$. Substituting (2.44) for $\rho^{(n)}(\mathbf{r}^n)$ in (2.43) we obtain

$$g^{(n)}(\mathbf{r}^n) = \frac{\rho^{(n-m)}(\mathbf{r}^{(n-m)}|\mathbf{r}^m)\rho^{(m)}(\mathbf{r}^m)}{\prod_{i=1}^{n} \rho^{(1)}(\mathbf{r}_i)}, \tag{2.46}$$

$$\frac{g^{(n)}(\mathbf{r}^n)}{g^{(m)}(\mathbf{r}^m)} = \frac{\rho^{(n-m)}(\mathbf{r}^{(n-m)}|\mathbf{r}^m)}{\prod_{i=m+1}^{n} \rho^{(1)}(\mathbf{r}_i)}. \tag{2.47}$$

In a uniform system, $\rho^{(1)}(\mathbf{r}) = \rho$ is independent of position and the product in the denominator can be replaced by ρ^{n-m} to give the result [141, 142]

$$\frac{g^{(n)}(\mathbf{r}^n)}{g^{(m)}(\mathbf{r}^m)} = \frac{\rho^{(n-m)}(\mathbf{r}^{(n-m)}|\mathbf{r}^m)}{\rho^{n-m}}. \tag{2.48}$$

When $m = n - 1$,

$$\frac{g^{(n)}(\mathbf{r}^n)}{g^{(n-1)}(\mathbf{r}^{n-1})} = \frac{\rho^{(1)}(\mathbf{r}_n|\mathbf{r}^{n-1})}{\rho}, \tag{2.49}$$

and so the ratio of the distribution functions tells us how $\rho^{(1)}(\mathbf{r}_n)$ deviates from its equilibrium value due to the presence of particles at $\mathbf{r}^{n-1}$. This is the structural interpretation of the distribution functions.

As an example, consider the case where $n = 2$. Noting from (2.43) that $g^{(1)} = 1$, (2.49) becomes

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{\rho^{(1)}(\mathbf{r}_2|\mathbf{r}_1)}{\rho}, \tag{2.50}$$

$$g^{(2)}(r) = \frac{\rho^{(1)}(r)}{\rho}, \tag{2.51}$$

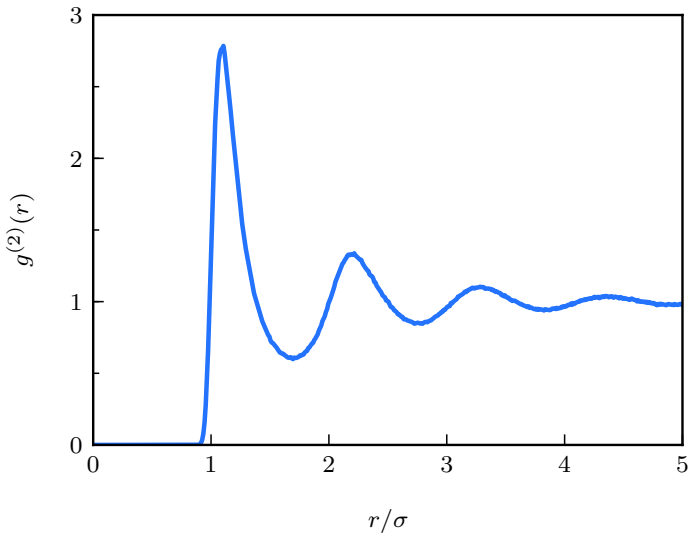


Figure 2.1: Example $g^{(2)}(r)$ for a 2D Lennard-Jones system (see Chapter 5), showing the pairwise structure of the fluid.

where on the final line we have noted that in a uniform system of isotropic particles, $\rho^{(1)}(\mathbf{r}_2|\mathbf{r}_1)$ and and therefore $g^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ depend only on the separation $r = |\mathbf{r}_1 - \mathbf{r}_2|$, and $\rho^{(1)}(r)$ is taken to implicitly indicate the presence of a particle at the origin. In this context, $g^{(2)}(r)$ is typically termed the *radial distribution function*, and (2.51) is the familiar result that $g^{(2)}(r)$ is the ratio of the local density at a fixed separation from a reference particle to the bulk density of the fluid. Figure 2.1 shows a typical $g^{(2)}(r)$, and clearly illustrates the pairwise structure of the fluid. At small separations, $g^{(2)}(r)$ is essentially zero: there are never pairs of particles at this separation due to the high associated energy cost. There are then a series of progressively less-well-defined peaks associated with coordination shells of particles about the reference particle. At larger separations, $g^{(2)}(r) \to 1$, showing that the positional order of the fluid is short-ranged, i.e. the presence of the reference particle is 'forgotten' within a short distance. This contrasts with the long-ranged positional order in crystalline solids, which can extend over macroscopic length scales.

2.4 The Potential Distribution Theorem

First published by Widom in 1963 [100] and Jackson and Klein in 1964 [101], the Potential Distribution Theorem (PDT) is best known as the basis for the test-particle insertion method for measuring the chemical potential in molecular simulations [63,64], which has also been applied

in colloidal experiments [13]. While both original articles derive expressions for $g^{(2)}(r)$ using parallel arguments, use of the PDT in measuring distribution functions has been very limited in favour of an alternative distance-histogram approach based on directly applying (2.51) [8, 64].[9] Here, we extend the derivations of Reference [100] to derive general expressions for many-body distribution functions. Although the extension of the PDT to many-body distribution functions is already known [137, 143, 144], here we focus on deriving expressions which will allow us to measure these functions by test-particle insertion in colloidal fluids and simulations.

In this section, we will again use $\mathbf{s}_i$ to indicate a fixed position and $\mathbf{r}_i$ for a coordinate which varies under integration, such that the conditional density function of (2.45) reads

$$\rho^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m) = \frac{\sum_{N=n}^{\infty} \frac{z^N}{(N-n)!} \int \mathrm{d}\mathbf{r}^{(N-n)} \exp\left[-\beta U_N(\mathbf{s}^n, \mathbf{r}^{(N-n)})\right]}{\sum_{N=m}^{\infty} \frac{z^N}{(N-m)!} \int \mathrm{d}\mathbf{r}^{(N-m)} \exp\left[-\beta U_N(\mathbf{s}^m, \mathbf{r}^{(N-m)})\right]}. \tag{2.52}$$

Extending the arguments of Widom [100], we split U_N in the numerator of (2.52) into a term representing all of the interactions between the $N-(n-m)$ particles at $\mathbf{s}^m$ and $\mathbf{r}^{(N-n)}$, and the additional potential energy $\Psi(\mathbf{s}^{(n-m)}; \mathbf{s}^m, \mathbf{r}^{(N-n)}) = \Psi(\mathbf{s}^{(n-m)})$ arising from interactions of the $n-m$ particles at $\mathbf{s}^{(n-m)}$ with each other and the other particles of the system. The numerator becomes

$$\sum_{N=n}^{\infty} \frac{z^N}{(N-n)!} \int \mathrm{d}\mathbf{r}^{(N-n)} \exp\left[-\beta U_{N-(n-m)}(\mathbf{s}^m, \mathbf{r}^{(N-n)})\right] \exp\left[-\beta \Psi(\mathbf{s}^{(n-m)})\right]. \tag{2.53}$$

We now make the substitution $M = N-(n-m)$ and renumber $\mathbf{r}^{(N-n)}$ as $\mathbf{r}^{(M-m)}$ to give

$$z^{n-m} \sum_{M=m}^{\infty} \frac{z^M}{(M-m)!} \int \mathrm{d}\mathbf{r}^{(M-m)} \exp\left[-\beta U_M(\mathbf{s}^m, \mathbf{r}^{(M-m)})\right] \exp\left[-\beta \Psi(\mathbf{s}^{(n-m)})\right], \tag{2.54}$$

which, when substituted into (2.52), yields

$$\begin{aligned}
\frac{\rho^{(n)}(\mathbf{s}^n)}{\rho^{(m)}(\mathbf{s}^m)} &= \frac{z^{n-m} \sum_{M=m}^{\infty} \frac{z^M}{(M-m)!} \int \mathrm{d}\mathbf{r}^{(M-m)} \exp\left[-\beta U_M(\mathbf{s}^m, \mathbf{r}^{(M-m)})\right] \exp\left[-\beta \Psi(\mathbf{s}^{(n-m)})\right]}{\sum_{N=m}^{\infty} \frac{z^N}{(N-m)!} \int \mathrm{d}\mathbf{r}^{(N-m)} \exp\left[-\beta U_N(\mathbf{s}^m, \mathbf{r}^{(N-m)})\right]} \\
&= z^{n-m} \left\langle \exp\left[-\beta \Psi(\mathbf{s}^{(n-m)})\right] \right\rangle_{\mathbf{s}^m} \\
&= z^{n-m} P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m),
\end{aligned} \tag{2.55}$$

[9] Use of the distribution function result has focussed on calculating the cavity distribution function $y^{(2)}(r)$. A full discussion is provided at the beginning of Chapter 4.

where we have replaced the left hand side by a ratio of density functions using (2.44). On the second line, we have noted that the right-hand side has the form of z^{n-m} multiplied by an ensemble average of $\exp[-\beta\Psi]$ over the subset of microstates with particles at $\mathbf{s}^m$, and on the final line we have simplified the notation by defining

$$P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m) = \left\langle \exp[-\beta\Psi(\mathbf{s}^{(n-m)})] \right\rangle_{\mathbf{s}^m}. \tag{2.56}$$

In the case that $n = 1$ and $m = 0$, (2.55) reduces to

$$\rho^{(1)}(\mathbf{s}_1) = zP^{(1)}(\mathbf{s}_1), \tag{2.57}$$

where we have noted that $\rho^{(0)} = 1$ using (2.39). This is Widom's result for inhomogeneous systems [145], and reduces to his original result in the uniform case [100].

As when we were deriving (2.47), we substitute (2.55) for $\rho^{(n)}(\mathbf{r}^n)$ in (2.43) to give

$$g^{(n)}(\mathbf{s}^n) = z^{n-m} g^{(m)}(\mathbf{s}^m) \frac{P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m)}{\prod_{i=m+1}^{n} \rho^{(1)}(\mathbf{s}_i)}. \tag{2.58}$$

Finally, we use (2.57) for $\rho^{(1)}(\mathbf{s}_i)$ to yield

$$\frac{g^{(n)}(\mathbf{s}^n)}{g^{(m)}(\mathbf{s}^m)} = \frac{P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m)}{\prod_{i=m+1}^{n} P^{(1)}(\mathbf{s}_i)}. \tag{2.59}$$

This is the central theoretical result of this thesis. Strikingly similar to (2.47), (2.59) is a hierarchy of equations which each express a ratio of n and m-body distribution functions as a ratio of ensemble averages rather than density functions. The ensemble averages may be evaluated by test-particle insertion, providing an alternative to the distance-histogram approach for measuring distribution functions from equilibrium configurations of fluids.

To demonstrate the use of (2.59), again consider the case where $n = 2$ and $m = 1$, which for a uniform system of isotropic particles yields

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{P^{(1)}(\mathbf{r}_2|\mathbf{r}_1)}{P}, \tag{2.60}$$

$$g^{(2)}(r) = \frac{P^{(1)}(r)}{P} \tag{2.61}$$

$$= \frac{\langle \exp[-\beta\Psi(r)] \rangle}{\langle \exp[-\beta\Psi] \rangle}, \tag{2.62}$$

where we have noted that $P^{(1)} = P$ is independent of position in the uniform case, and $P^{(1)}(\mathbf{r}_2|\mathbf{r}_1)$ only depends on the separation r. We have replaced $\mathbf{s}_i$ with $\mathbf{r}_i$ to emphasise the similarities with (2.51), and on the final line we have used the definition (2.56) to show explicitly that $g^{(2)}(r)$ is a ratio of ensemble averages, as noted by Puibasset and Belloni [111]. As with (2.51), $g^{(2)}(r)$ is expressed as a ratio of a local ensemble average, depending on the separation from a reference particle, to a bulk ensemble average, but this result suggests a completely different route to measure $g^{(2)}(r)$.

The bulk ensemble average is the same as that used to calculate the chemical potential, and is evaluated as follows [63]. Imagine the hypothetical insertion of an additional particle at a random point in the fluid: the additional potential energy Ψ of adding the particle can be calculated providing the interactions of the fluid are known. This is used in turn to calculate the exponential, and the ensemble average is evaluated by averaging over insertions at different positions in the fluid and for different equilibrium configurations. To measure the local ensemble average, the procedure is repeated, except only insertions at a separation r from existing particles in the system are considered. The ratio of the ensemble averages then yields $g^{(2)}(r)$. Further details of the implementation are provided in later chapters.

2.5 Links with Thermodynamics

As discussed in Section 2.3, density functions and distribution functions enable calculation of certain properties in fluids without knowledge of the full configurational PDF P_{c}. For uniform, pairwise-additive systems, we show in Appendix A.1 that the average potential energy per particle is given by [116]

$$\frac{\langle U \rangle}{N} = \frac{\rho}{2} \int \mathrm{d}\mathbf{r}_{12} u^{(2)}(\mathbf{r}_{12}) g^{(2)}(\mathbf{r}_{12}), \tag{2.63}$$

where the relative position $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$. For the two-dimensional isotropic systems used in this thesis, we may integrate out the angular dependence to obtain [116]

$$\frac{\langle U \rangle}{N} = \rho\pi \int_0^\infty r u^{(2)}(r) g^{(2)}(r) \mathrm{d}r, \tag{2.64}$$

with $r = |\mathbf{r}_{12}|$. This is known as the *energy equation*, and allows the average potential energy per particle to be calculated from $g^{(2)}(r)$ if $u^{(2)}(r)$ is known, illustrating the power of the distribution function approach.

A more involved derivation (Appendix A.2) gives rise to an expression for the pressure

$$\beta p = \rho - \frac{\beta\rho^2}{s}\int \mathrm{d}\mathbf{r} r \frac{\mathrm{d}u^{(2)}(r)}{\mathrm{d}r} g^{(2)}(r), \tag{2.65}$$

where s is again the system dimensionality. In the two-dimensional isotropic case this yields the *pressure equation* [116]

$$\frac{\beta p}{\rho} = 1 - \frac{\beta\rho\pi}{2}\int_0^\infty r^2 \frac{\mathrm{d}u^{(2)}(r)}{\mathrm{d}r} g^{(2)}(r)\mathrm{d}r, \tag{2.66}$$

which may be used to calculate the pressure if $g^{(2)}(r)$ and $u^{(2)}(r)$ are known. This provides a second route to the thermodynamics of the fluid, and its use in hard-disk systems is central to Chapter 4.

Finally, we also note that the thermodynamics of the fluid may be calculated from $g^{(2)}(r)$ using the *compressibility equation*, which can be derived by considering number fluctuations in the grand ensemble and reads [1]

$$\chi_T = \frac{\beta}{\rho}\left[1 + \rho\int \mathrm{d}\mathbf{r}\left(g^{(2)}(\mathbf{r}) - 1\right)\right] \tag{2.67}$$

$$= \frac{\beta}{\rho}\left[1 + 2\pi\rho\int_0^\infty \left(g^{(2)}(r) - 1\right)\mathrm{d}r\right], \tag{2.68}$$

where on the second line we have again specialised the equation to two-dimensional fluids of isotropic particles. In principle, this equation allows access to the thermodynamics without knowing the interactions, and applies even in systems with many-body interactions [37]. In practice, however, this equation is difficult to implement due to slowly-decaying oscillations of $g^{(2)}(r)$: while the magnitude of these oscillations is small, any noise in the measurements at large separations is increasingly magnified by the factor r. By contrast, the integrands of the energy and pressure equations contain factors depending on $u^{(2)}(r)$, which typically decays much more quickly than $g^{(2)}(r)$ and effectively imposes a cut-off on the domain of the integral [124].

Chapter 3

Models and Methods

In this chapter, we introduce the physics of the colloidal model systems used in this thesis and the experimental and simulation methods used in later chapters. We begin with a formal description of how a complex colloidal fluid with several components (the particles, the solvent and various ions) can be mapped onto a one-component fluid with the particles interacting via an effective many-body interaction. We consider the nature of this interaction for the model systems at hand: in particular, the DLVO interaction, which explains the particles' stability against aggregation, and the dipolar interactions between superparamagnetic particles in an external field. We describe how, provided the out-of-plane fluctuations are small, monolayers of sedimented colloidal particles can be treated as *quasi-two-dimensional* fluids, and include a brief description of their dynamics. We then turn to the experimental methods, including sample preparation procedures, the magnetic-field calibration and video-microscopy. We describe the particle detection, including a customised algorithm to discriminate between the superparamagnetic particles and (brighter) bacteria. Finally, we detail the simulation methods used in this thesis, including the underlying principles, and the local and event-chain Monte Carlo algorithms.

The experimental data for the hard-disk colloidal model system were acquired during Part II of my MChem degree. The image analysis for this system and the writing of the event-chain implementation were performed during the same project. These aspects are detailed again here for completeness. In both cases, the subsequent test-particle insertion analysis to measure $g^{(2)}(r)$ is new to this thesis and is the subject of Chapter 4.

3.1 Colloidal Model Systems

3.1.1 Colloidal Interactions

In this thesis, we use colloidal systems as experimental models of one-component monatomic fluids. This colloid-atom analogy originates in the work of Einstein and Perrin in the early twentieth century [74, 84], was reawakened in the 1970s with the scattering measurements of Pusey, Brown and co-workers [146–148], and has more recently been exploited to study a broad range of phenomena in condensed matter physics [4–23]. In reality, these models are complex multicomponent systems composed of colloidal particles in a solvent phase, which may also include various ions, polymers, etc. These solvent degrees of freedom (DOFs) are inaccessible to optical microscopy, since the species are too small ($\sim$1 Å) and move too quickly (on a time scale $\sim$1 ps) [1], and only the positions of the colloidal particles (corresponding scales $\sim$1 µm, $\sim$1 µs) can be measured.[1] We therefore integrate out the solvent DOFs from the grand partition function, obtaining a marginal PDF for the colloidal configurations in a procedure known as coarse graining [1]. Providing the marginal PDF is of the form of P_c in (2.8), we can apply the results of the last chapter to the colloidal configurations as we would with atomic fluids.

Coarse Graining

Following Reference [1], which gives a similar derivation in the canonical ensemble, we assume a two-component system of colloidal particles with positions $\{\mathbf{R}_i\}$ and solvent molecules with positions $\{\mathbf{r}_i\}$, ignoring the polydispersity of the colloidal particles. The multicomponent grand partition function is

$$\Xi = \sum_{N_\mathrm{c}=0}^{\infty} \frac{{z_\mathrm{c}}^{N_\mathrm{c}}}{N_\mathrm{c}!} \sum_{N_\mathrm{s}=0}^{\infty} \frac{{z_\mathrm{s}}^{N_\mathrm{s}}}{N_\mathrm{s}!} \int \mathrm{d}\mathbf{R}^{N_\mathrm{c}} \int \mathrm{d}\mathbf{r}^{N_\mathrm{s}} \exp\left[-\beta\left(U_\mathrm{cc}(\mathbf{R}^{N_\mathrm{c}}) + U_\mathrm{cs}(\mathbf{R}^{N_\mathrm{c}}, \mathbf{r}^{N_\mathrm{s}}) + U_\mathrm{ss}(\mathbf{r}^{N_\mathrm{s}})\right)\right], \tag{3.1}$$

where N_c, N_s, z_c, and z_s are the numbers of particles and reduced activities of each component, respectively, and the potential energy is split into particle-particle interactions U_cc, particle-solvent interactions U_cs and solvent-solvent interactions U_ss. We next note that the partial sum

[1] Such a separation of length and time scales also presents challenges in theoretical work and simulation [1, 72, 73].

over only the solvent DOFs is just the grand partition function $\Xi_\mathrm{s}(\mathbf{R}^{N_\mathrm{c}})$ of a one-component fluid of solvent molecules in an external field due to the colloidal particles

$$\Xi = \sum_{N_\mathrm{c}=0}^{\infty} \frac{{z_\mathrm{c}}^{N_\mathrm{c}}}{N_\mathrm{c}!} \int \mathrm{d}\mathbf{R}^{N_\mathrm{c}} \exp\left[-\beta U_\mathrm{cc}(\mathbf{R}^{N_\mathrm{c}})\right] \times \left[\sum_{N_\mathrm{s}=0}^{\infty} \frac{{z_\mathrm{s}}^{N_\mathrm{s}}}{N_\mathrm{s}!} \int \mathrm{d}\mathbf{r}^{N_\mathrm{s}} \exp\left[-\beta\left(U_\mathrm{cs}(\mathbf{R}^{N_\mathrm{c}}, \mathbf{r}^{N_\mathrm{s}}) + U_\mathrm{ss}(\mathbf{r}^{N_\mathrm{s}})\right)\right]\right] \tag{3.2}$$

$$= \sum_{N_\mathrm{c}=0}^{\infty} \frac{{z_\mathrm{c}}^{N_\mathrm{c}}}{N_\mathrm{c}!} \int \mathrm{d}\mathbf{R}^{N_\mathrm{c}} \exp\left[-\beta U_\mathrm{cc}(\mathbf{R}^{N_\mathrm{c}})\right] \Xi_\mathrm{s}(\mathbf{R}^{N_\mathrm{c}}) \tag{3.3}$$

$$= \sum_{N_\mathrm{c}=0}^{\infty} \frac{{z_\mathrm{c}}^{N_\mathrm{c}}}{N_\mathrm{c}!} \int \mathrm{d}\mathbf{R}^{N_\mathrm{c}} \exp\left[-\beta\left(U_\mathrm{cc}(\mathbf{R}^{N_\mathrm{c}}) + \Omega_\mathrm{s}(\mathbf{R}^{N_\mathrm{c}})\right)\right], \tag{3.4}$$

where on the final line we have used $\Xi = \exp[-\beta\Omega]$ from (2.13).

In (3.4), Ξ resembles that of a one-component system (2.6) with the solvent grand free energy Ω_s acting as an additional interaction between the colloidal particles, and the corresponding marginal PDF is of the required form (2.8). As a free energy, the additional interaction is state-dependent and many-body in nature, and the effective colloid-colloid interaction is not pairwise-additive [1, 149]. The additional interaction can only be calculated exactly in certain idealised scenarios [150]—more typically, Density Functional Theory (DFT) is used to derive approximate interactions [1]. The central aim of this thesis is to develop methods to measure these interactions from particle coordinates obtained by optical microscopy.

Charge Stabilisation

In a vacuum, the direct interaction between colloidal spheres would be due to the van der Waals attractions between their constituent atoms, which give rise to an attractive pair potential [1]

$$u_\mathrm{vdW}^{(2)}(r) = -\frac{A}{12}\left[\frac{\sigma^2}{r^2-\sigma^2} + \frac{\sigma^2}{r^2} + 2\ln\left(1-\frac{\sigma^2}{r^2}\right)\right], \tag{3.5}$$

where A is the Hamaker constant. Because $u_\mathrm{vdW}^{(2)}(r)$ diverges when the particles are in contact ($r = \sigma^+$), the particles would be expected to aggregate in the absence of an additional repulsive interaction. In a polar solvent, this can arise due to the additional interaction Ω_s. In brief, the colloidal particles acquire a charge due to the dissociation of chemical groups on their surface, with the translational entropy of the dissociated ions in solution overcoming the energetic cost

of charge separation [38]. The compromise between these effects gives rise to a charge density profile around each particle, known as the *electric double layer*, which has an excess of counter-ions and a deficit of co-ions compared with the bulk fluid. When the colloidal particles and their double layers approach, the confinement of the ions between their surfaces is entropically disfavoured and leads to an effective repulsion [39].

A quantitative treatment was given by Derjaguin and Landau [151] and by Verwey and Overbeek [152], and is known as DLVO theory. Modern derivations [1, 153, 154] are based on integrating out the ionic degrees of freedom as in (3.4), and calculating the grand free energy of the ions using DFT. Several assumptions are made [154]: the ions are treated as point charges which experience the mean electric field due to the other ions and the particles, i.e. their correlations are ignored. The solvent molecules are not treated explicitly, instead the solvent is characterised by a relative permittivity ϵ. If the deviations of the ion concentrations from their bulk values are assumed to be small (linearised Poisson-Boltzmann theory) [1,154], the resulting effective interaction is pairwise additive,[2] with a pair potential

$$u_{\mathrm{DL}}^{(2)}(r) = \frac{Z^2 e^2}{4\pi\epsilon\epsilon_0}\left(\frac{\exp(\kappa_{\mathrm{D}}\sigma/2)}{1+\kappa_{\mathrm{D}}\sigma/2}\right)^2 \frac{\exp(-\kappa_{\mathrm{D}} r)}{r}, \tag{3.6}$$

where ϵ_0 is the electric constant, Ze is the surface charge of the particles, and ${\kappa_{\mathrm{D}}}^{-1}$ is the *Debye length*, defined such that

$$\kappa_{\mathrm{D}} = \sqrt{\frac{\beta e^2}{\epsilon\epsilon_0}\sum_{\nu}\rho_\nu z_\nu^2}, \tag{3.7}$$

where the sum is over the ionic species ν, with ρ_ν their bulk number densities in the fluid and $z_\nu e$ their charges.

Providing Z is large enough, $u_{\mathrm{DL}}^{(2)}(r)$ dominates over $u_{\mathrm{vdW}}^{(2)}(r)$, which can be neglected [156].[3] In this work, we use a highly polar water-ethanol solvent and, although we do not add additional salt, the concentration of ions is expected to be high, leading to a small ${\kappa_{\mathrm{D}}}^{-1}$. Hence, while important in stabilising the particles against aggregation, $u_{\mathrm{DL}}^{(2)}(r)$ is short-ranged enough that

[2]Note that the grand free energy also includes a state-dependent term independent of the colloidal particles' positions, which has an impact on the phase diagram [154, 155].

[3]Although $u_{\mathrm{vdW}}^{(2)}(r)$ diverges at $r = \sigma^+$, for large Z the energy barrier due to the double layer repulsion $u_{\mathrm{DL}}^{(2)}(r)$ is insurmountable and the colloidal particles are *kinetically* stabilised against aggregation.

the particles interact as if they were hard spheres, with a pair potential

$$u_{\mathrm{HS}}^{(2)}(r) = \begin{cases} \infty, & r \leq \sigma, \\ 0, & r > \sigma. \end{cases} \tag{3.8}$$

Superparamagnetic Particles

Superparamagnetic particles acquire a magnetic dipole in the presence of an external magnetic field. Assuming the spheres are uniformly magnetised, they behave as point magnetic dipoles at the centre of the sphere with a magnetic dipole moment [157, 158]

$$m = \frac{V_{\mathrm{p}} \xi B}{\mu_0}, \tag{3.9}$$

in the same direction as the applied field, where B is the magnetic flux density, ξ is the dimensionless volume susceptibility, μ_0 is the magnetic constant, and $V_{\mathrm{p}} = \pi\sigma^3/6$ is the particle volume. Taking the relative permeability of the solvent to be 1, the pairwise interaction between spheres is [158]

$$u_{\mathrm{mag}}^{(2)}(\mathbf{r}_{ij}) = \frac{\mu_0}{4\pi} \left[\frac{\mathbf{m}_i \cdot \mathbf{m}_j}{r_{ij}^3} - 3\frac{(\mathbf{m}_i \cdot \mathbf{r}_{ij})(\mathbf{m}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5} \right], \tag{3.10}$$

where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ is the vector between the centres of particles i and j, $r_{ij} = |\mathbf{r}_{ij}|$ is its magnitude, $\mathbf{m}_i$ is the magnetic dipole moment of particle i.

In this thesis, we use quasi-two dimensional colloidal monolayers, with an applied external magnetic field perpendicular to the sample plane, such that $\mathbf{r}_{ij}$ is perpendicular to both $\mathbf{m}_i$ and $\mathbf{m}_j$ and the second term vanishes. Assuming that the particles are identical and that the magnetic field due to the induced magnetic moments is negligible, the interactions are pairwise additive with a repulsive dipolar interaction

$$\frac{u_{\mathrm{mag}}^{(2)}(r)}{k_{\mathrm{B}}T} = \frac{\alpha}{r^3}, \tag{3.11}$$

where we have divided both sides by the thermal energy $k_{\mathrm{B}}T$ and

$$\alpha = \frac{\pi\sigma^6\xi^2 B^2}{144 k_{\mathrm{B}} T \mu_0}. \tag{3.12}$$

Note that we have dropped the subscripts indicating the particles' identities. The strength of the repulsion can be altered by varying B, and data for several different pair potentials can be acquired using a single sample. In the absence of a magnetic field we expect a hard-sphere interaction (3.8), giving a total pair potential

$$\frac{u^{(2)}(r)}{k_{\mathrm{B}}T} = \begin{cases} \infty, & r \leq \sigma, \\ \alpha/r^3, & r > \sigma. \end{cases} \tag{3.13}$$

3.1.2 Colloidal Monolayers

When the mass density of the colloidal particles and the solvent are different, the particles experience an external field due to gravity. The importance of gravity for the colloidal system may be quantified by considering the height at which the gravitational potential energy of a particle is equal to the thermal energy $k_{\mathrm{B}}T$—this *gravitational height* (or *gravitational length*) is given by [159]

$$h_g = \frac{6k_{\mathrm{B}}T}{\Delta\tilde{\rho}g\pi\sigma^3}, \tag{3.14}$$

where the buoyant mass density $\Delta\tilde{\rho} = \tilde{\rho}_{\mathrm{p}} - \tilde{\rho}_{\mathrm{s}}$, with $\tilde{\rho}_{\mathrm{p}}$ and $\tilde{\rho}_{\mathrm{s}}$ the mass densities of the particles and the solvent respectively. A suitable choice of solvent allows the mass densities to be matched, such that the particles experience no external field [13]; experiments have also been performed under conditions of microgravity in space [160].

In this thesis, we instead allow the particles to sediment onto the lower surface of the sample cell to form a colloidal monolayer. The out-of-plane fluctuations are characterised by h_g; providing they are small enough, the particles are essentially confined to the surface and move in two dimensions only, resulting in a *quasi-two-dimensional* system [15]. While the dimensionality of a system is important in determining, e.g. its phase behaviour [34], two-dimensional fluids are described by the statistical mechanical framework in Chapter 2 and display many of the same phenomena observed in three dimensions. Importantly, they allow for simpler experiments without the need for density matching and confocal microscopy. In particular, we note that the methods we develop for measuring distribution functions and interactions are based on the PDT and apply regardless of the system dimensionality.

While we have established that such colloidal model systems display analogous configurational statistical mechanics to atoms, we have not explicitly considered their dynamics. Newtonian dynamics would be appropriate in the full microscopic description, but as with the configurational DOFs, the separation of the colloidal and solvent length and time scales provides for a more insightful description with the solvent DOFs integrated out. The colloidal particles are continually bombarded by solvent molecules, and while collisions with a particle occur from all directions, at a given moment they will sum to move the particle in a random direction. Since the solvent molecules move rapidly compared with the colloidal particles, the correlation of the random force on a particle decays very quickly and the particle's velocity is essentially δ-correlated [3]: it is different at each moment in time, and the particle undergoes a mathematical random walk.

The characteristic timescale of this motion is the Brownian time, τ_{B}, which is the mean time taken for the centre of a single particle to move a distance equal to its diameter [161]. In three dimensions this is given by $\tau_{\mathrm{B}} = \sigma^2/6D$, with the diffusion coefficient given by the Einstein relation $D = k_{\mathrm{B}}T/\zeta$ [3]. The friction coefficient ζ for a sphere is derived from hydrodynamics as $\zeta = 3\pi\eta\sigma$ [3], where η is the viscosity of the solvent, yielding

$$\tau_{\mathrm{B}} = \frac{\pi\eta\sigma^3}{2k_{\mathrm{B}}T}. \tag{3.15}$$

The timescale of the particle diffusion in two dimensions will be different due to the reduced dimensionality and the presence of the lower internal surface of the cell, and also depends on the density of the monolayer [162]. The calculation according to (3.15) should therefore only be used as a rough guide to determine the order of magnitude of the timescale of the colloidal motion. The experiments in this thesis rely on the system reaching thermal equilibrium to satisfy P_{c} of (2.8), and for efficient ensemble averaging we ideally want to average over equilibrium configurations that are not significantly correlated with each other. The dynamics need to be fast enough to satisfy these criteria, but slow enough to be accessible to optical microscopy, i.e. $\sim$1 s.

3.2 Experimental Methods

3.2.1 Preparing Colloidal Samples

The procedures for preparing colloidal monolayers are well-established in the Oxford Colloid Group and are taken or adapted from the DPhil theses of previous students [131, 163–165].

In this thesis, we use two different model systems composed of colloidal spheres. In Chapter 4 we use a model system which has been previously shown to behave structurally as hard disks [15]: in this case the pressure equation (2.66) depends only on the value of $g^{(2)}(r)$ at contact, which we will show can be readily measured by test-particle insertion. In Chapter 5, where we develop a method to measure $u^{(2)}(r)$, we use a model system composed of superparamagnetic spheres since $u^{(2)}(r)$ can be tuned by varying the strength of the external magnetic field.

In both cases, a sample is prepared with bulk number density $\rho_V = N/Ah$, chosen to give rise to an appropriate monolayer density $\rho_A = N/Ah$, where N is the number of particles, A is the area of the lower internal surface of the cell and h is the internal cell height. A stock sample obtained from the manufacturer is diluted with a 20%/80% v/v ethanol/water mixture, which has mass density $\tilde{\rho} = 974\,\mathrm{kg\,m^{-3}}$ and viscosity $1.81 \times 10^{-3}\,\mathrm{Pa\,s}$ at 293 K [166]. The presence of ethanol inhibits bacterial growth and lowers the surface tension of the solvent, improving the wetting of the cell surface by the sample and assisting in filling the cell [131]. A custom-made quartz glass cell (Hellma Analytics) with $h = 200\,\mu\mathrm{m}$ and lateral dimensions $9\,\mu\mathrm{m} \times 20\,\mu\mathrm{m}$ is cleaned in a 2-3% Hellmanex® solution at $60 - 70\,^{\circ}\mathrm{C}$, and then flushed out successively with deionised water and ethanol and dried with compressed air. Subsequent experiments within a short time period neglect the Hellmanex® cleaning, but retain the rest of the cleaning procedure. Finally, the cell is plasma cleaned for approximately 60 s to remove organic contaminants [164]; this process also modifies the surface to make it more hydrophilic, which assists with the filling of the cell and reduces the number of particles which stick to the surface. The sample is then injected into the cell, which is subsequently sealed using Blu Tack.

The high $\Delta\tilde{\rho}$ of the particles leads to rapid sedimentation, and within a few minutes the particles are confined to the lower glass surface to give a quasi-two-dimensional system. Noting that N is the same in the bulk density ρ_V and the monolayer density ρ_A, we obtain the relation $\rho_A = h\rho_V$. In practice, however, the pre-determined ρ_A is approximate and in all cases we

measure ρ_A by considering the number of particles in the microscope's field of view. More precise control of the density can be achieved by slightly diluting a sample which gives rise to a known ρ_A.

Hard Disks

In Chapter 4, we use carboxylic acid-functionalised melamine formaldehyde (MF) spheres (Microparticles GmbH) with mass density $\tilde{\rho} = 1510\,\mathrm{kg\,m^{-3}}$, advertised diameter 2.70 µm and advertised polydispersity 2.2%. The surface carboxylic-acid groups dissociate in aqueous solution (pH ≈ 7) to give a negative surface charge, leading to a DLVO potential with a small ${\kappa_\mathrm{D}}^{-1}$ which approximates to $u^{(2)}_\mathrm{HS}(r)$ of (3.8). The effective hard-disk diameter of this model system was characterised as 2.79 µm by Thorneywork et al. [15], who measured $g^{(2)}(r)$ and compared with DFT calculations [167]. At 293 K, $h_\mathrm{g} = 68\,\mathrm{nm}$ or 2.4% of the particle diameter, and therefore the monolayer is quasi-two-dimensional. The dynamics are fast enough ($\tau_\mathrm{B} = 15\,\mathrm{s}$) that equilibrium is rapidly established. Experiments were performed with a range of monolayer densities ρ_A.

Superparamagnetic Particles

In Chapter 5, we use carboxylic-acid functionalised polystyrene (PS) spheres (Dynabeads® M-270 Carboxylic Acid, Invitrogen, Thermo Fisher Scientific), with mass density $\tilde{\rho} = 1600\,\mathrm{kg\,m^{-3}}$ and advertised diameter 2.80 µm [168]. Again, the particles are stabilised against aggregation by a negative surface charge. The particles contain evenly distributed single-domain nanoparticles of iron oxides [168] and have no overall magnetic dipole moment in the absence of an external field, where the domains of the individual nanoparticles have random orientations. In the presence of an external magnetic field, these domains all align with the field direction, giving rise to superparamagnetic behaviour (Section 3.1.1) [163,164]. In Chapter 5, we find that the particles have a diameter of 3.04 µm in the solvent, giving $h_\mathrm{g} = 45\,\mathrm{nm}$ (1.5% of the particle diameter) and $\tau_\mathrm{B} = 20\,\mathrm{s}$. The model system therefore forms an equilibrium quasi-2D fluid.

We first sort the particles to remove aggregates and small particles, thus reducing their polydispersity. We then sonicate (VWR Ultrasonic Cleaner) a sample of the stock solution for at least 15 min to break up aggregated particles. We allow the particles to sediment for 5 s, and then 15 s, in a centrifuge (Eppendorf Centrifuge 5810) at 800 rpm: after each stage, we keep

the supernatant and discard the sediment to remove large particles and aggregates. We then centrifuge the sample for 120 s three times. Each time, the supernatant is discarded, and the sediment collected and diluted with solvent, removing small particles and impurities from the samples. We then dilute the stock solution to concentrations corresponding to the desired ρ_A. Immediately before we inject the samples into the cell, any remaining aggregates are broken up by successively sonicating and mixing the samples using a vortex mixer (VWR VV3).

Following sedimentation, we apply an external magnetic field perpendicular to the sample plane using a solenoid, as described below. For samples at high density or in a sufficiently strong magnetic field, some particles begin to sit directly on top of others, since the dipolar attraction associated with this configuration overcomes the associated gravitational cost. We therefore increase the field strength gradually, allowing sufficient time for the particles to spread out due to the dipolar repulsion, which provides a kinetic energy barrier to preserve the quasi-2D nature of the colloidal fluid. Nevertheless, the particles may leave the sample plane slightly when they approach, which may have a bearing on the measured $u^{(2)}(r)$.

3.2.2 Magnetic Field

We apply a magnetic field perpendicular to the sample plane using a solenoid attached to the microscope. The sample is placed at the centre of the solenoid, where B is expected to be directly proportional to the current I and was measured using a Gaussmeter (GM07, Hirst Magnetic Instruments Ltd.). Currents ranging from 0 A to 1 A were considered in increments of 0.01 A; three measurements of B were taken at each current since the value on the Gaussmeter fluctuated. The instrument offset was corrected for by taking measurements with the probe in both upward-facing and downward-facing orientations. The results in Figure 3.1 show the expected linear relationship, and were fitted with a linear least-squares fit to complete the calibration.

3.2.3 Video-Microscopy

We image the samples using an inverted bright-field microscope (Olympus CKX41) fitted with a 40× air objective (Olympus LCACHN, NA = 0.55), a CMOS Camera (Ximea xiQ), and an LED for illumination. Note that the experiments are expected to be reasonably insensitive to

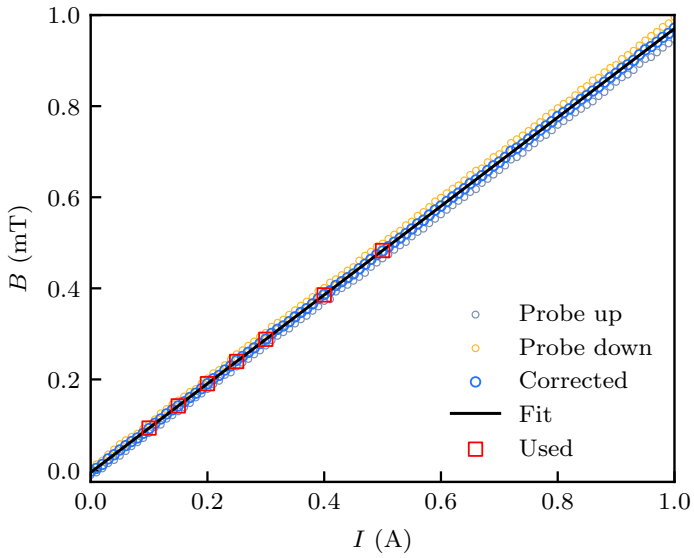


Figure 3.1: Calibration of the magnetic field using measurements taken both with the probe facing up and facing down, providing a corrected measurement. Also shown is the least-squares fit and the values used in the experiments in Chapter 5.

small amounts of particle drift due to the relatively short time series used. We carefully adjust the focal plane of the objective to optimise the images for particle tracking: a slight defocussing of the focal plane from the average plane of particle centres provides the necessary contrast, with the colloidal particles appearing as bright spots surrounded by a dark ring (see Figure 3.2) [36]. We capture images every second, and save them to the hard drive of the control computer as greyscale 8-bit TIFF files (2048 × 2048 pixels) pending further analysis.

3.2.4 Image Analysis

Part of a typical microscopy image of each model-system is shown in Figure 3.2. We use particle tracking routines based on Trackpy [169], a Python implementation of a standard particle tracking algorithm [24], which is detailed here. Particles are identified as local maxima (bright spots), and spurious detections are removed by imposing a minimum distance between neighbouring particles: when two maxima are detected within this distance, the brighter maximum is retained. The 'centre of mass' or *centroid* of the particle is then found using an iterative procedure.[4] For the superparamagnetic model system, some bacteria are found in the sample. While these are small and are likely to have only a small effect on the colloidal interactions,

[4]The brightness of each pixel is considered to be analogous to a mass density in this description.

they are very bright and would be detected preferentially over neighbouring particles. Unlike the particles, however, the bacteria are not surrounded by a dark ring, and so a customised preprocessing procedure is used to incorporate this information into the tracking.

Hard Disks: the Standard Algorithm

The hard disk model system is tracked using the standard algorithm, summarised in Figure 3.2. The original image is first preprocessed using a *bandpass filter*: pixel noise is smoothed by convolving the image with a gaussian filter with a length scale $\sim$1 pixel, and then a moving average over a length scale comparable to the diameter of the bright spot is subtracted to account for background variation. The filtered image is clipped according to a user-defined threshold, to give a series of gaussian-like blobs. Pixels are considered to be local maxima if they are brighter than all neighbouring pixels up to a separation consistent with the particle diameter: this detects the bright spots, but may also detect interstitial sites where regions of the background are brighter than the dark rings of nearby colloidal particles. These candidate positions are refined by taking the centroid of all pixels up to a distance corresponding to the radius of the bright spots, which is performed iteratively until convergence and leads to sub-pixel accuracy in the detection [24]. Finally, the peaks are characterised by their total brightness in the original images (i.e. the sum of their neighbouring pixel values), and this is used to remove spurious detections.

Superparamagnetic Particles: Incorporating the Dark Ring

This section largely follows the corresponding section in the Supplemental Material of Reference [97], on which Chapter 5 is based. © 2019 American Physical Society.

The superparamagnetic particles are tracked using a more sophisticated algorithm, also summarised in Figure 3.2, where the example image features a bacterium to illustrate how the procedure avoids its detection. The image is subjected to a bandpass filter (as above) and the filtered image is retained for use in the final refinement. For the detection steps, the image is adjusted to account for differences in the background, e.g. arising from dust on the microscope: each pixel is divided by the mean of the surrounding pixels within an appropriate cut-off distance. Two separate preprocessing procedures are then performed. The first accounts

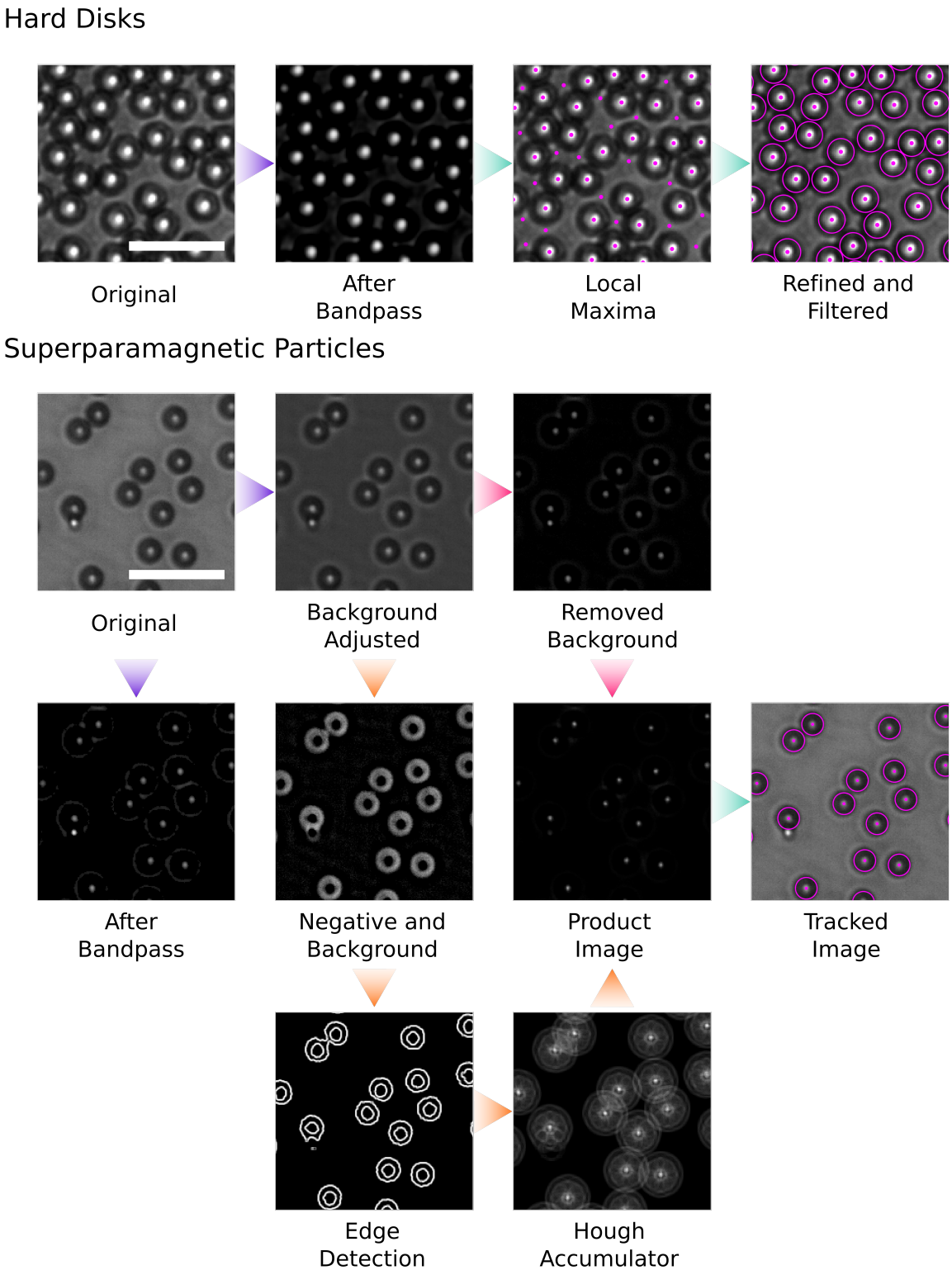


Figure 3.2: Particle Location. Hard Disks: images are preprocessed with a bandpass filter, then local maxima are detected. Their positions are refined and spurious detections are eliminated. Superparamagnetic particles: the image is adjusted and preprocessed to detect the bright spots (pink arrows) and dark rings (orange arrows). The local maxima of the product image are detected, and then refined based on the bandpass-filtered image. Detections corresponding to interstitial sites are eliminated (not shown). Scale bars are 10 µm.

for the bright spots, with the background of the image removed by clipping at the overall median pixel intensity. Note that from this image, the spot corresponding to the bacterium would be detected preferentially over its neighbouring particle since it is brighter. The second preprocessing procedure accounts for the dark rings. The image is first negated and then the background removed by clipping at the median intensity. It is then binarised, eroded by one pixel and its edges are detected using a Sobel filter[5]. A circular Hough transform with an appropriate radius is used to detect the outer edges of the dark rings; shown is the *accumulator*, where the bright spots correspond to the best candidates for the centres of the dark rings. This is then multiplied pixel-wise by the result of the first preprocessing step: in the product image, the bacteria are now significantly duller than their neighbouring colloidal particles. The local maxima of the product image are detected as above, and then refined based on the bandpass-filtered image. Finally, spurious particles corresponding to interstitial sites (which occur only rarely due to the low ρ_A) are removed using distance criteria, eliminating peaks with two or more neighbours within a preset distance.

3.3 Monte Carlo Simulations

Throughout this thesis, we apply our methods to particle coordinates generated from Monte Carlo (MC) simulations, in addition to those obtained from experiments. The simulation data may be important in their own right, such as those for hard disks, which are an important reference system in liquid state theory [1,167]. Their role in this thesis, however, is to provide a means of quickly generating particle coordinates for testing our methods for measuring interactions, with the advantage that the exact interaction potential is known, and the measurements can be checked. In Chapter 4, simulation data are used to measure $g^{(2)}(r)$ and find the hard-disk equation of state via the pressure equation (2.66), and in Chapter 5 they are used to verify the algorithm for measuring $u^{(2)}(r)$ before applying it in colloidal experiments. In Chapters 6 and 7, simulation data are used to demonstrate application of the methods in principle, and their application in colloidal experiments represents future work.

The principles and implementation of MC simulation are well described in many textbooks, e.g [63] and [171]; here we draw on both accounts, broadly following that of the latter (except

[5]The image processing operations, including the Sobel filter and the circular Hough transform, are performed using the Python library scikit-image [170].

for the event-chain MC scheme).

3.3.1 Importance Sampling and the Metropolis Algorithm

The ensemble average of a configurationally-dependent quantity $\mathcal{A}(\mathbf{r}^N; N)$ in the grand ensemble is given by

$$\langle \mathcal{A} \rangle = \mathrm{Tr}_\mathrm{c} \left[P_\mathrm{c} \mathcal{A} \right] . \tag{2.12}$$

Evaluating this quantity by direct integration is infeasible, since the domain of P_c has a very large number of dimensions, and P_c varies rapidly with changes in the particle coordinates [63]. Moreover, for configurations with strongly overlapping particles, the equilibrium PDF P_c is vanishingly small, and much effort would be wasted in calculating parts of the integral that make essentially zero contribution to the ensemble average [67]. Instead, importance sampling is used to generate configurations according to P_c: $\langle \mathcal{A} \rangle$ is then just a simple unweighted average over these configurations, with the weighting accounted for by the sampling algorithm.

The algorithm relies on setting up a Markov chain: a sequence of states, which here correspond to configurations of the system, with the property that each element of the sequence depends only on the state of its predecessor [171].[6] We construct the chain so that the probability distribution associated with a given element being in a particular state converges to the configurational PDF P_c. Once this *limiting distribution* is attained, $\langle \mathcal{A} \rangle$ may be estimated directly using the configurations associated with the elements of the chain.

To set up the chain, we consider the transition matrix $\mathbf{T}$, whose elements t_{mn} represent the probability that the next state in the sequence will be n if the current state is m.[7] These conditional probabilities must satisfy

$$\sum_n t_{mn} = 1, \tag{3.16}$$

[6] Each of the possible states of the Markov Chain should also be an element of a finite set [171], which in our case is the configuration space of the system (see Section 2.3). While technically this space is continuous in the particle coordinates r^N, it has been noted that the finite precision with which particle coordinates are stored means the system is effectively discretised [171].

[7] Note that for notational simplicity, in this section we specify configurations using the simple index n, rather than using the more precise specification $(N; \mathbf{r}^N)$.

where the sum runs over the elements of the configuration space of the system, and $\mathbf{T}$ is termed a stochastic matrix. For the limiting probability distribution to be $P_{\mathrm{c}}(N;\mathbf{r}^N)$, here denoted by the row vector $\mathbf{p}$ with elements p_n, $\mathbf{T}$ must satisfy the *global balance* condition

$$\mathbf{pT} = \mathbf{p}, \tag{3.17}$$

$$\sum_m p_m t_{mn} = p_n. \tag{3.18}$$

On the first line, the matrix product $\mathbf{pT}$ is the probability distribution of the next element of the chain, given that the current element has probability distribution $\mathbf{p}$. On the second line, the same equation in element notation shows explicitly how the transfer matrix elements t_{mn} are combined with the elements p_m of the current probability distribution, yielding the elements of the next probability distribution. Since $\mathbf{p}$ is the limiting distribution, the two successive probability distributions $\mathbf{p}$ and $\mathbf{pT}$ are equal—physically, this corresponds to a system in thermal equilibrium. To ensure that the full configuration space is sampled, $\mathbf{T}$ should also be *irreducible*, i.e. every configuration can be obtained from every other configuration given sufficient intervening steps of the chain, which is said to be *ergodic* [67]. Importantly, the properties of irreducible stochastic matrices ensure that the limiting distribution is independent of the initial state of the chain [64].

The global balance condition (3.18) can be satisfied by the stronger and simpler *detailed balance* condition [63],

$$p_m t_{mn} = p_n t_{nm}, \tag{3.19}$$

for all pairs of states $\{m, n\}$, which ensures that, once the limiting probability distribution is attained, the average number of moves from state m to n is equal to the average number of reverse moves. The off-diagonal elements of $\mathbf{T}$ are typically factorised as $t_{mn} = \alpha_{mn} s_{mn}$, where α_{mn} is the probability that a move from m to n is proposed in the algorithm and s_{mn} is the probability that it is accepted.[8] The ratio of the acceptance probabilities for the forward and reverse moves is then

$$\frac{s_{mn}}{s_{nm}} = \frac{\alpha_{nm}}{\alpha_{mn}} \frac{p_n}{p_m}, \tag{3.20}$$

[8]Note that this factorisation does not correspond to a matrix product, but an element-wise product.

which is satisfied with the criterion

$$s_{mn} = \min\left[1, \frac{\alpha_{nm}}{\alpha_{mn}}\frac{p_n}{p_m}\right]. \tag{3.21}$$

in the algorithm of Metropolis et al. [67]. Crucially, s_{mn} does not depend on the (generally unknown) grand partition function Ξ, since the factors of Ξ in (2.8) cancel in the ratio p_n/p_m. Finally, note that the diagonal elements of $\mathbf{T}$ correspond to the rejections of trial moves and are given by [171]

$$t_{mm} = 1 - \sum_{n\neq m} t_{mn}. \tag{3.22}$$

3.3.2 Local Monte Carlo

Canonical Ensemble

We begin with the canonical ensemble as a special case of the grand ensemble with a fixed number of particles [130], which is useful for simulations at a pre-determined density and for which the MC algorithm is particularly simple to implement. The configurational PDF for a system with $N^\star$ particles is obtained by multiplying P_c by the Kronecker delta $\delta_{NN^\star}$, constraining N to this fixed value. The ensemble average (2.12) is then

$$\langle \mathcal{A} \rangle = \frac{\mathrm{Tr}_\mathrm{c}\left[P_\mathrm{c}\delta_{NN^\star}\mathcal{A}\right]}{\mathrm{Tr}_\mathrm{c}\left[P_\mathrm{c}\delta_{NN^\star}\right]} \tag{3.23}$$

$$= \frac{\frac{z^{N^\star}}{N^\star!}\int \mathrm{d}\mathbf{r}^{N^\star} \exp\left[-\beta U_{N^\star}(\mathbf{r}^{N^\star})\right]\mathcal{A}(\mathbf{r}^{N^\star})}{\frac{z^{N^\star}}{N^\star!}\int \mathrm{d}\mathbf{r}^{N^\star} \exp\left[-\beta U_{N^\star}(\mathbf{r}^{N^\star})\right]} \tag{3.24}$$

$$= \frac{\int \mathrm{d}\mathbf{r}^{N^\star} \exp\left[-\beta U_{N^\star}(\mathbf{r}^{N^\star})\right]\mathcal{A}(\mathbf{r}^{N^\star})}{\int \mathrm{d}\mathbf{r}^{N^\star} \exp\left[-\beta U_{N^\star}(\mathbf{r}^{N^\star})\right]}, \tag{3.25}$$

and, dropping the superscript *, we infer that the canonical configurational PDF is

$$P_c^{(N)} = \frac{\exp[-\beta U_N(\mathbf{r}^N)]}{Q_N}, \tag{3.26}$$

where the canonical configurational integral

$$Q_N = \int \mathrm{d}\mathbf{r}^N \exp[-\beta U_N(\mathbf{r}^N)]. \tag{3.27}$$

This configurational PDF is independent of the chemical potential μ: the canonical ensemble is therefore specified by the thermodynamic variables (N, V, T) rather than (μ, V, T) as with the grand ensemble.

Following Metropolis et al. [67], the system is initiated by placing N particles in a simulation box in any configuration; often a regular lattice is used. A random particle is chosen, and a trial position chosen at random from a square centred on the original position. The side length of the square is an adjustable parameter given by 2δ, where δ is the maximum displacement of a particle in each direction. Periodic boundary conditions are typically employed to reduce finite-size effects and to avoid considering interactions with the walls of the container [64, 69]. The proposal probabilities of the forward and reverse moves are the same, i.e. $\alpha_{mn} = \alpha_{nm}$, and with the PDF (3.26), the acceptance criterion (3.21) becomes

$$s_{mn} = \min\left[1, \exp\left[-\beta(U_n - U_m)\right]\right]. \tag{3.28}$$

Particle movements are thus accepted automatically if they lead to a decrease in the total potential energy of the system, and are accepted with probability $\exp\left[-\beta(U_n - U_m)\right]$ if they lead to an increase.[9] For hard disks and hard spheres in the absence of external fields, the simple form of the pair potential (3.8) means that movements are rejected if they lead to overlap with another particle, and are accepted otherwise.

To optimise the efficiency of the algorithm, δ should be chosen carefully: if set too high, most proposed movements will lead to strong overlap with another particle and will be rejected, while if set too low, the configurations will not change much with each accepted move—in both cases, successive configurations will be highly correlated and it will take longer to reach equilibrium [67]. There are no formal criteria for selecting the optimum δ, but in practice it is often chosen so that the overall rate of acceptance is $\sim 50\%$ [64].

[9] That some moves which lead to an increase in the potential energy are accepted is important, since it ensures the ergodicity of the underlying Markov chain: considered with the protocol for proposing particle movements, it is clearly possible to reach any configuration from any other configuration in a finite number of moves.

Grand Ensemble

The full configuration space of the grand ensemble cannot be sampled using trial particle movements alone, and we must include additions and deletions of particles to allow N to vary. We first decide at random whether to attempt a particle movement, addition or deletion. In this thesis, we typically attempt particle movements, additions and deletions in the ratio 3:1:1; importantly, the probability of attempting an addition is the same as that for deletion [64]. The acceptance criterion for trial particle movements is the same as in the canonical case (3.28), as may be verified by substituting the grand ensemble PDF (2.8) into (3.21).

For trial additions, we attempt to add a particle at a random location in the simulation box, with probability density $1/V$, while for trial deletions, we attempt to delete a random particle from the system, with probability $1/N$. The proposal probability ratio is therefore $\alpha_{nm}/\alpha_{mn} = V/N + 1$, which when substituted with P_c into (3.21) yields [1]

$$s_{mn} = \min\left[1, \frac{V}{N+1}\frac{z^{N+1}\exp\left[-\beta U_n\right]}{z^N \exp\left[-\beta U_m\right]}\right] \tag{3.29}$$

$$= \min\left[1, \frac{Vz}{N+1}\exp\left[-\beta(U_n - U_m)\right]\right]. \tag{3.30}$$

For trial particle deletions, the proposal probability ratio is $\alpha_{nm}/\alpha_{mn} = N/V$, instead giving

$$s_{mn} = \min\left[1, \frac{N}{V}\frac{z^{N-1}\exp\left[-\beta U_n\right]}{z^N \exp\left[-\beta U_m\right]}\right] \tag{3.31}$$

$$= \min\left[1, \frac{N}{Vz}\exp\left[-\beta(U_n - U_m)\right]\right]. \tag{3.32}$$

In the cases of particle addition and deletion, we therefore consider the volume, number of particles and reduced activity, as well as the change in potential energy, when determining whether the trial move is accepted [63]. Note that the acceptance rate of insertion and deletion moves can be very low in dense systems, limiting their utility [64].

Multicomponent Systems

In Chapters 4 and 5, we perform simulations in multicomponent systems, and the extension of the algorithm is straightforward in both ensembles. In the canonical case, we simply add a step where we randomly decide which species to move. Movements of different species may be

attempted with differing frequencies and maximum displacements, without violating detailed balance.

When attempting additions and deletions of particles of a given species α, we simply replace z and N by the corresponding values for α in the acceptance criteria (3.30) and (3.32), yielding

$$s_{mn} = \min\left[1, \frac{Vz_\alpha}{N_\alpha + 1} \exp\left[-\beta(U_n - U_m)\right]\right] \tag{3.33}$$

and

$$s_{mn} = \min\left[1, \frac{N_\alpha}{Vz_\alpha} \exp\left[-\beta(U_n - U_m)\right]\right], \tag{3.34}$$

respectively. When deciding the proposal probabilities, it is critical that detailed balance is preserved. Although the rate of attempting additions and deletions for a particular species needs to be the same, there can be considerable differences between species if this assists in generated uncorrelated configurations [64].

3.3.3 Event-Chain Monte Carlo

For systems at high density, the local algorithms described above can be inefficient: a small δ is required for a reasonable acceptance rate, but leads to considerable correlations between the generated configurations, even after many moves [68, 172]. The acceptance rate of particle additions and deletions will also be very low, and alternative algorithms need to be sought. In recent years, workers in the group of Werner Krauth in Paris have developed rejection-free canonical MC algorithms [68, 173] which lead to much quicker decorrelations [174]. These algorithms have enabled larger simulations [175] and have been used to address the long-standing controversy of two-dimensional melting [176]. They also offer an alternative method to calculate the pressure from the simulation dynamics [173];[10] we therefore use the hard-disk implementation [68] in Chapter 4 and compare this measurement with that obtained by using the insertion measurement of $g^{(2)}(r)$ in the pressure equation (2.66). We provide a brief description of the hard-disk algorithm based on References [68] and [173] here, and refer the interested reader to these articles for full details.

[10] In common with other MC algorithms, the simulation dynamics are unphysical and do not correspond to the Newtonian dynamics of the classical fluid.

The simulation begins with an initial non-overlapping configuration at the desired density—in this work, the hard disks occupy the sites of a square lattice. A random disk is chosen and displaced in a random direction θ until it collides with another disk, which then takes over the displacement (moving in the same direction θ) until it collides with another disk, and so on [see Figure 3.3(a)]. The move is terminated when the total displacement of all disks in the chain is equal to a preset chain length l. Another random particle is then chosen, and the process is repeated. No moves are rejected, and this algorithm satisfies the detailed balance condition (3.19) [68]. In fact, θ can instead be drawn from an arbitrary PDF: this breaks detailed balance but the algorithm still satisfies the global balance condition (3.18), although at least two independent directions must be used for the simulation to remain ergodic [68,173]. The original authors note that their fastest implementation alternated between chains moving in the positive x and y directions [68]; in this work, we choose one of these directions at random for each chain. The chain length l is the only adjustable parameter and 'is typically chosen such that each chain consists of $\sim\sqrt{N}$ disks' [174]; we follow the initial authors in using $l \propto \sqrt{N}$ [177].[11] Further optimisation for each density is possible [175].

The pressure can be calculated directly from the simulation dynamics using [173]

$$\beta P = \rho \left\langle \frac{x_{\text{final}} - x_{\text{initial}}}{l} \right\rangle_{\text{chains}}, \tag{3.35}$$

where x_{initial} is the initial position of the first particle of the chain and x_{final} is the final position of the last particle, as shown in Figure 3.3(a). Care must be taken to account for the periodic boundary conditions, which is achieved by writing

$$x_{\text{final}} - x_{\text{initial}} = l + \sum_{\text{coll.}} \Delta x, \tag{3.36}$$

where Δx is the *excess displacement*, or the difference in the positions of the particle centres, along the chain direction, at collision, again shown in Figure 3.3(a) [173].

3.3.4 Cell Schemes

Calculating the required distances between the particles is very computationally intensive. For the local algorithms, the distances are needed to determine the potential energy differences in the

[11] We use $l = \sigma\sqrt{N}$, where σ is the hard-disk diameter.

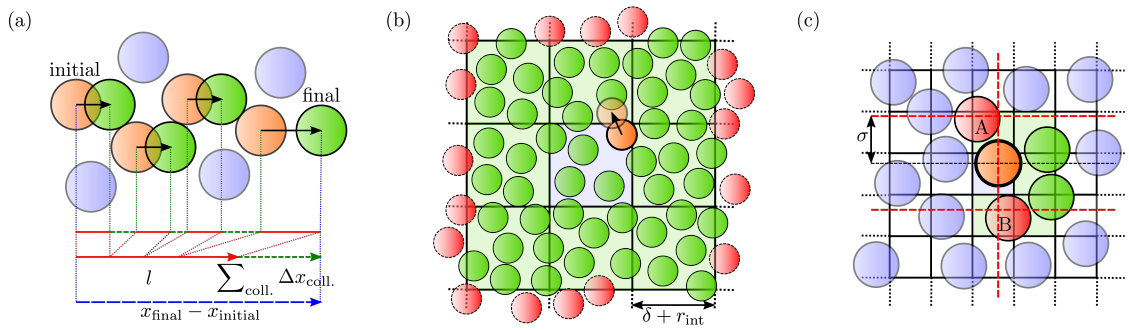


Figure 3.3: (a) The event chain algorithm. Adapted from J. Chem. Phys. **140**, 054116 (2014) [173], with the permission of AIP Publishing. An initial particle is displaced until it collides with another particle, which then takes up the motion and so on, up to a total chain length l (solid red lines). The excess displacements (dashed green lines) can be used in (3.36) to calculate the pressure. (b) The cell scheme used in the local algorithm. Only particles in neighbouring cells (solid green disks) need to be considered when deciding whether to accept a particle displacement; the same is true of particle additions and deletions. (c) The cell scheme used in the event-chain algorithm; each cell contains ∼1 particle. Only particles in six cells need to be considered, and some particles are removed from consideration immediately (red disks) since they are behind the moving particle (A), or too far away in the orthogonal direction (B). In this example, only collisions with the green particles need to be considered.

acceptance criteria, and in the event-chain case, they are needed to determine with which particle the moving particle will collide. To minimise the number of distance calculations required, a cell-scheme is used as shown in Figure 3.3(b)-3.3(c) (as detailed in [64] and [177]).

The simulation box is divided into square cells: for the simple Metropolis algorithm, the side length of the cell is the sum of the maximum displacement δ and the range of the interaction potential. For a given move (particle movement, addition or deletion), only interactions with particles in the neighbouring cells need to be considered when calculating the energy change, speeding up the calculation considerably [Figure 3.3(b)]. For the event chain algorithm, the side length is chosen so that there is an average of ∼1 particle per cell [177]. Particles in only 6 of the neighbouring cells need to be checked, depending on the chain direction [Figure 3.3(c)]. Particles *behind* the moving particle can be excluded from consideration, as can particles which are too far away in the orthogonal direction for a collision to occur.[12] In the case that the particle reaches the end of the neighbouring cell before colliding with another particle, it is first moved to the edge of that cell and then collisions are considered with particles in the new neighbouring

[12] A basic (and readable) Python program (without a cell scheme) for a hard-disk system of four particles can be found on Werner Krauth's website [178]. This was very useful as a starting point when implementing the algorithm.

cells.

Acknowledgements

Members of the Oxford Colloid Group, in particular Carla Fernández-Rico and François Lavergne, are thanked for practical assistance. Lachlan Alexander is thanked for useful discussions.

Chapter 4

Measuring the Contact Value Exactly in Hard-Disk Fluids

We use a test-particle insertion method based on the Potential Distribution Theorem to measure the pair distribution function at contact, $g^{(2)}(\sigma^+)$, in hard-disk fluids, which provides a direct link to the pressure. In contrast with the conventional distance-histogram method, no approximate extrapolation to contact is required. We use the insertion method to measure $g^{(2)}(r)$ in hard-disk simulations and colloidal experiments, and the equations of state (EOS) derived from the contact values agree well with the Scaled Particle Theory prediction up to high fluid packing fractions. For the simulation, this EOS agrees essentially perfectly with that measured using the event-chain dynamics, conclusively demonstrating the validity of the insertion approach in this region. We also provide the first experimental measurement of the cavity distribution function $y^{(2)}(r)$ *inside* the hard core, which will allow for a more complete comparison between colloidal model systems and liquid state theories. Finally, we further illustrate the method's flexibility by measuring the partial pair distribution functions of binary hard-disk mixtures in simulation.

The experimental data for this chapter were acquired, and the corresponding particle detection performed, during Part II of my MChem degree. The implementation of the event-chain simulation was written in the same project. This chapter is based on the following article [115]:

A. E. Stones, R. P. A. Dullens and D. G. A. L. Aarts. Communication: Contact values of pair distribution functions in colloidal hard disks by test-particle insertion. J. Chem. Phys. **148**, 241102 (2018).

4.1 Introduction

Systems of hard particles, where the only constraint is that the particles may not overlap with each other, are an important class of models in the condensed matter physics of continuous systems, playing a key role in the understanding of nematisation [82], crystallisation [65,81,179,180], two-dimensional melting [75, 176] and glass formation [181]. Since they introduce no additional energy scales beyond that associated with particle translation, which is already present in the perfect gas, their configurational statistical mechanics are temperature-independent and their phase behaviour highlights the role of entropy in ordering and self-assembly [182]. In liquid state physics, hard spheres (3D) and hard disks (2D) play a central role in the *van der Waals picture* [1, 130, 183] of fluids. In this picture, the fluid's structure is primarily determined by strong short-ranged interparticle repulsions, and the longer-ranged attractive forces give rise to a mean-field background potential, which provides the cohesive energy to hold the fluid together. This notion has led to various perturbation theories [1, 184, 185] and density functional theories (DFT) [1, 186], where hard spheres are used as a reference system whose properties may be calculated by Fundamental Measure Theory (FMT) [187–189] or molecular simulation [63, 64].

In Section 2.5, we saw that links between a fluid's structure and thermodynamics are provided by the compressibility equation (2.68) and, in the pairwise-additive case, the energy equation (2.64) and the pressure equation (2.66). For hard particles, the discontinuity in the pair potential leads to an especially simple form of the pressure equation, which for hard disks reads

$$\frac{\beta P}{\rho} = 1 + 2\phi g^{(2)}(\sigma^+), \tag{4.1}$$

where $\phi = \pi\sigma^2/4$ is the hard-disk packing fraction and σ is the hard-disk diameter. The quantity $g^{(2)}(\sigma^+)$ is the value of the pair distribution function at *contact*, i.e. the value of $g^{(2)}(r)$ as $r \to \sigma$ from above.[1] Knowledge of this *contact value* as a function of ϕ provides direct access to the hard-disk equation of state (EOS) [1, 116]; it is also important in theories of transport processes in gases [1, 190].

The conventional distance-histogram approach for measuring $g^{(2)}(r)$, based on (2.51), does

[1] A similar equation relates the pressure of hard sphere systems to the contact value of the one-body density $\rho^{(1)}$ at a hard wall [1].

not provide a direct measurement of the contact value, since space is discretised into bins of finite width. This implicitly assumes that $g^{(2)}(r)$ is slowly-varying and can be considered to be approximately constant over the width of a bin, an assumption which is clearly violated for hard particles at contact, where $g^{(2)}(r)$ is discontinuous. Instead, an approximate extrapolation is typically performed [15, 64, 176, 191–195] based on values of $g^{(2)}(r)$ measured away from contact, but this is somewhat speculative since the functional form of $g^{(2)}(r)$ in the contact region is generally unknown. We stress that while this failure of the distance-histogram method to measure contact values is particularly stark, it is actually symbolic of a broader failure of the distance-histogram method when the density or distribution function being measured varies rapidly. This will be seen in Chapter 6, where the distance-histogram method is used to measure $g^{(3)}$ in a system where $u^{(2)}(r)$ is continuous.

In this chapter, we will show that a test-particle insertion approach based on the Potential Distribution Theorem (PDT) can be used to measure $g^{(2)}(\sigma^+)$ without approximation in hard-disk systems. The method is based on Henderson's method [145, 196] for measuring the cavity distribution function $y^{(2)}(r)$, defined as

$$y^{(2)}(r) = g^{(2)}(r) \exp\left[\beta u^{(2)}(r)\right]. \tag{4.2}$$

While $g^{(2)}(r)$ and $u^{(2)}(r)$ are discontinuous at $r = \sigma$, the cavity distribution function remains continuous [1]. Previous use of this method has focussed on measuring bridge functions [104, 106–114], calculating Henry's law constants for dispheres [105] and applying zero separation theorems [103]; here we use it to obtain the contact value using $g^{(2)}(\sigma^+) = y^{(2)}(\sigma)$.

We will begin by specialising the PDT and the pressure equations to hard-disk systems, and then consider the advantages of the insertion approach in measuring $g^{(2)}(r)$. We will also discuss the physical meaning of $y^{(2)}(r)$ and the extension of these results to multicomponent systems. We then detail the numerical implementations of the distance-histogram and test-particle insertion methods, followed by the protocols used for the simulations and colloidal experiments. We obtain the contact values $g^{(2)}(\sigma^+)$ from the test-particle insertion measurements of $g^{(2)}(r)$, which we also compare to the distance-histogram results; we also provide the first experimental measurement of $y^{(2)}(r)$ inside the hard core. We compare the resulting equations of state to the Scaled Particle Theory (SPT) prediction [197, 198], finding excellent agreement throughout the fluid region. Moreover, the agreement of the simulation EOS derived from the contact values with

that calculated from the event-chain simulation dynamics using (3.35) is essentially perfect. Finally, we use the insertion approach to measure the partial pair distribution functions in a binary hard-disk simulation.

4.2 Theory

4.2.1 Hard Disks and the Potential Distribution Theorem

The potential energy U_N of a uniform, single-component hard-disk system is pairwise-additive (2.17), with the pair potential given by

$$u_{\mathrm{HS}}^{(2)}(r) = \begin{cases} \infty, & r \leq \sigma, \\ 0, & r > \sigma, \end{cases} \tag{3.8}$$

where σ is the hard-disk diameter. In this case (and its 3D counterpart), ensemble averages of the form $P = \langle \exp(-\beta\Psi) \rangle$, where Ψ is the change in U_N due to the insertion of a test particle, have a particular straightforward interpretation: $\Psi = \infty$ when the test particle overlaps with an existing particle, and $\Psi = 0$ in the absence of overlap. The exponential is then zero or unity, respectively, and the ensemble averages reduce to insertion probabilities [1]. The PDT for $g^{(2)}(r)$ (2.62) is simply

$$g^{(2)}(r) = \frac{P_{\mathrm{ins}}(r)}{P_{\mathrm{ins}}}. \tag{4.3}$$

To evaluate $g^{(2)}(r)$, we therefore need to find the probability of *successful* test-particle insertion at a separation r from an existing particle in the fluid, and then normalise it by the corresponding probability in the bulk fluid. Both probabilities can be evaluated using configurational snapshots at equilibrium.

4.2.2 Contact Values and the Pressure Equation

Recall that the pressure equation in a uniform system of isotropic particles is given in 2D by

$$\frac{\beta p}{\rho} = 1 - \frac{\beta\rho\pi}{2} \int_0^\infty r^2 \frac{\mathrm{d}u^{(2)}(r)}{\mathrm{d}r} g^{(2)}(r) \mathrm{d}r. \tag{2.66}$$

In the case of hard disks, the derivative of $u^{(2)}(r)$ is singular at $r = \sigma$, and so we instead rewrite (2.66) in terms of $y^{(2)}(r)$ [1],

$$\frac{\beta p}{\rho} = 1 - \frac{\beta\rho\pi}{2}\int_0^\infty r^2 \frac{\mathrm{d}u^{(2)}(r)}{\mathrm{d}r} \exp\left[-\beta u^{(2)}(r)\right] y^{(2)}(r)\mathrm{d}r \tag{4.4}$$

$$= 1 + \frac{\rho\pi}{2}\int_0^\infty r^2 \frac{\mathrm{d}}{\mathrm{d}r}\left\{\exp\left[-\beta u^{(2)}(r)\right]\right\} y^{(2)}(r)\mathrm{d}r, \tag{4.5}$$

where on the second line we have used the chain rule. With the hard-disk pair potential $u^{(2)}_{\mathrm{HS}}(r)$, the exponential is given by the Heaviside step function $\theta(r-\sigma)$, and its derivative is the Dirac delta $\delta(r-\sigma)$, yielding [1]

$$\frac{\beta p}{\rho} = 1 + \frac{\rho\pi}{2}\int_0^\infty r^2 \delta(r-\sigma) y^{(2)}(r)\mathrm{d}r \tag{4.6}$$

$$= 1 + \frac{\rho\pi\sigma^2}{2} y^{(2)}(\sigma) \tag{4.7}$$

$$= 1 + 2\phi g^{(2)}(\sigma^+). \tag{4.1}$$

On the final line we have noted that $y^{(2)}(r)$ and $g^{(2)}(r)$ are identical outside the core region, and thus $y^{(2)}(\sigma) = g^{(2)}(\sigma^+)$.

4.2.3 Advantages of the Insertion Approach

This section largely follows part of Reference [115] (with the permission of AIP Publishing), on which this chapter is based.

The conventional distance-histogram method for measuring $g^{(2)}(r)$ is based on directly applying [64]

$$g^{(2)}(r) = \frac{\rho^{(1)}(r)}{\rho}. \tag{2.51}$$

While measuring the bulk density $\rho = \langle N\rangle/V$ is straightforward, measuring $\rho^{(1)}(r)$ is more challenging: the average number of particles at an *exact* separation r from a reference particle is infinitesimally small, as is the corresponding volume element. Instead, the infinitesimal volume elements are replaced with small but finite volumes, in which the density can be measured. In this approximation, $\rho^{(1)}(r)$ is assumed to be slowly-varying and the measured $g^{(2)}(r)$ is in fact the convolution of the *true* $g^{(2)}(r)$ with a boxcar function depending on the size of these

volumes [15]. This assumption is reasonable in most cases, and the chosen volume sizes are a compromise: larger volumes result in reduced statistical noise but greater smoothing by the convolution and hence greater loss of structural information. Note that in low-density systems, very long experiments or simulations may be required to achieve acceptable resolution and statistics. This assumption is completely violated, however, for hard particles at contact, where the gradient of $g^{(2)}(r)$ is singular. Instead, $g^{(2)}(\sigma^+)$ must be determined by extrapolating values of $g^{(2)}(r)$ measured away from contact [15, 64, 176, 191–195].

By contrast, using the test-particle insertion approach based on (4.3), $g^{(2)}(\sigma^+)$ can be measured directly since insertions are attempted at precise separations from a reference particle, without introducing any approximations. Since an independent measurement is taken for each separation, the noise of the measured $g^{(2)}(r)$ depends only on the accuracy with which the insertion probabilities can be measured, allowing for a higher resolution measurement without incurring a noise penalty. Moreover, we will see that $y^{(2)}(r)$ can be measured at all separations, including those *inside* the hard core which are inaccessible to the distance-histogram method. The insertion method is especially advantageous at low densities, where good statistics are difficult to acquire in the distance-histogram approach but insertion probabilities are easily determined. At very high densities, the distance-histogram method is much less noisy, allowing for higher resolution, while measuring insertion probabilities is increasingly difficult [199].[2] In this sense, the two methods are complementary.

4.2.4 The Cavity Distribution Function

The cavity distribution function $y^{(2)}(r)$ can also be (and has been) measured directly by test-particle insertion [111, 196]. As noted in Section 4.1, for a general pair potential

$$y^{(2)}(r) = g^{(2)}(r) \exp\left[\beta u^{(2)}(r)\right] \tag{4.2}$$

$$= \frac{\langle \exp[-\beta\Psi(r)] \rangle}{\langle \exp[-\beta\Psi] \rangle} \exp\left[\beta u^{(2)}(r)\right], \tag{4.8}$$

where on the second line we have substituted (2.62). Consider the numerator of (4.8): $\Psi(r)$ is a sum of pairwise interactions between the test particle and the existing particles of the system,

[2] We will give a more complete explanation of the breakdown of the test-particle insertion approach at high densities in Section 6.4.1.

including the reference particle which we will label 1. Hence,

$$\Psi(r) = u^{(2)}(r) + \sum_{i=2}^{N} u^{(2)}(r_{i\mathrm{t}}), \tag{4.9}$$

where $r_{i\mathrm{t}}$ is the distance from the test particle t to particle i, and $r = r_{1\mathrm{t}}$. The pair potential $u^{(2)}(r)$ depends only on the separation r of the test and reference particles, and is independent of the positions of the other particles. It may therefore be removed from the ensemble average so that (4.8) becomes [111, 196]

$$y^{(2)}(r) = \frac{\langle \exp[-\beta\psi(r)] \rangle}{\langle \exp[-\beta\Psi] \rangle}, \tag{4.10}$$

where the terms depending on $u^{(2)}(r)$ have cancelled, and $\psi(r) = \Psi(r) - u^{(2)}(r)$ is the sum of interactions between the test particle and the existing particles, excluding the reference particle [see (4.10)]. This expression is very similar to the result (2.62) for $g^{(2)}(r)$, except that $\Psi(r)$ is replaced with $\psi(r)$, and so for hard disks [196]

$$y^{(2)}(r) = \frac{P^{\star}_{\mathrm{ins}}(r)}{P_{\mathrm{ins}}}, \tag{4.11}$$

where the superscript * indicates that the insertion probability should be calculated permitting overlap with the reference particle.

To understand the physical significance of $y^{(2)}(r)$, we first consider the definition of the distribution functions

$$g^{(n)}(\mathbf{r}^n) = \frac{\rho^{(n)}(\mathbf{r}^n)}{\prod_{i=1}^{n} \rho^{(1)}(\mathbf{r}_i)} \tag{2.43}$$

$$= \frac{z^n \exp\left[-\beta\Delta\Omega^{(n)}(\mathbf{r}^n)\right]}{\prod_{i=1}^{n} z \exp\left[-\beta\Delta\Omega^{(1)}(\mathbf{r}_i)\right]}, \tag{4.12}$$

where on the second line we have substituted the link (2.41) between the density functions and the grand free energy difference $\Delta\Omega^{(n)}(\mathbf{r}^n)$, which is associated with inserting n fixed particles at $\mathbf{r}^n$. In the uniform case with $n = 2$,

$$g^{(2)}(r) = \exp\left[-\beta\left(\Delta\Omega^{(2)}(r) - 2\Delta\Omega^{(1)}\right)\right], \tag{4.13}$$

where the factors z have cancelled and we have noted that, in a uniform system, $\Delta\Omega^{(2)}$ depends

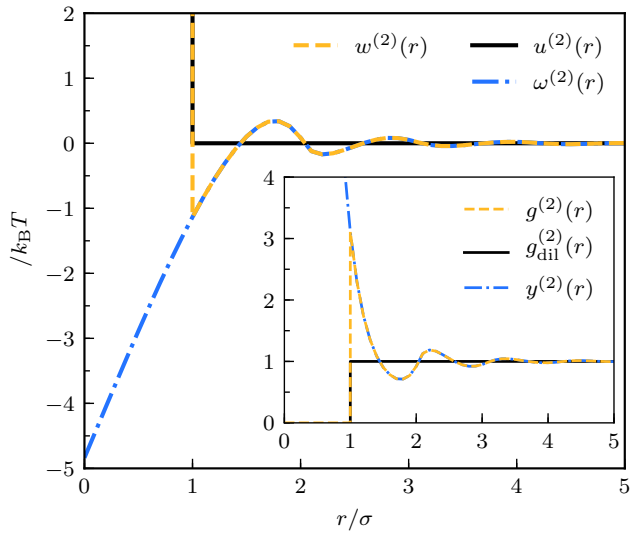


Figure 4.1: The potential of mean force $w^{(2)}(r)$ is the sum of the pair potential $u^{(2)}(r)$ and the indirect interaction $\omega^{(2)}(r)$, which arises due to presence of the other particles in the fluid. The measurements are for a hard-disk fluid with $\phi = 0.5$. Inset: Corresponding distribution functions: $g^{(2)}(r)$, $g^{(2)}_{\rm dil}(r)$ and $y^{(2)}(r)$. Note that $y^{(2)}(r)$ has a finite value at $r = 0$.

only on the separation r of the inserted particles and $\Delta\Omega^{(1)}$ is spatially-independent.

Considering large separations where $g^{(2)}(r) \to 1$, we can infer that $2\Delta\Omega^{(1)} = \Delta\Omega^{(2)}(\infty)$ is the grand free energy difference of inserting two particles at infinite separation. Hence, the grand free energy difference in (4.13) is the reversible work of bringing two particles from infinite separation up to a separation r [1], and is known as the *potential of mean force* $w^{(2)}(r)$, defined such that

$$g^{(2)}(r) = \exp\left[-\beta\left(\Delta\Omega^{(2)}(r) - \Delta\Omega^{(2)}(\infty)\right)\right] \tag{4.14}$$

$$= \exp\left[-\beta w^{(2)}(r)\right]. \tag{4.15}$$

In essence, $w^{(2)}(r)$ generalises the pair potential $u^{(2)}(r)$ by also accounting for the indirect interactions between two particles arising due to the presence of the other particles in the fluid. It may therefore be split into the direct interaction $u^{(2)}(r)$ and an indirect interaction $\omega^{(2)}(r)$ [1]:

$$g^{(2)}(r) = \exp\left[-\beta\left(u^{(2)}(r) + \omega^{(2)}(r)\right)\right], \tag{4.16}$$

$$y^{(2)}(r) = \exp\left[-\beta\omega^{(2)}(r)\right], \tag{4.17}$$

where on the final line we have used the definition of $y^{(2)}(r)$ in (4.2). This is illustrated in the

main plot of Figure 4.1, where $w^{(2)}(r)$ is shown as the sum of the pair potential $u^{(2)}(r)$ and the indirect interaction $\omega^{(2)}(r)$. At small separations $r \lesssim 1.8$, the overall effect of the other particles in this fluid is to push the pair closer together, and $\omega^{(2)}(r)$ is attractive.

The inset of Figure 4.1 shows the distribution functions which correspond to each of these interactions. In the absence of other particles, i.e. in infinite dilution, the pair distribution function is the simple Boltzmann inversion $g^{(2)}_{\text{dil}}(r) = \exp[-\beta u^{(2)}(r)]$ of the pair potential $u^{(2)}(r)$ [1]. At finite densities, additionally incorporating the indirect interactions $\omega^{(2)}(r)$ gives rise to the observed $g^{(2)}(r)$. The cavity distribution function $y^{(2)}(r)$ is analogous to $g^{(2)}(r)$, but excludes the direct interaction between the pair of particles, which is consistent with the modified insertion procedure given by (4.11). For hard particles, $y^{(2)}(r)$ is therefore the pair distribution function of *cavities* in the fluid where an additional particle could be accommodated [200]. The crucial difference with $g^{(2)}(r)$ is that, unlike the particles, such cavities can overlap and therefore $y^{(2)}(r)$ is continuous, with non-zero values even in the core region.

4.2.5 Multicomponent Systems

For multicomponent systems, the pair distribution function $g^{(2)}(r)$ generalises to a set of *partial pair distribution functions* $\{g^{(2)}_{\alpha\beta}(r)\}$ between particles of each pair of components $\{\alpha, \beta\}$.[3] The definitions of these functions and their relationships with the underlying density functions are analogous to the single-component case (see Section 2.3), and are generalised by using the multicomponent grand ensemble with fixed V and T, and a fixed chemical potential for each component, $\{\mu_\alpha\}$. Their interpretation and measurement is also analogous to the single-component case, and is summarised by the equations

$$g^{(2)}_{\alpha\beta}(r) = \frac{\rho^{(\beta)}_\alpha(r)}{\rho_\alpha} \tag{4.18}$$

and

$$g^{(2)}_{\alpha\beta}(r) = \frac{P^{(\beta)}_\alpha(r)}{P_\alpha}, \tag{4.19}$$

[3]Note that the higher order distribution functions $g^{(n)}$ generalise in a similar way: each of the n particles may be any one of the components of the system.

which generalise (2.51) and (2.62), respectively. For notational simplicity, we have dropped the superscript $^{(1)}$ on the conditional functions, instead indicating the identity of the reference particle.

From (4.18) we have the structural interpretation of $g^{(2)}_{\alpha\beta}(r)$ as the ratio of the density of α particles at a separation r from a reference β particle, $\rho^{(\beta)}_{\alpha}(r)$, to the density of α particles in the bulk fluid, ρ_α. It follows from the statistical mechanical derivation that $g^{(2)}_{\alpha\beta}(r) = g^{(2)}_{\beta\alpha}(r)$, and so the partial pair distribution function may also be interpreted in the same way with the roles of the α and β particles reversed. Using (4.19), $g^{(2)}_{\alpha\beta}(r)$ can also be considered a ratio of local and bulk ensemble averages, comparing the ensemble average $P^{(\beta)}_{\alpha}(r) = \langle \exp[-\beta\Psi_\alpha(r)]\rangle^{(\beta)}$, associated with inserting an α particle at a separation r from a reference β particle, with the corresponding ensemble average $P_\alpha = \langle \exp[-\beta\Psi_\alpha]\rangle$ in the bulk fluid. Note that in calculating Ψ_α, interactions of the inserted α particle with particles of *all* components are considered. For multicomponent hard-disk systems, the ensemble averages again reduce to insertion probabilities.

The pressure equation (2.66) may also be derived in the multicomponent case, and for hard disks is given by [201]

$$\beta P = \sum_{\alpha} \rho_\alpha + \frac{\pi}{2} \sum_{\alpha,\beta} \rho_\alpha \rho_\beta \sigma^2_{\alpha\beta} g^{(2)}_{\alpha\beta}(\sigma^+_{\alpha\beta}), \tag{4.20}$$

where the sums run over the components of the system and $\sigma^+_{\alpha\beta} = (\sigma_\alpha + \sigma_\beta)/2$ is the contact separation of two hard disks with diameters σ_α and σ_β. This provides the link between the pressure of the multicomponent system and the contact values of the partial pair distribution functions, and reduces to (4.1) in the single-component case. Again, (4.19) provides for an insertion method to measure $g^{(2)}_{\alpha\beta}(\sigma^+_{\alpha\beta})$ exactly.

The *partial cavity distribution functions* are defined by analogy with (4.2):

$$y^{(2)}_{\alpha\beta}(r) = g^{(2)}_{\alpha\beta}(r) \exp\left[\beta u^{(2)}_{\alpha\beta}(r)\right], \tag{4.21}$$

and may also be measured using test-particle insertion

$$y^{(2)}_{\alpha\beta}(r) = \frac{P^{\star(\beta)}_{\alpha}(r)}{P_\alpha}, \tag{4.22}$$

with the superscript * again indicating that the insertion probability is to be calculated without

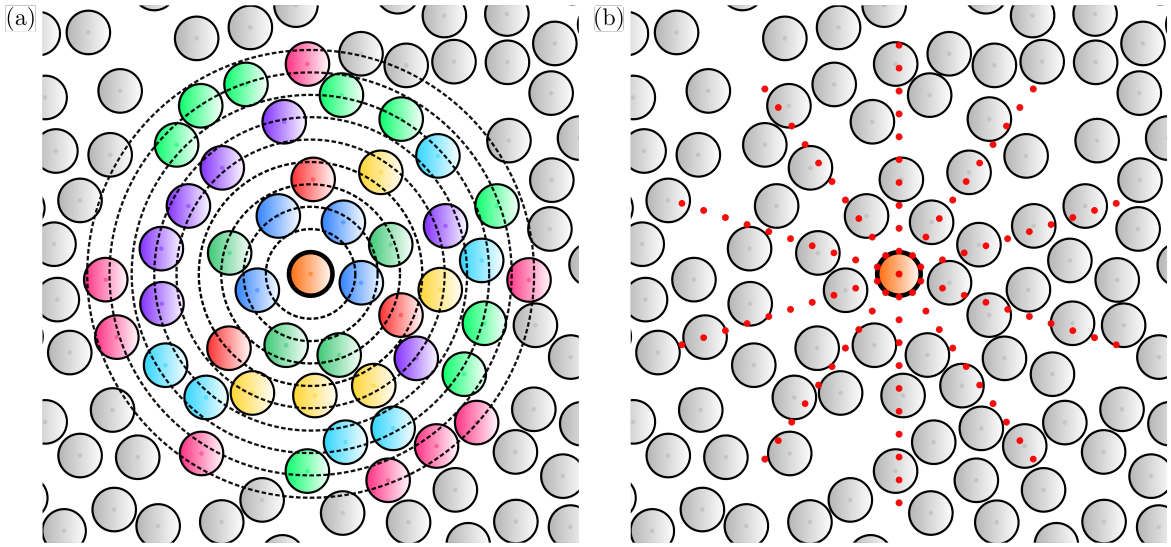


Figure 4.2: Cartoons illustrating the implementation of (a) the distance-histogram method and (b) the insertion method. For the distance-histogram method, particles are grouped by their separation r from the reference particle (bold, orange), up to a cut-off distance. The histogram-bin edges are shown as dashed lines, and the particles are coloured according to which bin they are in. For the insertion method, test-particle insertions are attempted at a range of separations from the reference particle; in this example, 10 insertions are attempted for each separation (red dots). In measuring $y^{(2)}(r)$, overlap of the test-particles with the reference particle is *not* considered. For both methods, the results are averaged over different reference particles and different fluid configurations.

considering overlap with the reference β particle.

4.3 Methods

4.3.1 Implementation

We implement the distance-histogram and insertion methods simultaneously to analyse configurational snapshots from simulations and experiments. We consider each particle in turn to be the reference particle: firstly, we calculate the relative coordinates of neighbouring particles (relative to the reference particle) and use them to construct a k-d tree [202, 203].[4] For the distance-histogram method, we find all particles within a cut-off distance, corresponding to the maximum r for which $g^{(2)}(r)$ is to be measured, and add them to a histogram based on their

[4] A k-d tree [202, 203] is an optimised data structure which allows for fast calculations of the distances to the nearest neighbouring particles. They are used throughout this thesis. We use the implementation in SciPy [204], which is based on an algorithm described by Maneewongvatana and Mount [205].

separation r from the reference particle [Figure 4.2(a)]. For the insertion method, we attempt test-particle insertions at a range of separations and for a number of different angles, taking care to exclude from consideration any overlap with the reference particle [Figure 4.2(b)]. For the experimental system where there are no periodic boundary conditions, we only consider those particles sufficiently far from the edge of the image as reference particles.

We then collate the results from all reference particles in all snapshots to measure the local density $\rho^{(1)}(r)$ and the local insertion probability $P^{\star}_{\text{ins}}(r)$. The bulk density ρ is known from the simulation input or measured separately, and used with $\rho^{(1)}(r)$ in (2.51) to obtain the distance-histogram measurement of $g^{(2)}(r)$. We perform a separate measurement of the bulk insertion probability P_{ins} by attempting test-particle insertions on a regular grid for each snapshot, and use it with $P^{\star}_{\text{ins}}(r)$ in (4.11) to calculate $y^{(2)}(r)$. We can then obtain the insertion measurement of $g^{(2)}(r)$, including the contact value $g^{(2)}(\sigma^{+})$, using (4.2).

An analogous analysis can be performed for multicomponent systems. In this case, we must measure the local densities and insertion probabilities of each component around reference particles of each component, and compare them with the corresponding bulk values. In the binary case, we therefore measure four local densities and four local insertion probabilities, as well as two bulk densities and two bulk insertion probabilities.

4.3.2 Protocols

The protocols of this section follow those in Reference [115] (with the permission of AIP Publishing), on which this chapter is based.

Simulation

For the single-component simulations, we used the Event-Chain Monte Carlo (ECMC) algorithm presented in Section 3.3.3, since this allows us to compare the EOS obtained from $g^{(2)}(\sigma^{+})$ using (4.1) to the value measured from the simulation dynamics using (3.35) [173].[5] In all cases, we used 1024 disks and adjusted the box-size to give the desired ϕ. After equilibration, we took

[5] In this chapter we perform simulations in the canonical ensemble. While the PDT was derived in the grand ensemble in Section 2.4, it may also be derived in the thermodynamic limit of the canonical ensemble [100]. While the simulations will include small finite-size effects, leading to an error $\sim 1/N$ in the measured $g^{(2)}(r)$ [1], these are not important for the conclusions of this chapter.

5×10^3 snapshots at intervals of 10^2 chains.

For the binary system, we used the local canonical ensemble algorithm of Section 3.3.2 to simulate 1000 disks of each diameter, with a diameter ratio $\sigma_\mathrm{l}/\sigma_\mathrm{s} = 4/3$. The box length was $50\sigma_\mathrm{l}$, giving a total packing fraction $\phi_\mathrm{total} = 0.49$. After 10^5 equilibration moves, we took 10^4 snapshots at intervals of 10^4 moves.

Experiment

We used the hard disk model system, preparing samples at a range of monolayer densities using the protocol discussed in Section 3.2.1. We followed the system using optical microscopy (Section 3.2.3), taking images every second for over half an hour. We tracked the particles using the standard algorithm of Section 3.2.4.

Analysis

For the single-component simulation data, we attempted insertions at 10 angles for r up to 10σ, with intervals of 0.1σ. We also performed a separate measurement of the bulk insertion probability in each case. For the distance-histogram method, the histogram bin widths were also 0.1σ, and we chose the bin locations such that $r = \sigma$ is at a bin edge. For the $\phi = 0.35$ simulation only, we performed an additional measurement with the bin locations chosen such that $r = \sigma$ is at a bin midpoint.

For the binary simulation data, we used 20 angles at radial distances separated by $0.1\sigma_{\alpha\beta}$, depending on the partial pair distribution function being measured. We also made a separate bulk insertion measurement for each component. We used the same interval for the histogram bins, which were located such that $r = \sigma_{\alpha\beta}$ is at a bin edge.

For the experimental data, we used 10 angles at radial distances separated by 0.05σ, where $\sigma = 2.79\,\mu\mathrm{m}$ was determined by Thorneywork et al. in Reference [15], who compared $g^{(2)}(r)$ measurements with FMT predictions [167]. Again, we also performed a separate measurement of the bulk insertion probability and the same interval was used for the histogram bins, which were located such that $r = \sigma$ is at a bin edge.

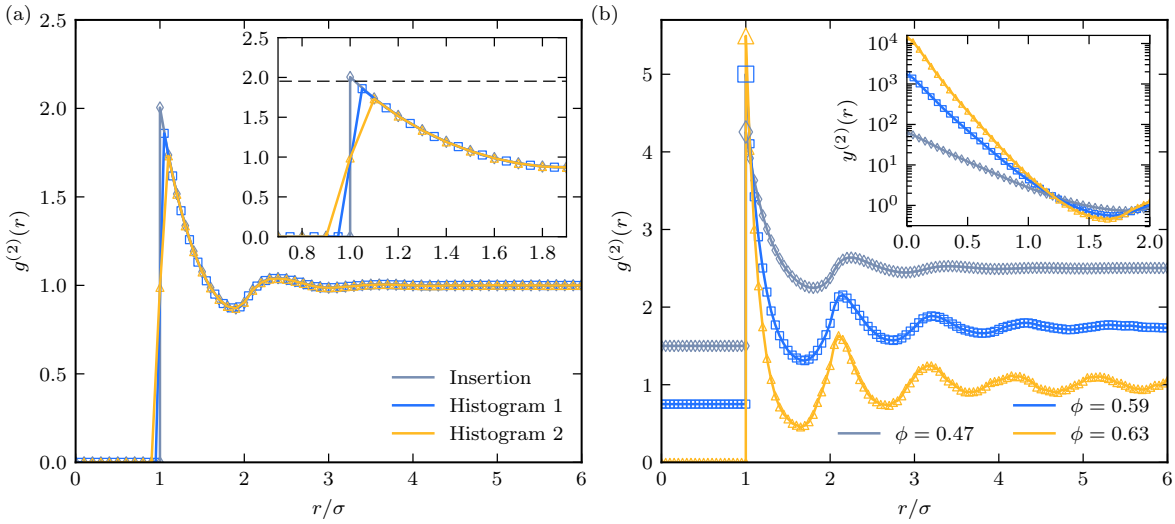


Figure 4.3: (a) Comparison of the $g^{(2)}(r)$ obtained in *simulation* ($\phi = 0.35$) by test-particle insertion (diamonds) and the distance-histogram method with two different bin locations (squares, triangles). Inset: zoom of the contact region; the dashed line is the SPT prediction of the contact value. (b) $g^{(2)}(r)$ measured using insertion in *experiment* at three different packing fractions; for clarity the results for $\phi = 0.59$ and $\phi = 0.47$ have been shifted by 0.75 and 1.50, respectively, and the contact values are displayed with larger symbols. Inset: corresponding $y^{(2)}(r)$ on a semi-log plot. Reproduced from [115], with the permission of AIP Publishing.

4.4 Results and Discussion

This section largely reproduces the corresponding section of Reference [115] (with the permission of AIP Publishing), on which this chapter is based.

4.4.1 Single-Component Hard Disks

Figure 4.3(a) shows $g^{(2)}(r)$ measured in a simulation of packing fraction $\phi = 0.35$ using both the insertion method and the distance-histogram method. We performed the distance-histogram method with two different sets of bin locations, in both cases failing to measure the contact value $g^{(2)}(\sigma^+)$. When $r = \sigma$ is at a bin midpoint, the measured $g^{(2)}(\sigma^+)$ is a significant underestimate since the bin overlaps with the core region ($r < \sigma$), where there are strictly no other particles due to the form of $u^{(2)}_{\mathrm{HS}}$. In the case that $r = \sigma$ is at a bin edge, although the measured $g^{(2)}(r)$ for the bin to the right of contact agrees well with the insertion method, the midpoint of the bin is not at $r = \sigma$ and an approximate extrapolation is required to estimate $g^{(2)}(\sigma^+)$ [64]. By contrast, the insertion method enables $g^{(2)}(\sigma^+)$ to be measured without any approximations,

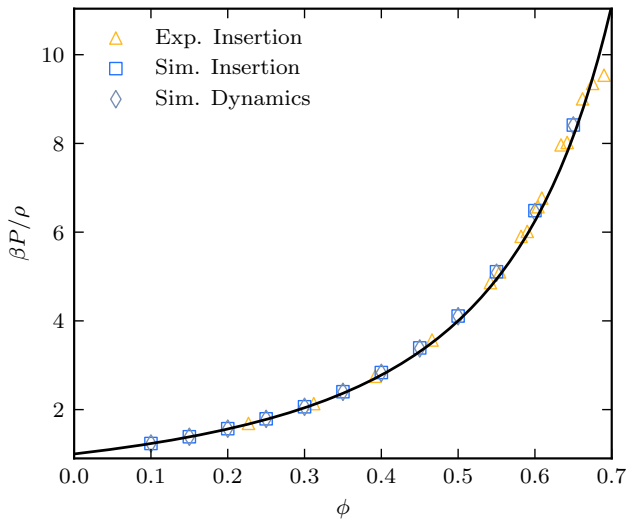


Figure 4.4: Compressibility factor $Z = \beta P/\rho$ calculated using the single-component contact values $g^{(2)}(\sigma^+)$ measured in experiment and simulation. For comparison, the values from the simulation dynamics are shown, as is the SPT equation of state (black line). Reproduced from [115], with the permission of AIP Publishing.

and the measured value agrees well with the SPT prediction [198] (dashed line), which is known to work well at this packing fraction [15]. Away from the contact region, the three analyses give essentially identical results, showing that the effect of the convolution is considerably less pronounced in regions where $g^{(2)}(r)$ is slowly-varying.

Figure 4.3(b) shows insertion measurements of $g^{(2)}(r)$ at three experimental packing fractions. Again, the discontinuity at contact is well-captured by construction, and $g^{(2)}(r)$ displays the typical damped oscillatory decay. The inset shows corresponding $y^{(2)}(r)$, including measurements inside the core (note again that $g^{(2)}(r) = y^{(2)}(r)$ outside the core). While $y^{(2)}(r)$ inside the core has been measured in simulations (using Henderson's method [107] or an alternative MC method [206]), this first experimental measurement opens the possibility of comparing results from experimental model systems and liquid state theories [167, 207] and thus evaluating their reliability. Furthermore, $y^{(2)}(r)$ is linked directly to thermodynamics via zero separation theorems [103, 208, 209] and $y^{(2)}(r)$ for hard particles plays a central role in certain perturbation theories [184, 185, 206].

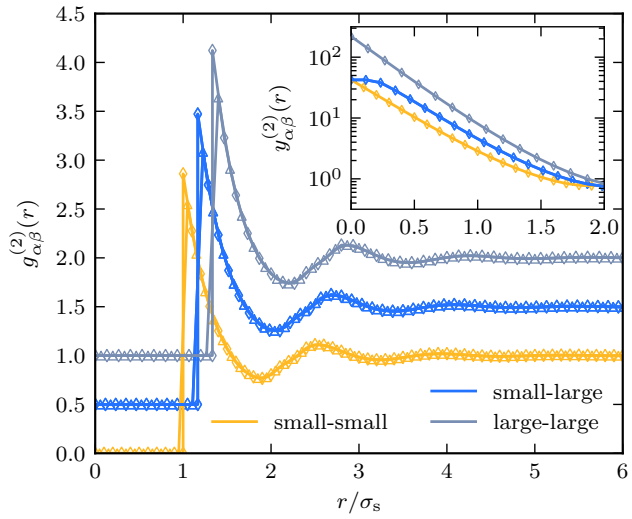


Figure 4.5: Binary hard-disk partial pair distribution functions $g^{(2)}_{\mathrm{ss}}(r)$, $g^{(2)}_{\mathrm{sl}}(r) = g^{(2)}_{\mathrm{ls}}(r)$ and $g^{(2)}_{\mathrm{ll}}(r)$ measured by insertion (diamonds) and the distance-histogram method (triangles) in simulation. The total packing fraction $\phi_{\mathrm{total}} = 0.49$, with an equal number of large and small disks with size ratio $\sigma_{\mathrm{l}}/\sigma_{\mathrm{s}} = 4/3$. For clarity, $g^{(2)}_{\mathrm{sl}}(r)$ and $g^{(2)}_{\mathrm{ll}}(r)$ have been shifted by 0.5 and 1.0, respectively. Inset: the corresponding partial cavity distribution functions on a semi-log plot. Reproduced from [115], with the permission of AIP Publishing.

4.4.2 Equation of State

The measured $g^{(2)}(\sigma^+)$ are then used in (4.1) to calculate the compressibility factor $Z = \beta P/\rho$ and hence the EOS, shown in Figure 4.4. The results from simulation and experiments agree well with the SPT prediction [198] up to high fluid packing fractions—Thorneywork et al. found the fluid-hexatic coexistence region to be at $\phi \approx 0.68 - 0.70$ in the same system [75]. This further confirms the validity of this system as an experimental model for hard disks, which has previously been shown by Thorneywork et al. who have measured the EOS using two independent routes [15, 75]. Although the SPT prediction is known to deviate from the simulation results at higher ϕ [210], essentially perfect agreement is found between the simulation EOS based on the measured $g^{(2)}(\sigma^+)$ and that calculated from the simulation dynamics using (3.35), conclusively demonstrating the validity of the insertion route.

4.4.3 Binary Hard Disks

Figure 4.5 shows the partial pair distribution functions measured from the binary hard-disk simulation using each method, with the insertion measurements of the partial cavity distribution

functions shown in the inset. The correct behaviour in the contact region is again captured by the insertion method, and using the contact values in (4.20) provides a direct link to the pressure of the binary fluid, which will allow for a more complete thermodynamic study of hard-disk and hard-sphere mixtures. In contrast to the distance-histogram method, the equality $g_{\mathrm{sl}}^{(2)}(r) = g_{\mathrm{ls}}^{(2)}(r)$ predicted by statistical mechanics is not strictly enforced by the insertion method [103], where the two functions are separately measured by considering the insertion of large particles about small particles and vice versa. Although numerically minor differences are observed due to noise, both results are plotted in Figure 4.5 where their differences are not distinguishable. The convergence of these functions indicates that enough data have been acquired to provide reliable statistics.

The measured partial cavity distribution functions display some noteworthy features which can be understood using (4.22) [211]. The coincidence of $y_{\mathrm{ss}}^{(2)}(r)$ and $y_{\mathrm{sl}}^{(2)}(r) = y_{\mathrm{ls}}^{(2)}(r)$ at $r = 0$ is understood by noting that $P_{\mathrm{s}}^{\star(\mathrm{s})}(0) = P_{\mathrm{s}}^{\star(\mathrm{l})}(0) = 1$, since in both cases the fixed particle prevents other particles from overlapping with the region where the test particle is inserted. From (4.22) it follows that $y_{\mathrm{ss}}^{(2)}(0) = y_{\mathrm{sl}}^{(2)}(0) = 1/P_{\mathrm{s}}$. Moreover, since in the case of a large fixed particle, $P_{\mathrm{s}}^{\star(\mathrm{l})}(r) = 1$ for $0 \leq r \leq (\sigma_{\mathrm{l}} - \sigma_{\mathrm{s}})/2$, the value of $y_{\mathrm{sl}}^{(2)}(r)$ is constant in this region. Note further that while $P_{\mathrm{l}}^{\star(\mathrm{s})}(r) < 1$ over this range, the equality $y_{\mathrm{sl}}^{(2)}(r) = y_{\mathrm{ls}}^{(2)}(r)$ is maintained since the bulk insertion probability for large particles is less than that for small particles (i.e. $P_{\mathrm{l}} < P_{\mathrm{s}}$).

4.5 Conclusion and Outlook

We have applied Henderson's method [196] for measuring $y^{(2)}(r)$ to provide a test-particle insertion measurement of $g^{(2)}(r)$ in hard-disk simulations and colloidal experiments. In contrast with the conventional distance-histogram method, the contact value $g^{(2)}(\sigma^{+})$ can be measured exactly and without an approximate extrapolation. We have used the measured contact values to obtain the hard-disk EOS, finding excellent agreement with SPT for both the simulations and experiments and further confirming the validity of the colloidal system as a model for hard disks. The simulation EOS agrees perfectly with that measured using the simulation dynamics, conclusively demonstrating the validity of the insertion approach. We have also obtained the first experimental measurement of $y^{(2)}(r)$ *inside* the hard core, which will enable further comparisons between liquid state theory and experiments. Finally, we have shown that the method

can also be used to measure partial pair distribution functions in a binary hard-disk simulation.

The methods of this chapter readily extend to higher dimensions, and will enable a more complete and accurate thermodynamic study of hard-sphere mixtures. In principle, they may also be extended to investigate more complex hard-particle fluids, such as those composed of hard rods or of molecules made of fused hard spheres [171]. The insertion result (2.62) is general and applies equally well in systems with continuous interparticle interactions. While the interactions are generally unknown in colloidal fluids, combining the insertion and distance-histogram methods provides an elegant route to measure the pair potential $u^{(2)}(r)$, which we will develop in the next chapter. Finally, in Chapter 6, we show that the insertion method can be extended to measure many-body distribution functions $g^{(n)}$ (with $n > 2$), where the sub-optimal resolution of the distance-histogram method is still more pronounced.

Acknowledgements

Members of the Oxford Colloid Group, especially François Lavergne, are thanked for practical assistance. Bob Evans, Alice Thorneywork and Andrew Parry are thanked for useful discussions.

Chapter 5

Measuring the Pair Potential from Equilibrium Configurations

We develop a model-free measurement of the pair potential $u^{(2)}(r)$ from equilibrium configurational data, which is based on enforcing consistency between the distance-histogram and insertion measurements of $g^{(2)}(r)$. We first demonstrate the accuracy and versatility of the method in simulations of simple fluids, recovering the input $u^{(2)}(r)$ in all cases. We then apply the method experimentally in the superparamagnetic colloidal model system, obtaining the anticipated dipolar repulsion with the correct dependence on the applied magnetic field. Finally, we extend the method to measure the full set of pair potentials in multicomponent fluids, and discuss how an estimate of the uncertainty of the measurements can be obtained. Compared with existing techniques, the experimental and computational requirements are minimal, and the method will enable routine measurements of pair potentials in colloidal systems. The method will also be useful in deriving coarse-grained interactions for use in simulations of more complex and biological fluids.

This chapter is based on the following article [97]:

A. E. Stones, R. P. A. Dullens and D. G. A. L. Aarts. Model-free measurement of the pair potential in colloidal fluids using optical microscopy. Phys. Rev. Lett. **123**, 098002 (2019).

5.1 Introduction

The widespread use of colloidal systems both in industry and as experimental models in condensed matter physics derives from their diverse and tuneable interactions (see Chapter 1). It is somewhat surprising, then, that these interactions are usually only approximately known and based on theoretical assumptions, and are not routinely measured. Existing measurements come with certain drawbacks: methods based on inverting distribution functions are either only valid in infinite dilution [90], or rely on additional assumptions [212–215] or large numbers of computer simulations [91–94]. Direct methods [86,88,89] are often technically demanding, requiring specialist equipment, and measure the interactions away from the equilibrium fluid context, e.g. between two particles or a particle and a wall. By contrast, in this chapter we develop a versatile, fast, model-free method which allows an effective pair potential to be determined solely from equilibrium fluid configurations obtained using optical microscopy.

We saw in Chapter 4 that the pair distribution function $g^{(2)}(r)$ can be measured using a distance-histogram method based on

$$g^{(2)}(r) = \frac{\rho^{(1)}(r)}{\rho}, \tag{2.51}$$

or by test-particle insertion using

$$g^{(2)}(r) = \frac{\langle \exp[-\beta\Psi(r)] \rangle}{\langle \exp[-\beta\Psi] \rangle}. \tag{2.62}$$

In that chapter we focussed on the enhanced resolution of the insertion method, which there allowed for a measurement of the contact value $g^{(2)}(\sigma^+)$ in hard disks. Here, we focus on another key difference between the two methods: the distance-histogram method requires only equilibrium particle coordinates, while to employ the insertion method we must also know the interparticle interactions to calculate Ψ. We assume a uniform, pairwise-additive fluid of isotropic particles, such that the insertion energy Ψ is a sum of interactions depending on the pair potential $u^{(2)}(r)$. In this case, Henderson showed that $u^{(2)}(r)$ giving rise to a particular $g^{(2)}(r)$ is unique [216], and Chayes and Chayes later showed that even in the presence of many-body interactions, a unique, state-dependent *effective pair potential* can be found which gives rise to

a particular $g^{(2)}(r)$ [37, 217].[1] By matching $g^{(2)}(r)$ from insertion with the distance-histogram measurement, we provide an elegant route to measure $u^{(2)}(r)$.

We begin by explaining the predictor-corrector (PC) scheme used to match $g^{(2)}(r)$ and detail its numerical implementation, including for multicomponent fluids. After briefly outlining the protocols used for the simulations and colloidal experiments, we will test the method in MC simulations, where the measured $u^{(2)}(r)$ agree excellently with simulation input. We then apply the method to the superparamagnetic colloidal model system, finding excellent agreement with the form of $u^{(2)}(r)$ anticipated in (3.13). While this experimental demonstration is for a quasi-two dimensional, effective one-component fluid composed of particles with isotropic interactions, we stress that the method can be extended to study multicomponent fluids and anisotropic interactions in any dimension. We next apply the method in a three-component simulation, where all six input pair potentials are recovered. Finally, we estimate the uncertainty of the method and discuss its sensitivity to the input data.

5.2 Theory

5.2.1 The Predictor-Corrector Scheme

We achieve the practical step of matching $g^{(2)}$ from insertion with that from the distance histogram method using an iterative predictor-corrector (PC) scheme. Such schemes have been widely used in inverting $g^{(2)}(r)$ via molecular simulation [91, 92, 218–221] in procedures now known as Iterative Boltzmann Inversion (IBI) [95] and Inverse Monte Carlo (IMC) [222], and have been implemented in open-source software packages [96, 223]. In these schemes, an initial guess for the pair potential is either obtained by simple Boltzmann inversion [91]

$$u_0^{(2)}(r) = -k_{\mathrm{B}}T \ln \left[g^{(2)}(r) \right], \tag{5.1}$$

which is formally only valid in the limit of infinite dilution, or by employing a closure approximation [92]. A simulation is then performed with $u_0^{(2)}(r)$ to *predict* the pair distribution function

[1] We saw in Section 3.1.1 that colloidal interactions derived from integrating out molecular degrees of freedom generally involve many-body contributions. A more complete discussion of effective pair potentials is given in Section 7.2.2.

$g_0^{(2)}(r)$. In general, this measurement will be different from the *true*[2] pair distribution function $g_{\mathrm{h}}^{(2)}(r)$, but can be used to *correct* $u_0^{(2)}(r)$ to give an improved pair potential $u_1^{(2)}(r)$. In this work, we use a corrector due to Schommers [91],

$$u_{j+1}^{(2)}(r) = u_j^{(2)}(r) - k_{\mathrm{B}}T \ln \left[\frac{g_{\mathrm{h}}^{(2)}(r)}{g_j^{(2)}(r)} \right], \tag{5.2}$$

although alternative correctors have been suggested [92, 220]. These simulation-based inversion schemes can give state-dependent results even in the pairwise-additive case [95], which has led to the development of schemes based on simultaneously optimising $g^{(2)}(r)$ at several state points [96].

The crucial difference between these schemes and ours is that, rather than predicting $g_j^{(2)}(r)$ by performing a simulation and using the distance-histogram method, we instead use the test-particle insertion method on the *original* particle coordinates. This provides for a very efficient iterative scheme: by reusing the same insertion points for each iteration, the distances between the insertion points and the particles can be calculated in advance, and only $u_j^{(2)}(r)$ needs to be updated between iterations. There is no need for a separate time-consuming simulation for each iteration, although we should also note that iterative schemes have been used *on-the-fly* in MC simulations [224, 225]. Moreover, the insertion-based approach incorporates more information from the original configurations, since they are used continually in the insertion measurement of $g_j^{(2)}(r)$ for each iteration, rather than just initially to calculate $g_{\mathrm{h}}^{(2)}(r)$. Note, however, that approaches based on test-particle insertion break down at very high densities due to inadequate sampling [199], which precludes use of this method to measure $u^{(2)}(r)$ in high-density systems.

The corrector (5.2) originates in the observation that the cavity distribution function

$$y^{(2)}(r) = g^{(2)}(r) \exp\left[\beta u^{(2)}(r)\right], \tag{4.2}$$

is comparatively insensitive to the form of $u^{(2)}(r)$ [218].[3] For each iteration we therefore make

[2] In this chapter, we describe $g_{\mathrm{h}}^{(2)}(r)$ as the *true* pair distribution function, although we noted in Section 4.2.3 that $g_{\mathrm{h}}^{(2)}(r)$ is actually a convolution of the *true* $g^{(2)}(r)$ with a boxcar function [15].

[3] As discussed in Section 4.2.4, $y^{(2)}(r)$ is related to the indirect interactions of two particles via other particles in the system.

the approximation that

$$y_h^{(2)}(r) \approx y_j^{(2)}(r). \tag{5.3}$$

Substituting (4.2) into (5.3) we obtain

$$u^{(2)}(r) \approx u_j^{(2)}(r) - k_{\mathrm{B}}T \ln \left[\frac{g_{\mathrm{h}}^{(2)}(r)}{g_j^{(2)}(r)} \right], \tag{5.4}$$

and we take the right-hand side to be the next iteration of the pair potential $u_{j+1}^{(2)}(r)$ to give (5.2). This is then used to calculate $g_{j+1}^{(2)}(r)$ by insertion, and the procedure is performed iteratively until convergence, which is verified by monitoring the value of

$$\chi_j^2 = \sum_k \left[g_j^{(2)}(r_k) - g_{\mathrm{h}}^{(2)}(r_k) \right]^2, \tag{5.5}$$

where the r_k are the mid-points of the bins used in the distance-histogram method.

5.2.2 Multicomponent Systems

The extension of the PC scheme to the multicomponent case is straightforward. A pairwise-additive fluid containing n components will have a set of $n(n+1)/2$ pair potentials $\{u_{\alpha\beta}^{(2)}(r)\}$, each corresponding to the interaction between particles of a pair of components $\{\alpha, \beta\}$. In Chapter 4 we introduced the *partial* pair distribution functions $\{g_{\alpha\beta}^{(2)}(r)\}$ and showed that they too can be measured by both the insertion and distance-histogram methods. To extend the PC scheme, we simply substitute $u_{\alpha\beta}^{(2)}(r)$ and $g_{\alpha\beta}^{(2)}(r)$ for their single-component counterparts in (5.2) to give the corrector[4]

$$u_{\alpha\beta,j+1}^{(2)}(r) = u_{\alpha\beta,j}^{(2)}(r) - k_{\mathrm{B}}T \ln \left[\frac{g_{\alpha\beta,\mathrm{h}}^{(2)}(r)}{g_{\alpha\beta,j}^{(2)}(r)} \right]. \tag{5.6}$$

The scheme is essentially identical: for each iteration, $g_{\alpha\beta,j}^{(2)}(r)$ is measured for each pair of components by insertion, and then used with the corresponding $g_{\alpha\beta,\mathrm{h}}^{(2)}(r)$ in (5.6) to obtain the

[4]Simulation-based PC schemes can also be extended 'to more than one potential/distribution function pair' [95] in a similar manner.

corrected pair potential $u^{(2)}_{\alpha\beta,j+1}(r)$. The convergence is monitored using

$$\chi^2_j = \sum_{\alpha,\beta} \sum_k \left[g^{(2)}_{\alpha\beta,j}(r_k) - g^{(2)}_{\alpha\beta,\mathrm{h}}(r_k) \right]^2, \tag{5.7}$$

where the first sum runs over all pairs of components.[5]

5.3 Methods

Except for the multicomponent system, the implementation and protocols of this section follow those in the Supplemental Material of Reference [97], on which this chapter is based. The implementation, in particular, is described in more detail here. © 2019 American Physical Society.

5.3.1 Implementation

Distance-Histogram Method

For each of the simulation or experimental snapshots, we calculate the distances between each particle and its neighbours up to a cut-off distance d using a k-d tree [202, 203] (see note 4 on page 56). The cut-off distance should be large enough to capture all features of the pair potential, such that $u^{(2)}(r) \approx 0$ beyond d. We then compute a histogram of these distances using bins centred at mid-points r_k, and sum the results for all snapshots together. For the simulation analysis, we then divide each of these values by the relevant bin volume and the total number of particles considered to obtain a discretised $\rho^{(1)}(r)$. Dividing by a measurement of the bulk density ρ yields $g^{(2)}_{\mathrm{h}}(r)$ according to (2.51).

There are no PBCs in the experimental case, and so we first divide the histogram values by the total number of particle pairs considered. Each of these values then corresponds to the probability that a random pair of particles are separated by a distance which falls in the relevant bin. To obtain $g^{(2)}_{\mathrm{h}}(r)$, we then divide these values by the analogous probability for a system of non-interacting particles, which we calculate from the PDF of the distance between two random

[5] Note that there are two insertion measurements of a given $g^{(2)}_{\alpha\beta}(r)$, with each component fulfilling the role of the inserted particle or the reference particle. These are separately optimised with respect to $g^{(2)}_{\alpha\beta,\mathrm{h}}(r)$, and so both are included in the sum.

points in a box [226].

Insertion Method: Weighted-Histogram Approach

While (2.62) can be applied directly as in Chapter 4, with test-particle insertions attempted at fixed separations from existing particles of the fluid, the need to store the distances between the insertion points and their neighbouring particles in the memory makes this computationally infeasible. Noting that many of the insertion points are very close to each other, we instead employ an approximate approach, attempting insertion on a grid of points over the simulation box or field-of-view. This approach is in the same spirit to one discussed by Puibasset and Belloni in the more complicated context of a dipolar fluid [111]. While this approximation sacrifices the exact resolution of the insertion method, this is not important here since the resolution of the measured $u^{(2)}(r)$ is already limited by that of the distance-histogram result.

For each iteration, we evaluate the insertion energy Ψ for each point by using the distances to neighbouring particles (up to d) in the trial pair potential $u_j^{(2)}(r)$; since the distances are pre-calculated, this is very fast. We then approximate the local ensemble average $\langle \exp[-\beta\Psi(r_k)] \rangle$ by averaging over all points which are separated from particles by a distance in the bin centred at r_k, i.e. by constructing a *weighted histogram*; in this way, we consider each insertion point as being at an approximate separation from many different particles. We obtain the bulk ensemble average $\langle \exp[-\beta\Psi] \rangle$ by averaging over all of the insertion points, and use (2.62) to calculate $g_j^{(2)}(r)$.

Predictor-Corrector Scheme

For each iteration, we use the insertion result $g_j^{(2)}(r)$ in (5.2) to obtain the next iteration of the pair potential, $u_{j+1}^{(2)}(r)$. The corrector cannot be employed immediately since $g_{\mathrm{h}}^{(2)}(r)$ and $g_j^{(2)}(r)$ are often measured as zero for small r, and their logarithms cannot be taken.[6] We therefore recast (5.2) as

$$e_{j+1}^{(2)}(r) = e_j^{(2)}(r) \frac{g_{\mathrm{h}}^{(2)}(r)}{g_j^{(2)}(r)}, \tag{5.8}$$

[6]With hard particles $g^{(2)}(r)$ is exactly zero in the core. For continuous $u^{(2)}(r)$ such as the Lennard-Jones potential it is finite but *extremely* small, and so is measured as zero in practice.

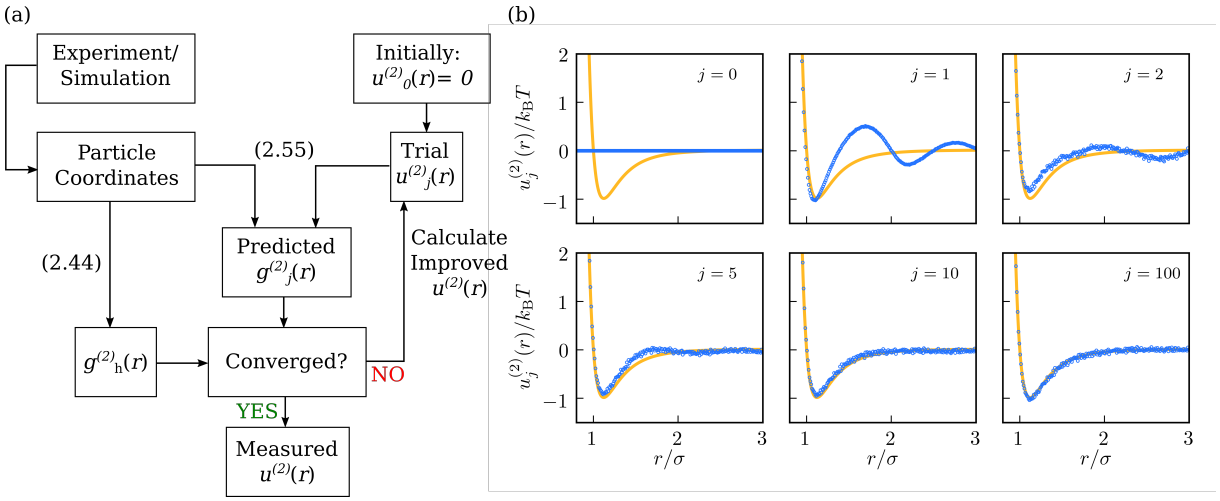


Figure 5.1: (a) Flowchart summarising the iterative scheme for measuring $u^{(2)}(r)$ by test-particle insertion. (b) Progress of the iterative scheme applied to configurations from a Lennard-Jones simulation, comparing the trial $u_j^{(2)}(r)$ after j iterations (blue circles) with the simulation input (yellow line).

where $e_j^{(2)}(r) = \exp[-u_j^{(2)}(r)/k_BT]$. We first replace zero values of $g_j^{(2)}(r)$ by a small non-zero value, in our case 10^{-20}, to avoid dividing by zero. The resulting $e_{j+1}^{(2)}(r)$ is zero wherever $g_h^{(2)}(r)$ is zero and so these values are too replaced by 10^{-20}, before the corrected pair potential is calculated as $u_{j+1}^{(2)}(r) = -k_BT \ln[e_{j+1}^{(2)}(r)]$.[7] Finally, note that we begin with the trial potential $u_0^{(2)}(r) = 0$ everywhere ($e_0^{(2)}(r) = 1$) to emphasise the model-free nature of our approach, but since this always gives an insertion measurement $g_0^{(2)}(r) = 1$, $u_1^{(2)}(r)$ has the form given in (5.1) anyway. The iterative scheme is summarised by the flowchart in Figure 5.1, which also shows the progress of the iterative scheme applied to configurations from a Lennard-Jones simulation.

A major advantage of using the insertion method to predict $u_j^{(2)}(r)$ is that the distances between the insertion points and the particles can be calculated in advance, and there is no need to perform a time-consuming simulation for each iteration. To store these distances requires careful memory management, and so we only store them at the precision required for the calculation.[8] We further improve the computational efficiency by using a lookup table for $u_j^{(2)}(r)$, which has a higher resolution than that of the $g^{(2)}(r)$ histograms. To achieve an efficient implementation, we use sparse matrices [204] as detailed below.

[7] As a consequence of replacing zero with a non-zero value, an upper bound is placed on the measured pair potential (here $\sim 46k_BT$). In practice, this only alters the value deep inside the strongly-repulsive core region of the particle. Since the physically-important quantity is actually the interparticle force, the particles are essentially precluded from getting this close anyway.

[8] In practice, this does not limit the accuracy of the calculation, since for the experiments there is an error associated with locating the particles anyway.

We calculate the distances between each insertion point and its neighbouring particles using a k-d tree, and use these to construct a comparison matrix $\mathbf{C}$ and a lookup matrix $\mathbf{L}$. The columns of these matrices correspond to the intervals of the $g^{(2)}(r)$ histograms and the lookup table, respectively, and the rows correspond to the insertion points (insertion points at the same location but in different snapshots are considered separately). The matrix elements are the number of particles in the corresponding interval around each insertion point. The current $u_j^{(2)}(r)$ is stored as a series of values for the midpoints r_k of the histogram bins; for each iteration, the lookup table is calculated by linear interpolation and stored as a column vector $\mathbf{u}_j$. The insertion energy for each insertion point is then calculated as the matrix product $\mathbf{L}\mathbf{u}_j$, which is exponentiated element-wise to give the $\exp[-\Psi/k_\mathrm{B}T]$, stored as a column vector $\mathbf{e}_j$. The bulk ensemble average is simply the mean of the elements of $\mathbf{e}_j$, while the local ensemble average for each r_k is evaluated by first taking the matrix product $\widetilde{\mathbf{C}}\mathbf{e}_j$, and then dividing each element by the sum of the corresponding column of $\mathbf{C}$. We use (2.62) to calculate $g_j^{(2)}(r)$ for use in (5.8).

For multicomponent systems, the implementation is very similar: we initially calculate the distance-histogram partial pair distribution functions $\{g_{\alpha\beta,\mathrm{h}}^{(2)}(r)\}$, and then use the insertion results $\{g_{\alpha\beta,j}^{(2)}(r)\}$ to find the next iteration of the pair potentials $\{u_{\alpha\beta}^{(2)}(r)\}$. We calculate a separate $\mathbf{L}_\alpha$ and $\mathbf{C}_\alpha$ for each component, using the same insertion points and intervals in each case: the matrix elements are the number of particles of the component in the corresponding interval around each insertion point. After calculating the lookup vectors for each component pair, $\{\mathbf{u}_{\alpha\beta,j}\}$, we calculate the insertion energy for each component α at each point as $\sum_\beta \mathbf{L}_\beta \mathbf{u}_{\alpha\beta,j}$ and obtain $\mathbf{e}_{\alpha,j}$ as above. The bulk ensemble average for insertion of α particles is the mean of the elements of $\mathbf{e}_{\alpha,j}$, and the local ensemble average for inserting α particles around reference β particles is $\widetilde{\mathbf{C}}_\beta \mathbf{e}_{\alpha,j}$ divided by the sum of the corresponding column of $\mathbf{C}_\beta$. These averages are then used to calculate $\{g_{\alpha\beta,j}^{(2)}(r)\}$ for use in (5.7). Note that there are two insertion measurements for a given $g_{\alpha\beta}^{(2)}(r)$: one corresponds to insertion of α particles around β particles, and the other to insertion of β particles around α particles. Both are optimised separately with respect to $g_{\alpha\beta,\mathrm{h}}^{(2)}(r)$ to obtain two pair potentials, $u_{\alpha\beta}^{(2)}(r)$ and $u_{\beta\alpha}^{(2)}(r)$, which should agree with each other.

5.3.2 Protocols

Simulation

We used the Grand Canonical Monte Carlo scheme (Section 3.3.2), since in the grand ensemble, $g^{(2)}(r) \to 1$ exactly for large r, while in the canonical ensemble $g^{(2)}(r) \to 1-1/N$ [1], which might affect the comparison of the two methods in the iterative scheme. For the single-component simulations, we used a two-dimensional square box of side length 50σ, with PBCs employed to reduce finite-size effects. Initially, we placed 10 particles in the box at random and then allowed the density to increase for 10^6 equilibration steps; we then took 10^3 snapshots at intervals of 10^4 steps. Details of the input pair potentials and parameters are given in Table 5.1: we used a wide range of pair potentials of interest in liquid state theory and colloidal science, including hard disks, attractive and repulsive square-well potentials, attractive and repulsive hard-core Yukawa potentials, and a hard-core two-Yukawa potential with competing short-range attractions and long-range repulsions [227]. In all cases, the particle size parameter $\sigma = 1$ and the pair potential was truncated at $r = 2.5\sigma$, and shifted so that $u^{(2)}(2.5\sigma) = 0$ and $u^{(2)}(r)$ is continuous at this separation. Note that we also show the simulation chemical potential μ', which is defined such that

$$\exp[\beta\mu'] = z\sigma^2, \tag{5.9}$$

where z is the reduced activity. For all simulations, the inverse temperature $\beta = 1$.

For the multicomponent simulation, we used three components, labelled α, β and γ. The six interactions were of the Lennard-Jones form (see Table 5.1); the parameters for the same-component interactions were $\sigma_{\alpha\alpha} = 1$, $\epsilon_{\alpha\alpha} = 0.5$; $\sigma_{\beta\beta} = 1.25$, $\epsilon_{\beta\beta} = 1$ and $\sigma_{\gamma\gamma} = 1.5$, $\epsilon_{\gamma\gamma} = 1.5$. The parameters for the inter-component interactions were given by the Lorentz and Berthelot mixing rules [1], e.g.

$$\sigma_{\alpha\beta} = \frac{\sigma_{\alpha\alpha} + \sigma_{\beta\beta}}{2}, \tag{5.10}$$

$$\epsilon_{\alpha\beta} = \sqrt{\epsilon_{\alpha\alpha}\epsilon_{\beta\beta}}. \tag{5.11}$$

The pair potentials were again truncated and shifted at, e.g. $r = 2.5\sigma_{\alpha\beta}$. The inverse temper-

Table 5.1: The input pair potentials and parameters used in the simulation. Also shown is the density of each simulation. © 2019 American Physical Society.

Pair Potential	$u^{(2)}(r)$	Parameters	μ'	$\rho\sigma^2$
Lennard-Jones	$4\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$	$\sigma = 1$	1	0.677
Hard disk	$\infty \quad r \leq \sigma$ $0 \quad r > \sigma$	$\sigma = 1$	2	0.481
Attractive square-well	$\infty \quad r \leq \sigma$ $\epsilon \quad \sigma < r \leq \lambda\sigma$ $0 \quad r > \lambda\sigma$	$\sigma = 1$ $\lambda = 1.5$ $\epsilon = -1$	-1	0.539
Repulsive square-well	$\infty \quad r \leq \sigma$ $\epsilon \quad \sigma < r \leq \lambda\sigma$ $0 \quad r > \lambda\sigma$	$\sigma = 1$ $\lambda = 1.5$ $\epsilon = 1$	1	0.264
Attractive Yukawa	$\infty \quad r \leq \sigma$ $\left(\frac{\epsilon\sigma}{r}\right)\exp\left[-\kappa\left(\frac{r}{\sigma}-1\right)\right] \quad r > \sigma$	$\sigma = 1$ $\kappa = 1.5$ $\epsilon = -1$	1	0.580
Repulsive Yukawa	$\infty \quad r \leq \sigma$ $\left(\frac{\epsilon\sigma}{r}\right)\exp\left[-\kappa\left(\frac{r}{\sigma}-1\right)\right] \quad r > \sigma$	$\sigma = 1$ $\kappa = 1.5$ $\epsilon = 1$	1	0.306
Hard-core two-Yukawa (HCTY)	$\infty \quad r \leq \sigma$ $\left(\frac{\epsilon\sigma}{r}\right)\left\{-\exp\left[-z\left(\frac{r}{\sigma}-1\right)\right] + A\exp\left[-y\left(\frac{r}{\sigma}-1\right)\right]\right\} \quad r > \sigma$	$\sigma = 1$ $\epsilon = 6$ $A = 0.8$ $z = 4$ $y = 2.5$	1	0.367

ature $\beta = 1$, and the simulation chemical potentials were $\mu'_\alpha = -1$, $\mu'_\beta = -2$ and $\mu'_\gamma = -3$.[9] We used a square box of side length $100\sigma_{\alpha\alpha}$, which was initially populated by attempting 1500 particle additions with the components chosen at random. During the first 2×10^6 steps, we adjusted the maximum displacement for each component such that around 50% of movements were accepted [228]; the system was then equilibrated for another 5×10^5 moves. We took a total of 10^3 snapshots at intervals of 10^5 steps. For each step, we selected the component at random, with equal probability for each component.

Experiment

We used the model system composed of superparamagnetic particles, preparing samples of two different number densities using the protocol discussed in Section 3.2.1. We carried out measurements for a range of different external magnetic fields, each corresponding to a different pair potential. The system was followed using optical microscopy (Section 3.2.3), with images taken every second for an hour, and the particles were tracked using the customised procedure of Section 3.2.4. We summarise the experimental data used in the analysis in Table 5.2.

Analysis

When analysing the simulation data, we used 10^4 insertion points in each of the obtained snapshots. For the single-component analysis, we used a pair-potential cut-off of 5σ, a comparison interval of 10^{-2} and a lookup interval of 5×10^{-3}. We performed 250 iterations of the PC scheme in each case. For the multicomponent analysis, we used a pair-potential cut-off of $5\sigma_{\alpha\alpha}$, a comparison interval of 2×10^{-2} and a lookup interval of 2×10^{-3}, performing 100 iterations. In all cases, the PC scheme converged.

When analysing the experimental data, we used all the snapshots indicated in Table 5.2 for the distance-histogram calculation, and every tenth snapshot for the insertion analysis, with 10^4 insertion points in each. We used a pair-potential cut-off of 34.6 µm, a comparison interval of 0.138 µm and a lookup interval of 0.014 µm. Some particles at the image edges were missed by the detection algorithm, and so we excluded a border of 13.8 µm from the analysis. We performed 500 iterations of the PC scheme, achieving convergence in all cases.

[9] All of the simulation chemical potentials are defined with respect to $\sigma_{\alpha\alpha}$, i.e. $\exp[\beta\mu'_\gamma] = z_\gamma\sigma_{\alpha\alpha}$, since we carried out the simulation in these units.

Table 5.2: Summary of the experimental data used in the analysis. We use $\sigma = 3.04\,\mu m$, based on the extracted pair potentials. Note that the number density varies due to fluctuations or changing the location of the field of view. © 2019 American Physical Society.

Sample	Field (mT)	Snapshots	$\rho\sigma^2$	Sample	Field (mT)	Snapshots	$\rho\sigma^2$
1	0.093	3600	0.139	2	0.093	3600	0.175
1	0.142	3600	0.138	2	0.142	3600	0.179
1	0.191	3600	0.143	2	0.191	3600	0.176
1	0.239	2865	0.144	2	0.239	3600	0.177
1	0.288	3600	0.144	2	0.288	3600	0.180
1	0.386	3600	0.143	2	0.386	3600	0.192
1	0.483	3600	0.110	2	0.483	3600	0.177

5.4 Results and Discussion

The first two parts of this section largely follow part of Reference [97], on which this chapter is based. © 2019 American Physical Society.

5.4.1 Simulations: Testing the Method

The pair potentials measured from the simulation data are compared with the simulation input in Figure 5.2. Figure 5.2(a) shows that for the continuous Lennard-Jones potential, the agreement between the measured and input $u^{(2)}(r)$ is excellent; the remaining noise could *in principle* be reduced by performing longer simulations. We stress that the measurement is *model-free*, and no particular form for the pair potential is assumed. Figure 5.2(b) illustrates that the PC scheme is successful in enforcing consistency between the two results for $g^{(2)}(r)$: after convergence, the insertion measurement agrees essentially perfectly with that of the distance-histogram method. This convergence is quantified in Figure 5.2(c), where χ_j^2, which is a measure of the difference between the two $g^{(2)}(r)$, is found to converge after $\sim$100 iterations.

Figures 5.2(d-i) show the results for the other pair potentials; in each case the agreement is excellent, and convergence was achieved in even fewer iterations than in the Lennard-Jones case. The only outlying points are found at the discontinuities in the input $u^{(2)}(r)$, and arise because the corresponding discontinuities in $g^{(2)}(r)$ are neither fully captured by the distance-

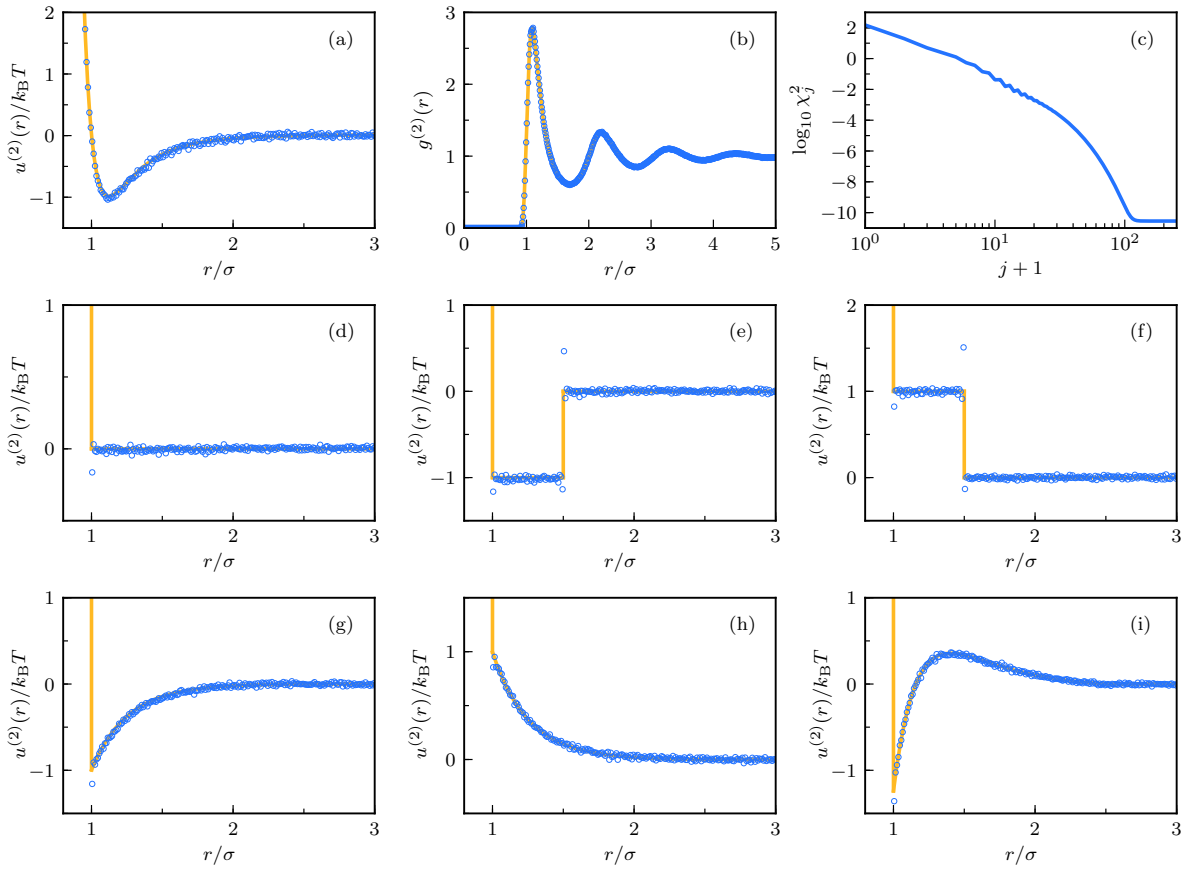


Figure 5.2: (a) Measured Lennard-Jones $u^{(2)}(r)$ (blue circles), compared with the simulation input (yellow line). (b) Corresponding $g^{(2)}(r)$ from the distance-histogram method (yellow line) and by test-particle insertion with the measured pair potential (blue circles). (c) Log-log plot showing convergence of the PC scheme. Remaining panels show $u^{(2)}(r)$ measured for (d) a hard-disk potential, (e) a hard-core attractive square-well potential, (f) a hard-core repulsive square-well potential, (g) an attractive Yukawa potential, (h) a repulsive Yukawa potential, (i) a hard-core two-Yukawa (HCTY) potential. Reproduced from [97]. © 2019 American Physical Society.

histogram method nor the approximate (weighted-histogram) insertion method. Nevertheless, the form of each potential is very clear from the measurements, since the agreement away from the discontinuities is essentially perfect.

5.4.2 Experiments: Superparamagnetic Colloidal Particles

Figure 5.3 shows the pair potentials measured from the experimental data. Recall that at separations beyond the hard-core region, a repulsive-dipolar pair potential is expected such that

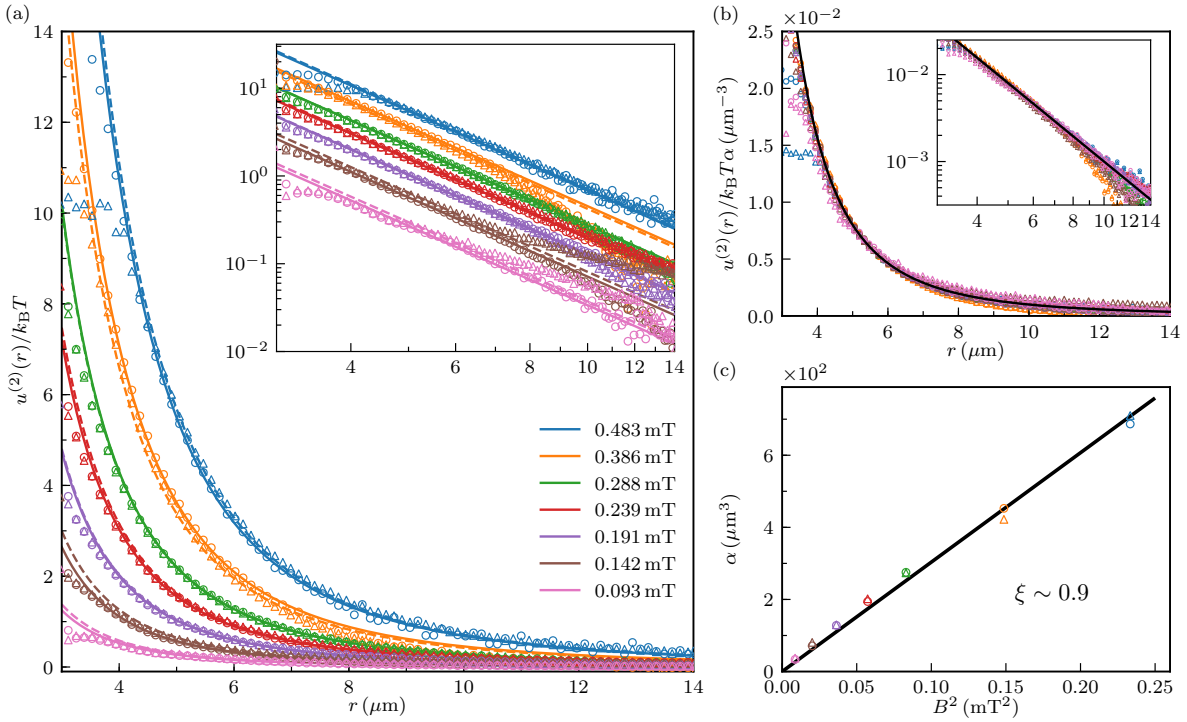


Figure 5.3: (a) Repulsive dipolar pair potentials measured for two samples at different number densities (Sample 1—triangles and dashed lines, Sample 2—circles and solid lines) for different B; the lines are fits to (5.12). Inset: corresponding log-log plot where the pair potentials appear as parallel straight lines with slope -3. (b) After dividing by the power law prefactor α, the measured pair potentials collapse onto a single curve. (c) Plot of α against B^2—as anticipated, the data fall on a straight line. Reproduced from [97]. © 2019 American Physical Society.

$$u^{(2)}(r) = \begin{cases} \infty, & r \leq \sigma, \\ \alpha/r^3, & r > \sigma, \end{cases} \tag{3.13}$$

with

$$\alpha = \frac{\pi\sigma^6\xi^2B^2}{144k_BT\mu_0}. \tag{3.12}$$

A similar model system has been used in several important studies by Maret and co-workers [60–62], who determined α by comparing $g^{(2)}(r)$ from experiments with those from simulations. With our model-free approach, we are not only able to measure the parameter α, but to confirm that (3.13) is the correct *form* of $u^{(2)}(r)$.

The $u^{(2)}(r)$ measured by the PC scheme are shown in Figure 5.3(a): in all cases $u^{(2)}(r)$ is

clearly repulsive, with the strength of the repulsion increasing with B. The measured $u^{(2)}(r)$ decay as a power law with exponent -3, evidenced by the straight lines on the log-log plot (inset). For a given B, the measurements agree very well between the two samples, showing that at these densities three-body and higher-order interactions are not important for this system. The uncertainty of these measurements and their sensitivity to the input data are discussed in the next section.

For the fitting, the dipolar repulsion of (3.13) was recast as

$$\ln\left[\frac{u^{(2)}(r)}{k_{\mathrm{B}}T}\right] = \ln\alpha - 3\ln r, \tag{5.12}$$

and a least-squares minimisation was performed to fit $\ln u^{(2)}$ as a function of $\ln r$, with α as the only fitting parameter. In each case, the data between 3.94 µm and 7.13 µm were used for the fit, since the log-log plot shows that the measured $u^{(2)}(r)$ are well-described by a dipolar repulsion in this region. After dividing each $u^{(2)}(r)$ by α from the corresponding fit, the data collapse onto a single curve as illustrated in Figure 5.3(b). In Figure 5.3(c), we plot the fitted α against B^2, finding excellent agreement with the straight line predicted in (3.12). A least-squares fit gives the dimensionless volume susceptibility $\xi \sim 0.9$, but we note that this value depends sensitively on the particle diameter ($\xi \propto \sigma^{-3}$). Here we use $\sigma = 3.04\,\mu\mathrm{m}$, based on the measured $u^{(2)}(r)$.

5.4.3 Multicomponent Simulation

Figure 5.4 shows the results of the multicomponent PC scheme applied to the three-component simulation. The partial pair distribution functions for each pair of components, measured by the distance-histogram method, are shown in Figure 5.4(a) and Figure 5.4(b). While each one gives an indication of the length scale of the strong repulsive part of the corresponding pair potential, they do not reveal the finer details of the attractive parts. The convergence of the PC scheme is shown in Figure 5.4(c). Remarkably, all six pair potentials are recovered, and are compared with the simulation input in Figures 5.4(d-i). In Figures 5.4(g-i), we show two measured pair potentials which correspond to matching each of the insertion measurements (with each component considered as either the inserted or reference particle) to the distance-histogram result. The measured pair potentials agree with each other in all three cases, demonstrating the internal consistency of the PC scheme.

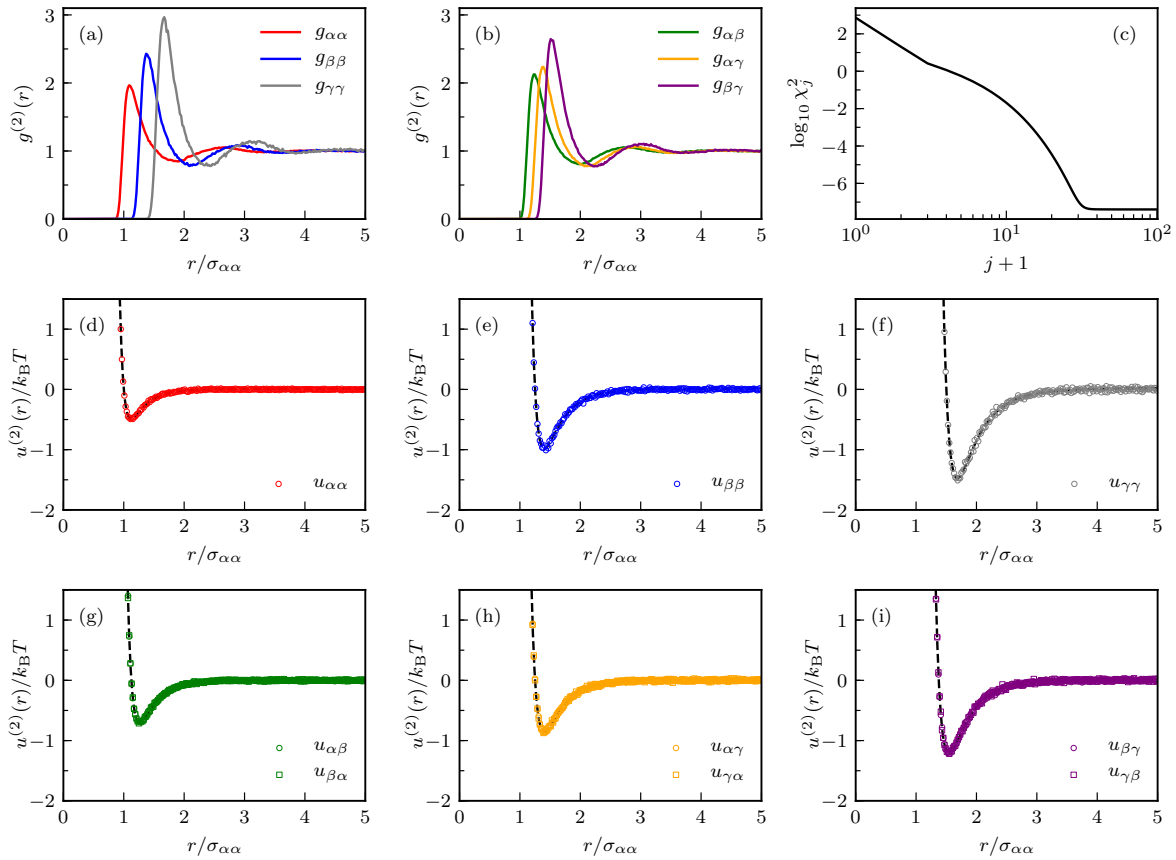


Figure 5.4: (a) Measured partial pair distribution functions between particles of the same species and (b) different species. (c) The extended PC scheme of (5.6) converges. (d-i) Measured pair potentials (symbols) compared with the simulation input (dashed lines); in the case of different species, measurements using both possible insertion approaches are included.

5.5 Uncertainty in the Measured Pair Potential

This section essentially reproduces the corresponding section in the Supplemental Material of Reference [97], on which this chapter is based. © 2019 American Physical Society.

5.5.1 Uncertainty Estimation Using the Measurement Noise

Calculating the uncertainty in the measured $u^{(2)}(r)$ is challenging because, in principle, every iteration of the PC scheme introduces noise from both the distance-histogram and insertion measurements, yet the uncertainty should stop growing on convergence. The dependence of the uncertainty on the separation r is non-trivial: for each iteration, the entire $u_j^{(2)}(r)$ is used to calculate $g_j^{(2)}(r)$ and obtain an improved pair potential, and so each separation cannot be

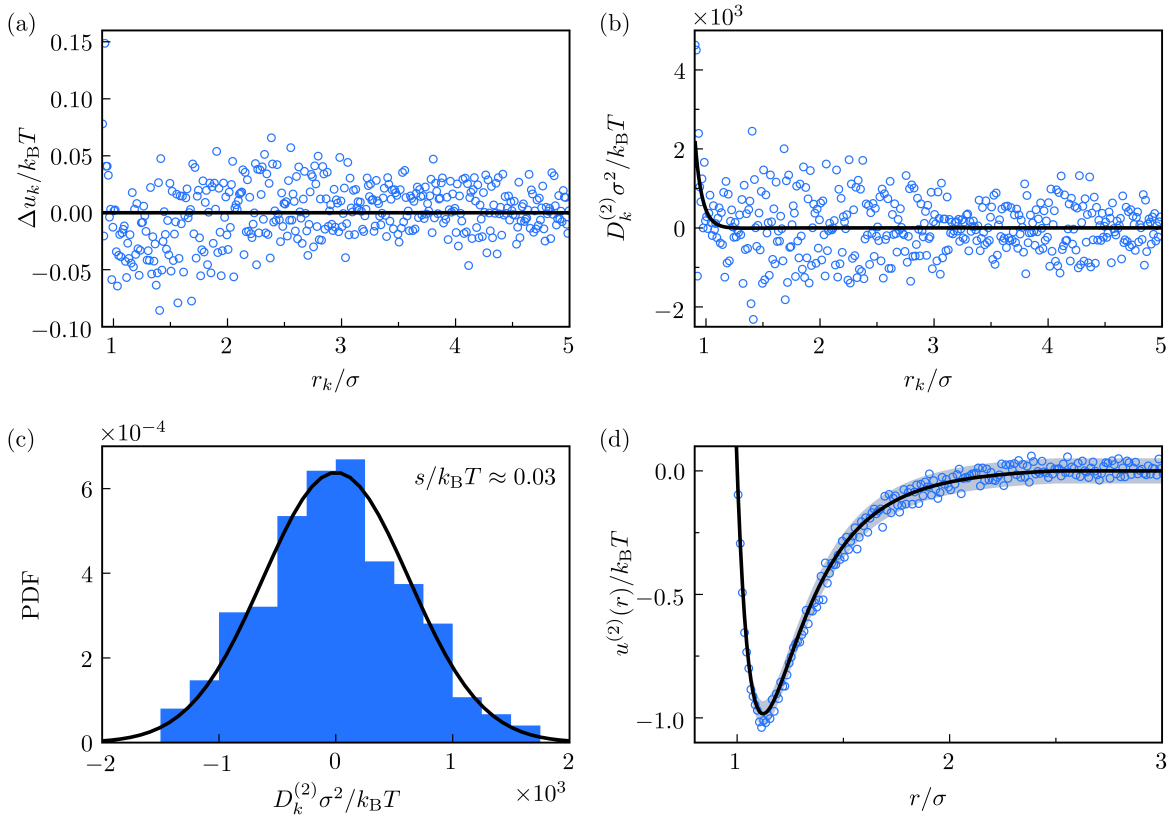


Figure 5.5: (a) The deviation of the measured Lennard-Jones $u^{(2)}(r)$ from the simulation input as a function of separation. (b) The numerical second derivative of the measured $u^{(2)}(r)$ (blue circles) compared with that of the true $u^{(2)}(r)$ (black line). (c) A histogram of the second derivatives for separations $r_k > 2\sigma$, and the estimated normal distribution (black line). (d) Comparison of part of the measured (blue circles) and input (black line) $u^{(2)}(r)$. The shaded area corresponds to $u^{(2)}(r) \pm 2s$. Reproduced from [97]. © 2019 American Physical Society.

considered independently.

When implementing the scheme, $u_j^{(2)}(r)$ is discretised and we actually measure a set of values u_k corresponding to each r_k. The deviation from the true pair potential $\widetilde{u}_k$ is $\Delta u_k = u_k - \widetilde{u}_k$; we assume here that Δu_k is drawn from a normal distribution which is independent of separation and characterised by a standard deviation s and mean 0. We further assume that the deviations of neighbouring points are uncorrelated. A numerical estimate of the second derivative at r_k is given by

$$D_k^{(2)} = \frac{1}{w^2}\left(u_{k+1} + u_{k-1} - 2u_k\right) \tag{5.13}$$

$$= \frac{1}{w^2}\Big[(\widetilde{u}_{k+1} + \widetilde{u}_{k-1} - 2\widetilde{u}_k) + (\Delta u_{k+1} + \Delta u_{k-1} - 2\Delta u_k)\Big], \tag{5.14}$$

Table 5.3: The uncertainty in the simulation pair potentials estimated using the second-derivative method compared with that estimated directly using the input pair potential.
© 2019 American Physical Society.

Pair potential	$s/10^{-2}k_{\mathrm{B}}T$	
	via $D_k^{(2)}$	Direct
Lennard-Jones	2.6	2.1
Hard disk	1.0	0.8
Attractive square-well	1.6	1.3
Repulsive square-well	1.6	1.3
Attractive Yukawa	1.5	1.6
Repulsive Yukawa	1.2	1.3
Hard-core two-Yukawa (HCTY)	1.0	0.8

where w is the bin width. Where the true pair potential is smooth, the first term on the right-hand side of (5.14) is expected to be small compared with the second term and is thus neglected. The second term is a normally-distributed random variable with standard deviation $\sqrt{6}s/w^2$, and is used to estimate s.[10]

Figure 5.5 illustrates the use of this method in the Lennard-Jones simulation, where we can compare to the exact input pair potential. For separations $r \geq 2\sigma$ ($u^{(2)}(r) \approx 0$), the deviation from the true pair potential is approximately symmetric about 0, as shown in Figure 5.5(a). Figure 5.5(b) verifies the assumption that the second derivative of the true pair potential makes a negligible contribution to the second derivative of the measured pair potential $D_k^{(2)}$ at these separations. The corresponding $D_k^{(2)}$ are found in Figure 5.5(c) to obey the anticipated normal distribution with $s \approx 0.03k_{\mathrm{B}}T$. In Figure 5.5(d), we compare the measured $u^{(2)}(r)$ to the simulation input and find the deviation to be well-captured by a spread of $\pm 2s$. In Table 5.3, we compare these estimates of s for each simulation to those obtained by considering the deviation from the true pair potentials, finding reasonable agreement in all cases.

For each of the experimentally-measured pair potentials, we performed the same analysis, considering separations larger than 5 µm, and obtained a normal distribution analogous to that of Figure 5.5(c). We show the estimated uncertainties in Table 5.4; $s < 0.1k_{\mathrm{B}}T$ in all cases.

[10]Note that any bias due to the correlations in neighbouring $D_k^{(2)}$ is ignored.

Table 5.4: Estimates of the uncertainties in the experimental pair potential measurements.
© 2019 American Physical Society.

Field (mT)	$s/10^{-2}k_{\mathrm{B}}T$	
	Sample 1	Sample 2
0.093	0.4	0.3
0.142	0.5	0.4
0.191	0.5	0.6
0.239	0.7	0.7
0.288	0.8	1.2
0.386	1.2	3.0
0.483	2.0	6.7

5.5.2 Sensitivity to the Input Data

Using fewer configurations in the analysis results in more noise in the measured $g_{\mathrm{h}}^{(2)}(r)$ and each measurement of $g_j^{(2)}(r)$, in turn giving rise to more noise in the measured $u^{(2)}(r)$ and potentially resulting in the PC scheme failing to converge. To investigate the effect of limited data, we reanalysed the experimental samples corresponding to the least (Sample 1, 0.093 mT) and greatest (Sample 2, 0.483 mT) structure. The data were split into five parts of equal length and analysed separately, with each part effectively corresponding to a shorter experiment.

The measured $u^{(2)}(r)$ are shown in Figure 5.6. While the qualitative repulsive behaviour is found even when only part of the data is analysed, the finer quantitative details are lacking. In particular, we note the strong differences at large separations, which would likely preclude a confident prediction of the power-law exponent. Especially in the case of least structure, there is considerable advantage in using all of the data simultaneously in the PC scheme, rather than splitting the data in this way and calculating the average result. In principle, more data can always be obtained by performing longer experiments or simulations, and then incorporated into the PC scheme using more memory.

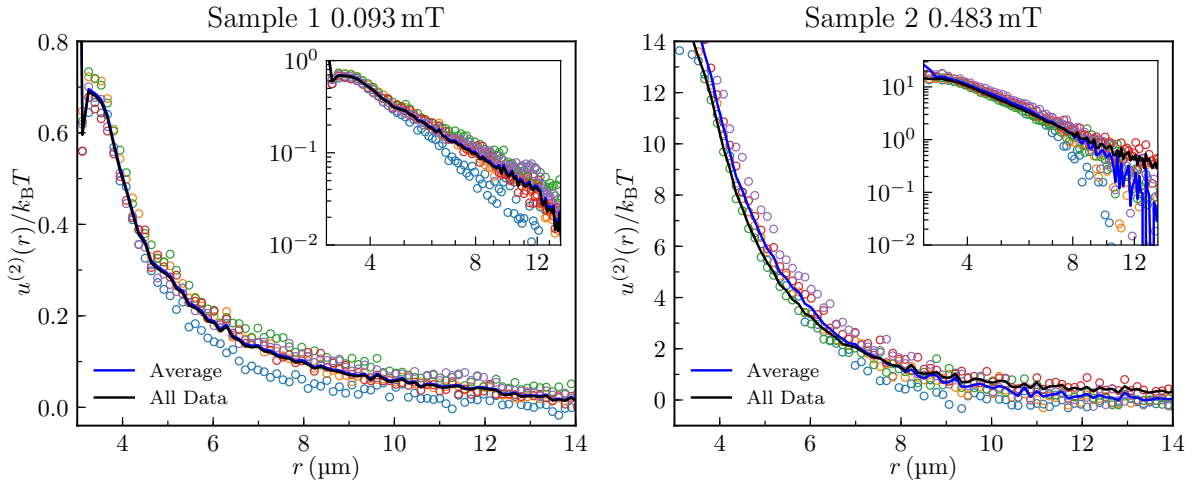


Figure 5.6: Comparison of $u^{(2)}(r)$ measured by considering all of the data (black line) with those measured by separately considering five parts of the data (coloured circles), and their average (blue line). Inset: the results shown on a log-log plot. Reproduced from [97]. © 2019 American Physical Society.

5.6 Conclusion and Outlook

We have developed a novel model-free method for measuring pair potentials from equilibrium configurational data, which is based on enforcing consistency between the distance-histogram and insertion measurements of $g^{(2)}(r)$ using an iterative predictor-corrector scheme. We have first demonstrated the accuracy and versatility of the method in simulations using a wide range of pair potentials, recovering the simulation input in all cases. We have then applied the method experimentally using the superparamagnetic colloidal model system, obtaining the anticipated dipolar repulsion with the correct dependence on the applied external magnetic field. We have also extended the method to measure the full set of pair potentials in multicomponent fluids and demonstrated its application in a three-component simulation. Finally, we have provided a method for estimating the uncertainty in the measured potential and have discussed the sensitivity of the method to the quantity of input data.

In contrast with existing pair potential measurements, the experimental and computational requirements of our method are minimal, requiring only a simple transmission microscope to obtain equilibrium particle coordinates (two dimensions). The method has clear relevance in industry, where knowledge of interparticle interactions is crucial in product formulation, and will enable coarse-grained pair potentials to be derived for use in simulations of more complex

fluids and biological systems [99, 220, 229]. The extension of the method to multicomponent systems suggests its experimental application in colloidal mixtures, as well as further extensions to three dimensions (using confocal microscopy) and to systems with anisotropic interactions. In Chapter 7, we will see that applying this method in a system with three-body interactions yields an *effective pair potential*, which is able to reproduce $g^{(2)}(r)$ but not higher order distribution functions such as $g^{(3)}$. Extending the PC scheme to measure three-body interactions requires an insertion measurement of $g^{(3)}$, which is the focus of the next chapter.

Acknowledgements

Members of the Oxford Colloid Group, especially Carla Fernández-Rico, are thanked for useful discussions and practical assistance.

Chapter 6

Measuring Many-Body Distribution Functions in Simple Fluids

We extend the test-particle insertion method for measuring $g^{(2)}(r)$ (Chapter 4) to measure many-body distribution functions $g^{(n)}$. We show that, in principle, a given $g^{(n)}$ can be measured using n different insertion methods, which correspond to test-insertion of between 1 and n particles. We apply each of the insertion methods to measure $g^{(2)}(r)$ and $g^{(3)}(r,s,t)$ in a simple uniform fluid, finding generally excellent agreement of the methods with each other and the distance-histogram approach, although the multiple-insertion methods are increasingly noisy at higher densities due to sampling difficulties. The improved resolution of the insertion approach is particularly useful in regions of $g^{(3)}(r,s,t)$ where the three particles are close together. The single-insertion method is expected to be useful in assessing the validity of approximations used to close the Yvon-Born-Green (YBG) hierarchy of integral equations. We also provide an approximate implementation to efficiently measure the full $g^{(3)}(r,s,t)$ up to a certain distance cut-off, which we will use in Chapter 7 to extend the predictor-corrector scheme of Chapter 5 to measure three-body interactions.

6.1 Introduction

We saw in Chapter 2 that the distribution functions $g^{(n)}$ are central to the statistical mechanical description of classical fluids, and can be used to quantify their structure [1]. In particular, the pair distribution function $g^{(2)}$ provides access to the thermodynamics of a fluid through the compressibility equation and, for pairwise-additive systems, the pressure and energy equations [1, 64, 116]. While this statement is formally correct, even simple monatomic liquids such as argon cannot be described adequately with pairwise interactions alone [133, 134, 230–232], and incorporating many-body interactions into the pressure and potential energy requires knowledge of higher-order distribution functions, where $n \geq 3$ [135]. Importantly for this thesis, measuring many-body interactions, using a scheme analogous to that of Chapter 5, requires consideration of higher-order distribution functions. Measuring $u^{(2)}$ and $u^{(3)}$, for example, requires matching the insertion and distance-histogram measurements of both $g^{(2)}$ and $g^{(3)}$; in general, all distribution functions up to $g^{(n)}$ must be matched to fully measure the interactions in a system with n-body interactions [127, 128].

Even in the pairwise-additive case, theoretical calculations of $g^{(2)}$ require consideration of many-body correlations [125, 233]; expressions for higher-order distribution functions in terms of lower ones can be used to close the Yvon-Born-Green (YBG) hierarchy in integral equation theories [119–126]. Properties depending on thermodynamic derivatives, such as the heat capacity, are expressed as integrals over many-body distribution functions [125, 135, 234], as is the configurational entropy [235–237]. Finally, a description of fluid structure at the pairwise level alone is inherently limited: simulations [135, 238] and colloidal studies [62, 118, 239] have found that the triplet distribution function $g^{(3)}$ shows crystal-like order at densities far lower than $g^{(2)}$ does. Disruption of this incipient crystalline order by impurities [118] or locally-preferred five-fold structures [240–244] has been discussed in relation to glass formation.

Previous measurements of many-body distribution functions from particle coordinates in simulations [238, 245–252] and colloidal experiments [62, 118, 239] have used the conventional distance-histogram approach. While this method usually provides adequate resolution for $g^{(2)}$ in uniform fluids, which depends only on the interparticle separation r, for many-body distribution functions its shortcomings are increasingly pronounced. Even in the uniform case, $g^{(3)}$ depends on three variables, and doubling the spatial resolution reduces the bin volumes by a factor of

eight, significantly increasing the statistical noise and limiting the available resolution [251]. For configurations where the particles approach closely, which are especially important in evaluating closure approximations [135, 249–255], the distribution functions vary rapidly and approximate interpolations must be performed—in this work, we find both linear and cubic interpolations to be inadequate for $g^{(3)}$.[1] Normalisation of the histograms is also increasingly complex, e.g. for $g^{(3)}$, the interparticle separations must satisfy triangle inequalities, which can intersect the bins used in a naive distance-histogram scheme [247, 251].

In this chapter, we address these problems by showing that the hierarchy of equations

$$\frac{g^{(n)}(\mathbf{s}^n)}{g^{(m)}(\mathbf{s}^m)} = \frac{P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m)}{\prod_{i=m+1}^{n} P^{(1)}(\mathbf{s}_i)}, \tag{2.59}$$

can be used to measure $g^{(n)}$ by test-particle insertion [100, 115]. A given $g^{(n)}$ can be obtained by multiple routes, which correspond to insertion of between 1 and n test particles. The normalisation is straightforward and, for insertion of n or $n-1$ test particles, the measurements are exact: in principle, $g^{(n)}$ can be measured with arbitrary resolution and for specific particle configurations. To begin, we develop the insertion methods explicitly for $g^{(2)}$ and $g^{(3)}$, and discuss the Kirkwood Superposition Approximation (KSA) closure of the YBG hierarchy. After detailing the implementation of the methods, we compare the distance-histogram measurements to those obtained using different numbers of insertions. In particular, we discuss the improved resolution of the insertion methods, which is useful in assessing closure approximations, and the increased difficulty of using multi-particle insertion methods at higher densities.

[1] These configurations are analogous to the hard-disk contact configurations considered in Chapter 4, where the distance-histogram method failed to correctly capture the rapidly-varying $g^{(2)}(r)$.

6.2 Theory

6.2.1 Two-Body Distribution Function

As in Section 2.4, we specialise (2.59) to the two-body case ($n = 2$) for a uniform, isotropic system, but here consider the results for all $m < n$:

$$g^{(2)}(\mathbf{s}_1, \mathbf{s}_2) = \frac{P^{(2)}(\mathbf{s}_1, \mathbf{s}_2)}{P^{(1)}(\mathbf{s}_1)P^{(1)}(\mathbf{s}_2)}, \qquad m = 0, \tag{6.1a}$$

$$= \frac{P^{(1)}(\mathbf{s}_2|\mathbf{s}_1)}{P^{(1)}(\mathbf{s}_2)}, \qquad m = 1, \tag{6.1b}$$

where we have noted that $g^{(0)} = g^{(1)} = 1$ and we recall the definition

$$P^{(n-m)}(\mathbf{s}^{(n-m)}|\mathbf{s}^m) = \left\langle \exp[-\beta\Psi(\mathbf{s}^{(n-m)})] \right\rangle_{\mathbf{s}^m}. \tag{2.56}$$

In a uniform system of isotropic particles, $P^{(1)}(\mathbf{s}_i) \equiv P$, while $g^{(2)}(\mathbf{s}_1, \mathbf{s}_2)$, $P^{(2)}(\mathbf{s}_1, \mathbf{s}_2)$ and $P^{(1)}(\mathbf{s}_2|\mathbf{s}_1)$ depend only on the separation $|\mathbf{s}_2 - \mathbf{s}_1| = r$, such that

$$g^{(2)}(r) = \frac{P^{(2)}(r)}{P^2}, \qquad m = 0, \tag{6.2a}$$

$$= \frac{P^{(1)}(r)}{P}, \qquad m = 1. \tag{6.2b}$$

As discussed in Section 2.4 and demonstrated in Chapters 4 and 5, the second of these equations can be evaluated directly using test-particle insertion: the numerator is evaluated by attempting insertions at a fixed separation r from a reference particle,[2] and the denominator is evaluated by considering insertions at all positions of the fluid.

The $m = 0$ result, however, suggests an alternative insertion route. The numerator $P^{(2)}(r)$ can be evaluated by considering test-insertion of a *pair* of particles with a fixed separation r at random positions in the fluid. The ensemble average is evaluated in the same manner as in the single-particle insertion case: first the additional potential energy Ψ is calculated by considering the interactions of the particles with each other and the existing particles of the fluid.[3] This is then used to calculate the exponential average in (2.56), averaging over equilibrium

[2] To preserve the simplicity of the notation, the conditional nature of $P^{(1)}(r)$, and later $P^{(2)}(r, s, t)$ and $P^{(1)}(r, s, t)$, is taken to be implicit.

[3] If the interaction between the inserted particles was *not* included when evaluating the ensemble

configurations of the fluid and insertion positions, including the orientation of the pair. Once $P^{(2)}(r)$ is determined, it is normalised by dividing by the *square* of the ensemble average P associated with inserting a single particle. This second insertion method is also formally exact and allows $g^{(2)}(r)$ to be obtained at arbitrary resolution.

6.2.2 Three-Body Distribution Function

Analogously to the two-body case, the hierarchy (2.59) suggests that $g^{(3)}$ can be measured by three different insertion routes, corresponding to insertion of up to three test particles. Substituting $n = 3$ into (2.59), for a uniform system of isotropic particles we obtain

$$g^{(3)}(r,s,t) = \frac{P^{(3)}(r,s,t)}{P^3}, \qquad m = 0, \tag{6.3a}$$

$$= \frac{P^{(2)}(r,s,t)}{P^2}, \qquad m = 1, \tag{6.3b}$$

$$= \frac{P^{(1)}(r,s,t)}{P} g^{(2)}(r), \qquad m = 2. \tag{6.3c}$$

In deriving these equations, we have noted that in this case, $g^{(3)}$ is simply a function of the triangle formed with a particle at each vertex; this triangle describes the relative positions of the particles and has the same role as the separation r in the two-body case. While there are several ways to specify a given triangle, in this thesis we use the three side lengths r, s and t.

Figure 6.1 summarises the application of the above equations to measure $g^{(3)}(r,s,t)$. In the first case, $P^{(3)}(r,s,t)$ is evaluated by test-insertions of a triangle of three particles at a random position and orientation in the fluid [Figure 6.1(a)]. In the second case, $P^{(2)}(r,s,t)$ is measured using two-particle test-insertions with an existing particle as the third vertex [Figure 6.1(b)]; in the third case $P^{(1)}(r,s,t)$ is obtained by test-insertions of a single particle around pairs of existing particles separated by one of the side lengths [Figure 6.1(c)]. In this final case, (6.3c) shows that is also necessary to multiply by $g^{(2)}(r)$, where r is the separation of the pair of existing particles. Since we have seen that $g^{(2)}(r)$ may be also be measured by single-particle test-insertions, this illustrates the principle that the hierarchy (2.59) can be used to express the $g^{(n)}$ as products of ensemble averages corresponding only to single-particle test-insertions, analogous to a *bootstrap* operation described by Percus [139]. The first two cases, which correspond to n and $n-1$ insertions, respectively, are exact and allow $g^{(3)}(r,s,t)$ to be measured for precise

average, the cavity distribution function $y^{(2)}(r)$ (see Section 4.2.4) would be calculated instead.

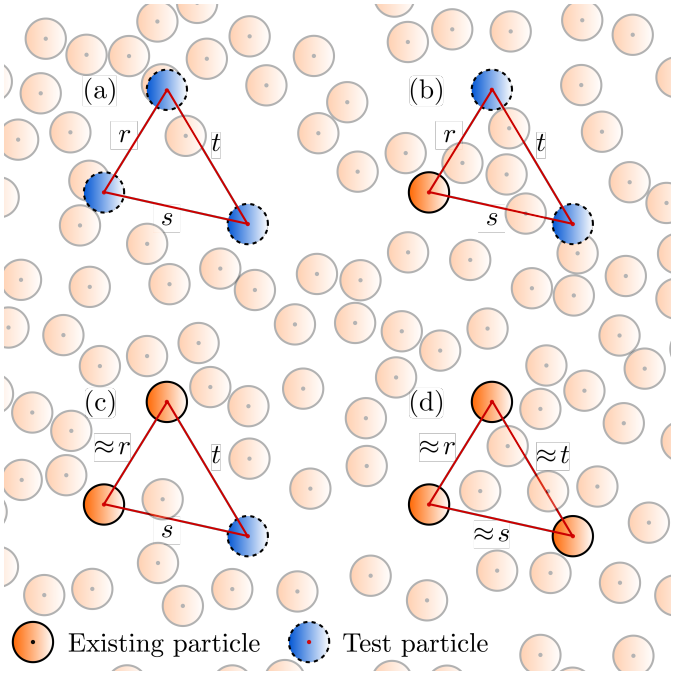


Figure 6.1: Applying (2.59) permits $g^{(3)}(r,s,t)$ to be measured by insertion of up to three test particles, (a)-(c), in addition to the conventional distance-histogram approach, (d). Methods (a) and (b) are exact, while (c) and (d) require histograms in one and three variables, respectively.

triangles, while the third requires a histogram in the side length r only, significantly improving the statistics compared with the full distance-histogram approach [Figure 6.1(d)]. For all of the insertion methods, normalisation is simply a straightforward division by the appropriate power of P, whereas the triangle inequalities must be carefully taken into account when using the distance-histogram method [247, 251].

Since $g^{(3)}(r,s,t)$ is a function of three variables, it is more difficult to visualise than $g^{(2)}(r)$. Typically only a subset of $g^{(3)}(r,s,t)$ is shown, either where the side lengths are functions of only one variable, e.g. for equilateral or isosceles triangles [238, 245–248, 250–252], or where two particles are held at fixed separations and $g^{(3)}(r,s,t)$ is shown as function of the position of the third [62, 118, 239, 249]. The insertion approach is efficient in both cases, since $g^{(3)}(r,s,t)$ can be measured directly for the subset of triangles without having to consider all triplets of particles in the system. We also provide an approximate implementation, analogous to the weighted-histogram approach of Section 5.3.1, for measuring the *complete* insertion $g^{(3)}(r,s,t)$ up to appropriate cut-off distances—this will be required in Chapter 7 when measuring three-body interactions.

6.2.3 Kirkwood Superposition Approximation

As discussed in Section 6.1, calculating $g^{(2)}$ theoretically requires consideration of many-body correlations even in the pairwise-additive case; this is clearly seen by considering the first non-trivial member of the YBG hierarchy for uniform, pairwise-additive fluids [135, 196]

$$-k_{\mathrm{B}}T\nabla_1 g^{(2)}(r_1) = g^{(2)}(r_1)\nabla_1 u^{(2)}(r_1) + \rho\int \mathrm{d}\mathbf{r}_2 g^{(3)}(\mathbf{0},\mathbf{r}_1,\mathbf{r}_2)\nabla_1 u^{(2)}(r_{12}), \tag{6.4}$$

where we have used an alternative specification of the triangle with particles at the origin $\mathbf{0}$, and $\mathbf{r}_1$ and $\mathbf{r}_2$. This integro-differential equation relates $g^{(2)}$ (unknown) to $g^{(3)}$ (unknown) and $u^{(2)}$ (known); its solution requires an additional *closure* relation, expressing $g^{(3)}$ in terms of the other two functions. An exact closure would yield $g^{(2)}$ and solve the classical problem of liquid state theory in the pairwise-additive case; while no exact closure is known, many approximate closures of varying degrees of sophistication have been employed [135, 249–254]. Closures for higher members of the hierarchy have also been considered, where, e.g. $g^{(4)}$ is expressed in terms of $g^{(3)}$ [120–123, 135].

Consider again equation (2.47)

$$\frac{g^{(n)}(\mathbf{r}^n)}{g^{(m)}(\mathbf{r}^m)} = \frac{\rho^{(n-m)}(\mathbf{r}^{(n-m)}|\mathbf{r}^m)}{\prod_{i=m+1}^{n}\rho^{(1)}(\mathbf{r}_i)}. \tag{2.47}$$

Noting that $g^{(1)}(\mathbf{r}_1) = 1$, in the case where $n = 3$ and $m = 1$ we obtain

$$g^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) = \frac{\rho^{(2)}(\mathbf{r}_2,\mathbf{r}_3|\mathbf{r}_1)}{\rho^{(1)}(\mathbf{r}_2)\rho^{(1)}(\mathbf{r}_3)}. \tag{6.5}$$

Multiplying and dividing by $\rho^{(1)}(\mathbf{r}_2|\mathbf{r}_1)\rho^{(1)}(\mathbf{r}_3|\mathbf{r}_1)$ yields [141]

$$g^{(3)}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) = \frac{\rho^{(2)}(\mathbf{r}_2,\mathbf{r}_3|\mathbf{r}_1)}{\rho^{(1)}(\mathbf{r}_2|\mathbf{r}_1)\rho^{(1)}(\mathbf{r}_3|\mathbf{r}_1)}\frac{\rho^{(1)}(\mathbf{r}_2|\mathbf{r}_1)}{\rho^{(1)}(\mathbf{r}_2)}\frac{\rho^{(1)}(\mathbf{r}_3|\mathbf{r}_1)}{\rho^{(1)}(\mathbf{r}_3)} \tag{6.6}$$

$$= g^{(2)}(\mathbf{r}_2,\mathbf{r}_3|\mathbf{r}_1)g^{(2)}(\mathbf{r}_1,\mathbf{r}_2)g^{(2)}(\mathbf{r}_1,\mathbf{r}_3), \tag{6.7}$$

where on the final line we have used (2.50) and the conditional distribution function

$$g^{(2)}(\mathbf{r}_2,\mathbf{r}_3|\mathbf{r}_1) = \frac{\rho^{(2)}(\mathbf{r}_2,\mathbf{r}_3|\mathbf{r}_1)}{\rho^{(1)}(\mathbf{r}_2|\mathbf{r}_1)\rho^{(1)}(\mathbf{r}_3|\mathbf{r}_1)}, \tag{6.8}$$

which quantifies the correlation of two particles, with positions $\mathbf{r}_2$ and $\mathbf{r}_3$, in the presence a third particle at $\mathbf{r}_1$. We will consider a very simple closure approximation known as the Kirkwood Superposition Approximation (KSA) [1, 116, 233], in which the effect of the third particle is ignored by replacing $g^{(2)}(\mathbf{r}_2, \mathbf{r}_3|\mathbf{r}_1)$ with $g^{(2)}(\mathbf{r}_2, \mathbf{r}_3)$ [256]:

$$g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \approx g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) g^{(2)}(\mathbf{r}_1, \mathbf{r}_3) g^{(2)}(\mathbf{r}_2, \mathbf{r}_3), \tag{6.9}$$

$$g^{(3)}(r, s, t) \approx g^{(2)}(r) g^{(2)}(s) g^{(2)}(t). \tag{6.10}$$

On the final line, we have rewritten the KSA in terms of the side lengths r, s, and t. Although this simple closure is known to give poor results [135, 256], here we use it to illustrate the advantages of the insertion approach in regions where closure approximations typically break down, and note that comparisons with more sophisticated closures can be performed [135, 249–254].

6.3 Methods

6.3.1 Implementation

Distance-Histogram Method

The distance-histogram measurement of $g^{(2)}(r)$ is performed according to the implementation in Section 5.3.1; the implementation for $g^{(3)}(r, s, t)$ is similar, with some additional complications arising in the normalisation. The method is based on directly applying (6.5), which, for a uniform fluid of isotropic particles, reads

$$g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \frac{\rho^{(2)}(\mathbf{r}_2, \mathbf{r}_3|\mathbf{r}_1)}{\rho^2}, \tag{6.11}$$

$$g^{(3)}(r, s, t) = \frac{\rho^{(2)}(r, s, t)}{\rho^2}, \tag{6.12}$$

where we have noted that $\rho^{(1)}(\mathbf{r}) \equiv \rho$ in a uniform system. On the final line, we have again used the side lengths r, s and t to specify the triangle, and have taken the presence of a particle at the origin in $\rho^{(2)}(r, s, t)$ to be implicit. As with $g^{(2)}(r)$, we use a distance cut-off d to avoid considering triangles with large side lengths, for which the short-ranged structure of the fluid has essentially decayed away. We find triangles with two side lengths $r \leq d$ and $s \leq d$; the third

side length $t \leq 2d$ due to the triangle inequality $t \leq r + s$. We then use the side lengths of these triangles to construct a 3D histogram, using the same bin width for each variable.

The symmetry of $g^{(3)}(r, s, t)$ with respect to interchange of the side lengths means that each triangle can be included six times (once for each permutation of $\{r, s, t\}$), although, in practice, we include those triangles with $t > d$ only twice due to the distance cut-off. We normalise the histogram using a two-dimensional version [239] of the procedure outlined in [247], which is used to calculate the volume of the region in each bin which satisfies the triangle inequalities. We then divide by the number of reference particles considered to obtain $\rho^{(2)}(r, s, t)$; dividing this by ρ^2 (measured separately) then gives $g^{(3)}(r, s, t)$ according to (6.12).

As discussed in Section 4.2.3, the distance-histogram method replaces infinitesimal volume elements with small but finite volumes, which leads the method to break down when the distribution function varies rapidly. This breakdown is more challenging to overcome when measuring $g^{(3)}(r, s, t)$, since increasing the resolution carries a more severe noise penalty due to the dependence on three variables. Unfortunately, the regions where $g^{(3)}(r, s, t)$ varies most rapidly are exactly those which are important in assessing closure approximations, and correspond to all three particles being close together. In these regions, the effect of the third particle on the correlations of the other two cannot simply be ignored, as in the KSA. The need to use larger bins to achieve reasonable statistics means that we cannot assume that the value of $g^{(3)}(r, s, t)$ is the same everywhere in the bin; instead, we obtain values away from the bin centres using linear and cubic interpolations [204, 257].

Insertion Method: Subsets Depending on a Single Variable

As discussed in Section 6.2.1, equations (6.2a) and (6.2b) provide two different routes to measuring $g^{(2)}(r)$ by test-particle insertion. For the single-insertion route, we measure $P^{(1)}(r)$ by attempting one insertion around each particle for each of the separations considered; we choose the angle of the inserted particle relative to the existing particle at random. We perform a separate measurement of P by attempting insertions at random points in the fluid, and then use (6.2b) to calculate $g^{(2)}(r)$. For the two-insertion route, we measure $P^{(2)}(r)$ by attempting insertions of pairs of particles with each separation at random locations and with random orientations. By considering the particles as separate entities, we can determine P using the same insertions. We then use (6.2a) to calculate $g^{(2)}(r)$.

In Section 6.2.2, we noted that often only subsets of $g^{(3)}(r,s,t)$ are considered where the side lengths are functions of a single variable. We consider three such subsets: the set of equilateral triangles, where $g^{(3)}(r,r,r)$ depends on a single side length r, and two different sets of isosceles triangles. In the first isosceles set (set A), two side lengths are fixed as r_m, corresponding to the maximum of $g^{(2)}(r)$, and the third side length s is allowed to vary.[4] In the second set (set B), one of side the lengths is fixed as r_m and the length s of the other two sides is varied. Equations (6.3a) to (6.3c) provide three insertion routes, which were summarised in Figure 6.1.

We implement the single-insertion route slightly differently in the equilateral and isosceles cases. In the equilateral case, we find all pairs of particles which are separated by less than a cut-off distance d, and perform test-insertions at the two points which complete an equilateral triangle. We then measure $P^{(1)}(r,r,r)$ by constructing a weighted histogram of these measurements. In the isosceles cases, we find pairs of particles separated by $\sim r_\mathrm{m}$ and, for each value of s, attempt either four (set A) or two (set B) test-insertions to complete the corresponding triangle [see Figure 6.1(c)].[5] Note that the triangle inequality limits the possible values to $s \leq 2r_\mathrm{m}$ (set A) or $s \geq r_\mathrm{m}/2$ (set B), and that for set A, we choose the second side length corresponding to r_m to be consistent with the exact separation of the pair. Averaging over all snapshots yields $P^{(1)}(r_\mathrm{m}, r_\mathrm{m}, s)$ (set A) or $P^{(1)}(r_\mathrm{m}, s, s)$ (set B). In all three cases, we measure P by attempting insertions at random points in the fluid and, along with the single-insertion measurement of $g^{(2)}(r)$, use it in (6.3c) to calculate $g^{(3)}$.

For the two-insertion route, we attempt insertions of a pair of particles around an existing particle, choosing the particles' positions to form a triangle with the appropriate side lengths [see Figure 6.1(b)]. We construct the triangles with a random orientation and, for each existing particle, attempt to insert one pair per separation [r (equilateral) or s (isosceles)]; these are then averaged to calculate $P^{(2)}$ for each case. For the three-insertion route, we measure $P^{(3)}$ by attempting insertions of triplets of particles (with appropriate side lengths) at random locations and with random orientations [see Figure 6.1(a)]. For both routes, we combine these measurements with a separate measurement of P, calculating $g^{(3)}$ according to either (6.3b) or (6.3a), respectively.

[4] This is analogous to the rolling-contact case in hard-sphere systems, where two particles are simultaneously in contact with a third [123, 255, 258–260].

[5] While pairs of particles with separations $\sim s$ may also be considered, we do not include them in this implementation.

Insertion Method: Weighted-Histogram Approach

As noted in Section 6.2.2, we also implement an approximate measurement of $g^{(3)}(r,s,t)$ which is analogous to the weighted-histogram insertion measurement of $g^{(2)}(r)$ implemented in Section 5.3.1. As in the two-body case, this approach sacrifices the considerable resolution advantage of the insertion methods discussed above, but is much more efficient in terms of speed and memory, and allows for a complete measurement of $g^{(3)}(r,s,t)$, up to an appropriate cut-off distance, which will be required in Chapter 7. We begin by calculating the insertion energies Ψ for a set of random points in the fluid—we use the same points for all three insertion methods. Similar to the two-body case, where each insertion point could be considered as an insertion about multiple reference particles, in measuring $g^{(3)}(r,s,t)$ each insertion point can partake in multiple triangles formed by combinations of insertion points and existing particles. We then use a weighted-histogram procedure to calculate $P^{(1)}(r,s,t)$, $P^{(2)}(r,s,t)$ and $P^{(3)}(r,s,t)$ for each bin.

For the one-insertion route, we begin by finding all triangles formed by two particles and one insertion point [see Figure 6.1(c)] satisfying the distance criteria, which we choose to be the same as those for the distance-histogram method. We use these triangles to construct two histograms, again using the same bins used in the distance-histogram method.[6] The *number histogram* is just a straightforward histogram of the triangles, while the *insertion histogram* weights each triangle by $\exp[-\beta\Psi]g^{(2)}(r)$, where r is the separation of the two existing particles and Ψ is the insertion energy of the inserted particle. Note that we use the distance-histogram measurement of $g^{(2)}(r)$, but an insertion measurement could equally be used. We divide the insertion histogram by the number histogram to obtain $P^{(1)}(r,s,t)g^{(2)}(r)$ for each bin, and evaluate P by averaging $\exp[-\beta\Psi]$ over all the insertion points. Finally, we calculate $g^{(3)}(r,s,t)$ using (6.3c).

For the two-insertion route, we find the triangles formed by a single particle and two insertion points, while for the three-insertion route we consider triangles formed by three insertion points—once again, these triangles must satisfy the distance criteria. In the insertion histogram, we weight these triangles by $\exp[-\beta\Psi]$; Ψ is the total insertion energy of the two or three inserted particles, which includes their interactions with the existing particles and with each other.

[6] As with the distance-histogram method, the symmetry of $g^{(3)}(r,s,t)$ means that the same triangle can be included in the histograms multiple times.

Dividing by the number histogram yields $P^{(2)}(r,s,t)$ or $P^{(3)}(r,s,t)$, respectively, which we then divide by the appropriate power of P, giving $g^{(3)}(r,s,t)$ according to (6.3b) or (6.3a).

6.3.2 Protocols

Simulation

We generated equilibrium configurations using the Grand Canonical Monte Carlo scheme (Section 3.3.2). The interparticle interactions were pairwise according to a WCA Potential [261]:

$$u^{(2)}_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon, & r \leq 2^{\frac{1}{6}}\sigma, \\ 0, & r > 2^{\frac{1}{6}}\sigma, \end{cases} \tag{6.13}$$

where r is the interparticle separation. We used a square simulation box of side length 50σ, with the parameters $\sigma = 1$, $\epsilon = 1$, and inverse temperature $\beta = 1$. We include results for two simulations, with simulation chemical potentials $\mu' = 1.25$ and $\mu' = 4$ [see (5.9) of Section 5.3.2], which gave rise to mean numbers of particles $\overline{N} \approx 1017$ ($\rho\sigma^2 \approx 0.41$) and $\overline{N} \approx 1507$ ($\rho\sigma^2 \approx 0.60$), respectively.[7] For the particle movements, we use maximum displacements δ of 0.6 and 0.28, respectively, which corresponded to acceptance probabilities of approximately 50%. We initially placed 10 particles at random positions in the box, before allowing the density to increase during 10^7 equilibration steps. We then took 5×10^4 snapshots at intervals of 10^4 steps.

Analysis

We used all 5×10^4 snapshots throughout the analysis. For the distance-histogram measurements of $g^{(2)}(r)$, we used a distance cut-off $d = 10\sigma$ and a bin width 0.01σ; for $g^{(3)}(r,s,t)$ we used $d = 5\sigma$ and a bin width 0.05σ.

For the single-insertion $g^{(2)}(r)$ measurements we used separations corresponding to the centres of the bins used in the $g^{(2)}(r)$ distance-histogram measurements (every 0.01σ), up to 5.005σ. We used 10^4 insertion points per snapshot to measure P. For the two-insertion $g^{(2)}(r)$ measurements we used separations corresponding to the centres of the bins used in the $g^{(2)}(r,s,t)$

[7] A third, higher density, simulation was performed with $\overline{N} \approx 1936$ ($\rho\sigma^2 \approx 0.77$), but the full analysis was only completed after the submission of the thesis [262].

distance-histogram measurements (every 0.05σ), up to 5.025σ; in each snapshot, we inserted 10^3 pairs per separation.

For the single-insertion $g^{(3)}$ measurements, we used $d = 5\sigma$ with a bin width of 0.01σ in the equilateral case. For the isosceles cases, we determined that $r_{\mathrm{m}} \approx 1.105\sigma$ in the $\rho\sigma^2 \approx 0.41$ simulation, and considered all pairs with separations $1.10\sigma \leq r < 1.11\sigma$; $r_{\mathrm{m}} \approx 1.085\sigma$ for the $\rho\sigma^2 \approx 0.60$ simulation, and we considered all pairs with separation $1.08\sigma \leq r < 1.09\sigma$. The separations s were centres of the bins with bin width 0.01σ, with the triangle inequality limiting the possible values to $s \leq 2r_{\mathrm{m}}$ (set A) or $s \geq r_{\mathrm{m}}/2$ (set B). For the two-insertion and three-insertion measurements, the separations [r (equilateral) or s (isosceles)] were centres of the bins with bin width 0.05σ. For the three-insertion measurements, in each snapshot we inserted 10^3 triangles for each separation. In all cases, we used 10^4 insertion points per snapshot to measure P.

Finally, for the measurements of $g^{(3)}(r, s, t)$ using the weighted-histogram approach, we used 4×10^3 insertion points in each snapshot. For the histogram, the same bins were used as in the $g^{(3)}(r, s, t)$ distance-histogram measurement.

6.4 Results and Discussion

6.4.1 Subsets Depending on a Single Variable

Comparison of Insertion Approaches

We show the results for $g^{(2)}(r)$ and the subsets of $g^{(3)}$ depending on a single variable in Figure 6.2, for the fluid with $\rho\sigma^2 \approx 0.41$, and Figure 6.3, for $\rho\sigma^2 \approx 0.60$; in the insets we show a zoom of the region around the primary maximum of each measurement. In Chapter 4, we found excellent agreement between the distance-histogram and single-insertion measurements of $g^{(2)}(r)$ for hard-disk fluids, with the only exceptions occurring at contact where $g^{(2)}(r)$ varies rapidly. The results in panel (a) of each figure show that this agreement extends to continuous pair potentials, and confirm the expectation that the results of the two-insertion method agree with those of the single-insertion method.

We show the results for $g^{(3)}$ in panels (b)-(d); note that in the isosceles cases [panels (c)

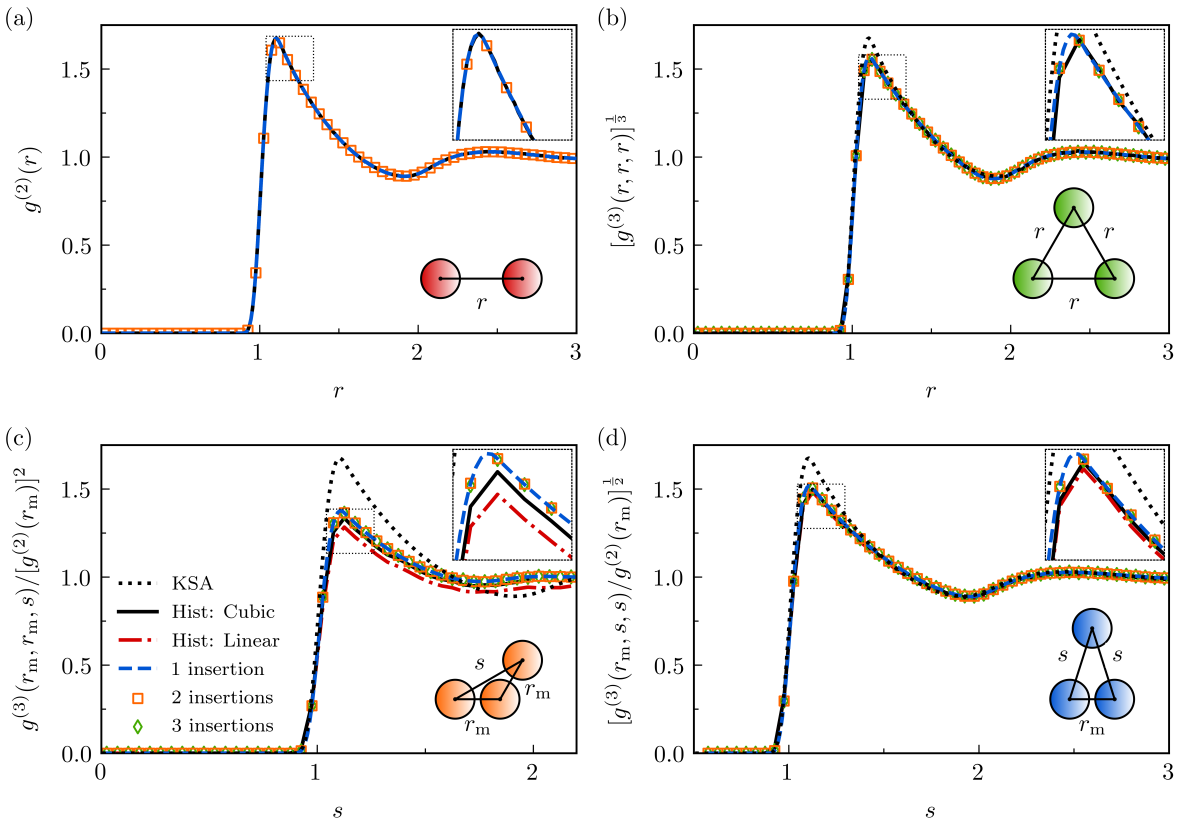


Figure 6.2: Comparison of $g^{(2)}$, (a), and $g^{(3)}$ for equilateral and isosceles triangles, (b)-(d), measured in the lower-density fluid ($\rho\sigma^2 \approx 0.41$) by the distance-histogram method and each insertion method. The insets show a zoom of each plot around the primary maximum. The ordinates of (b)-(d) facilitate comparison with the KSA prediction $g^{(2)}$. In (c) and (d), the side length r_m corresponds to the maximum of $g^{(2)}(r)$ and was not at the centre of a bin used in the distance-histogram method—instead, the results of linear and cubic interpolation are shown. Note that no such interpolation was required in (b) (see note 9 on page 101).

and (d)], the range of values for the side length s is limited by the triangle inequality. We find the agreement of the insertion results with each other, and with those of the distance-histogram method (with the cubic interpolation), to be generally excellent, confirming the validity of (2.59) and the insertion methods described above. In the high-density case, while the two-insertion and three-insertion results are more noisy, the underlying function appears to show the expected agreement with the other results. This increased noise is also apparent in the two-insertion measurement of $g^{(2)}(r)$.

The increased noise of multiple-insertion methods at higher densities is symptomatic of the well-documented breakdown of test-particle insertion approaches in high-density systems [199]. In brief, ensemble averages of the form $P = \langle \exp[-\beta\Psi] \rangle$ are increasingly difficult to evaluate by

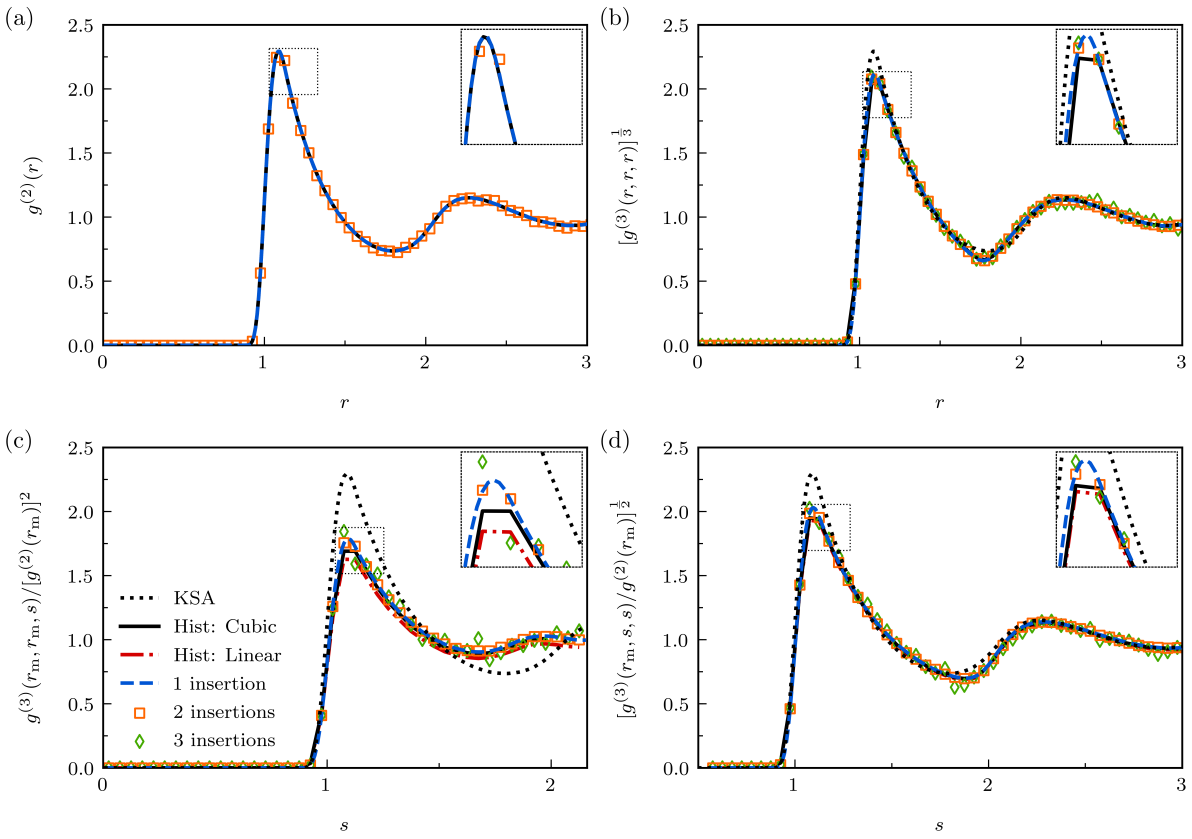


Figure 6.3: The same measurements shown in Figure 6.2, except for the higher-density fluid ($\rho\sigma^2 \approx 0.60$).

test-particle insertion since small or negative values of Ψ are sampled only very rarely, despite making an important contribution to the ensemble average. In essence, it becomes 'too difficult' to insert an additional particle, and this problem is exacerbated when inserting multiple particles. In Chapter 4, we found that the single-particle insertion approach works well to measure $g^{(2)}(r)$ in most of the fluid region, and we expect this to hold for $g^{(3)}(r,s,t)$ and higher-order distribution functions. We note, however, that the single-particle insertion method is only formally exact in the case of $g^{(2)}(r)$. Moreover, when the side lengths all correspond to strong overlap, the absence of such pairs in the equilibrium configurations means that $P^{(1)}(r,s,t)$ cannot be measured. This is the case for equilateral triangles at small r [panel (b)]. In practice this limitation is of no consequence, since at such separations $g^{(2)}(r) \approx 0$ and so $g^{(3)}(r,s,t) \approx 0$ according to (6.3c).[8]

[8]For continuous $u^{(2)}(r)$ such as that used here, $g^{(2)}(r)$ for strongly-overlapping r is finite but *extremely* small, and so is measured as zero in practice.

Resolution

The isosceles cases [panels (c) and (d)] illustrate the effects of the poor resolution of the distance-histogram method when measuring $g^{(3)}(r, s, t)$ where, to obtain adequate statistics, larger bins are required than when measuring $g^{(2)}(r)$. Since r_{m} does not coincide with a bin centre, an approximate interpolation in (r, s, t)-space is necessary to provide values for comparison with the insertion measurements and the KSA.[9] We show the results of two interpolation schemes: although the cubic interpolation outperforms the simple linear interpolation, we still find marked deviation from the insertion results, particularly for set A [panel (c)] where the particles are always close together and so $g^{(3)}$ varies rapidly over the bin width. Moreover, the cubic interpolation fails when s is too close to the edge of the domain (i.e. 0 or the limits given by the triangle inequality), since here the scheme requires values for bins where $g^{(3)}(r, s, t)$ is not defined. We expect these issues with resolution and interpolation to be still more pronounced for higher-order distribution functions, which depend on more variables and have more complicated domains.

By contrast, while the two-insertion and three-insertion measurements may be noisy at high-density due to the *sampling* issues discussed above, no issues arise with resolution and these methods are formally exact. The single-insertion method requires a histogram in only *one* of the side lengths, and so we expect any resolution issues to be no more pronounced than with the distance-histogram method for $g^{(2)}(r)$. We performed the single-insertion measurement with $5\times$ greater resolution than with the other methods, and this measurement can be considered the benchmark for comparison. The implementation of the single-insertion measurement is particularly efficient: the method does not suffer from sampling problems to the same degree as the multiple insertion-methods, and, e.g. in the isosceles cases, insertions are only attempted around pairs with separation $r \approx r_{\mathrm{m}}$. This generally leads to fewer attempted insertions per snapshot than with the two-insertion and three-insertion methods, with the insertions being, in some sense, very 'well-targeted'.

Kirkwood Superposition Approximation

For each plot, we choose the ordinate to facilitate comparison with $g^{(2)}(r)$ [panel (b), equilateral] or $g^{(2)}(s)$ [panels (c) and (d), isosceles], which is the KSA prediction. We find the KSA to be

[9] In the equilateral case, the side lengths do coincide with the bin centres, and no interpolation is required.

an excellent approximation when the particles are well-separated, which corresponds to large separations for the equilateral triangles [panel (b)] and isosceles set B [panel (d)]. When the particles are all close together, i.e. for the primary maxima of these cases (and the subsequent minima for the higher density), or in isosceles set A [panel (c)], the KSA prediction is much worse. This is consistent with the KSA assumption that the conditional two-body distribution function is unaffected by the presence of a third particle, which is reasonable when the third particle is far away but increasingly poor as it approaches the other two. An improved closure approximation is required to capture the behaviour of $g^{(3)}(r, s, t)$ at such separations. While we do not examine more sophisticated closure approximations here, these results illustrate the utility of the insertion approach in this task: the regions where a more sophisticated closure approximation is required are also those where the poor resolution of the distance-histogram method has the greatest effect.

6.4.2 Subsets with Two Particles at a Fixed Separation

We show the results for subsets where two of the particles are held at a fixed separation as density plots in Figure 6.4; these measurements were performed with the weighted-histogram implementation. In each case, we follow Reference [118] and choose the separations to be the bin centres closest to the maxima and minima of $g^{(2)}(r)$. As with $g^{(2)}(r)$, the structural interpretation of $g^{(3)}$ in a uniform fluid is found by considering (2.50) when $n = 3$ [142]:

$$\frac{g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)}{g^{(2)}(\mathbf{r}_1, \mathbf{r}_2)} = \frac{\rho^{(1)}(\mathbf{r}_3 | \mathbf{r}_1, \mathbf{r}_2)}{\rho}. \tag{6.14}$$

We therefore plot $g^{(3)}(r, x_3, y_3)/g^{(2)}(r)$, which corresponds to the deviation of $\rho^{(1)}(x_3, y_3)$ from the uniform density ρ due to the presence of the fixed pair. For a given r, we use the side lengths (s, t) to calculate the cartesian coordinates (x_3, y_3) of the third particle; since these points do not form a regular grid, we use a Delaunay triangulation and a cubic spline to interpolate the function [263]. A uniform grid in (s, t)-space corresponds to a poor sampling of points near the x_3-axis, which is most apparent from the slightly-flattened edges of the black circular regions. We use an upper and lower cut-off on the colour scale to ensure that the interesting features are well-illustrated. We plot the results of each method in a single quadrant, but these correspond to the results for the whole space due to the reflection symmetry of $g^{(3)}(r, x_3, y_3)$ in the lines $x_3 = 0$ and $y_3 = 0$.

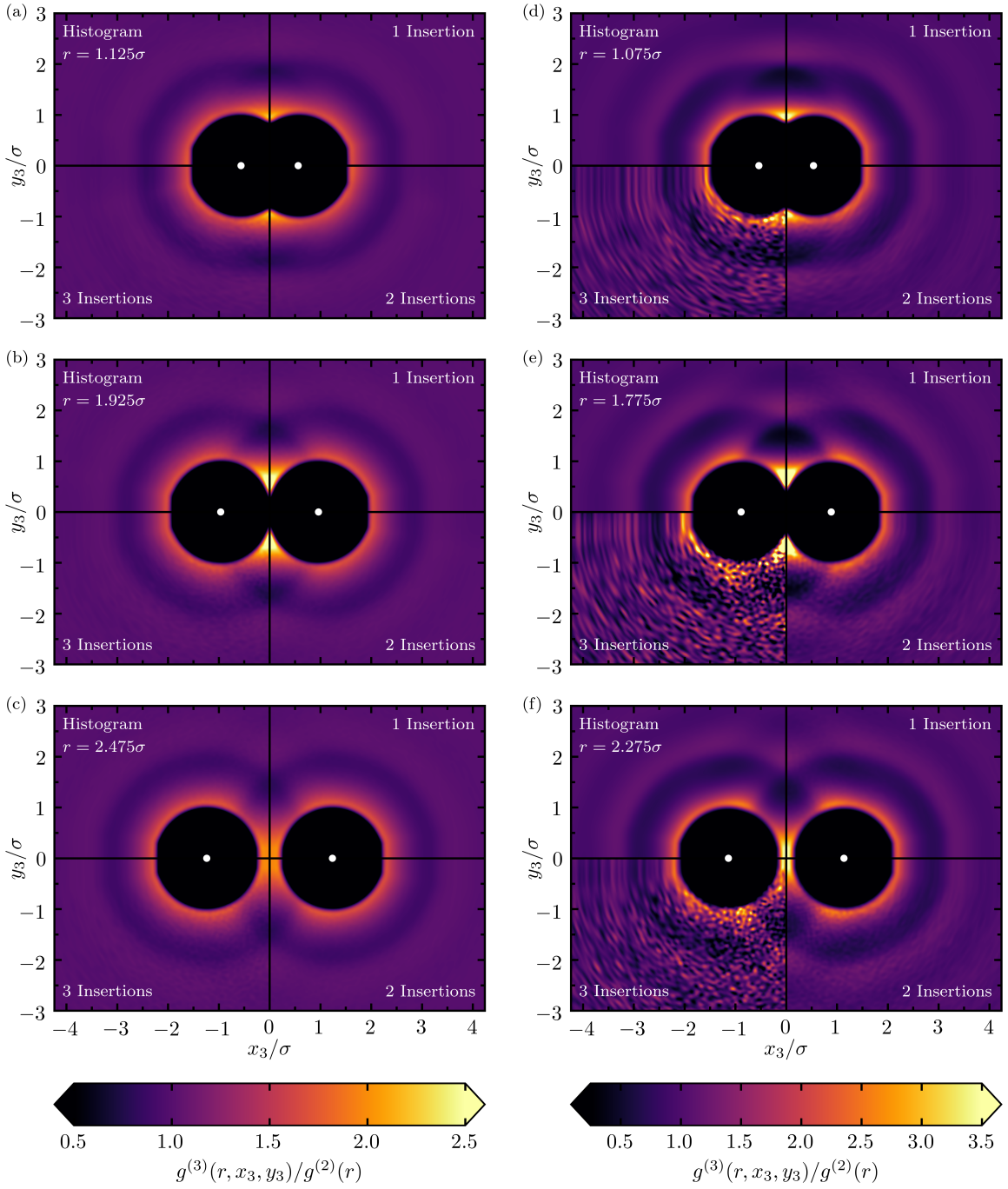


Figure 6.4: The conditional one-body density (normalised by the bulk density) of a third particle [with coordinates (x_3, y_3)] around a pair of particles (white circles) with separation r, corresponding to an extremum in $g^{(2)}(r)$. (a)-(c) are for the lower-density fluid ($\rho\sigma^2 \approx 0.41$), and (d)-(f) are for the higher-density fluid ($\rho\sigma^2 \approx 0.60$). In each case, measurements are shown for the distance-histogram method and the three insertion methods. Note that $g^{(3)}(r, x_3, y_3)$ has reflection symmetry in the lines $x_3 = 0$ and $y_3 = 0$.

For the low-density simulation [Figure 6.4(a)-(c)], we find excellent agreement between the results of each method, which further validates the insertion methods and demonstrates the legitimacy of the weighted-histogram approach. At higher density [Figure 6.4(d)-(f)], the broad features of each measurement are in agreement, but the noise of the multi-insertion methods is increasingly apparent, especially in the three-insertion case. The agreement between the distance-histogram and single-insertion measurements suggests that a PC scheme analogous to that of Chapter 5, but for measuring three-body interactions, will be feasible by matching these results. Note that at these relatively low densities, the measurements show little indication of incipient crystalline order.

6.5 Conclusion and Outlook

We have shown that the Potential Distribution Theorem provides n possible insertion methods for measuring an n-body distribution function $g^{(n)}(\mathbf{r}^n)$, and have demonstrated their use in measuring $g^{(2)}(r)$ and $g^{(3)}(r,s,t)$ in a simple uniform fluid. We have provided two implementations of each insertion method: a targeted approach which allows certain subsets of $g^{(3)}$ to be measured efficiently, and an approximate weighted-histogram approach which allows for a full measurement of $g^{(3)}(r,s,t)$ up to a certain cut-off distance. We have found that the agreement of the insertion methods with each other and the distance-histogram method is generally excellent. In particular, the insertion-based methods offer improved resolution and a more straightforward normalisation than the distance-histogram method, although methods based on inserting multiple particles suffer from increased noise at higher densities. While in this chapter the demonstration was for a uniform, single-component, pairwise-additive fluid, we stress that the methods can be extended to inhomogeneous fluids, as well as fluids with multiple components and many-body interactions.

The improved resolution of the insertion approach is most advantageous in regions where the particles are close together: we therefore expect the insertion methods to be especially useful in examining the validity of closure approximations used in integral equation theories. Furthermore, they will facilitate investigations into expressions for higher-order distribution functions, such as $g^{(4)}$, which can be used to close the YBG hierarchy at higher levels [120–123, 135]. Finally, the particular success of the single-insertion approach, combined with its

approximate implementation allowing for a full $g^{(3)}(r, s, t)$ measurement, suggests the possibility of extending the PC scheme of Chapter 5 to measure three-body interactions. We turn to this task in Chapter 7.

Chapter 7

Measuring Three-Body Interactions from Equilibrium Configurations

We incorporate systems with three-body interactions into the framework developed in this thesis, considering two different approaches. We show that applying the predictor-corrector (PC) scheme of Chapter 5 in systems with three-body interactions permits a coarse-grained *effective pair potential* $u_{\rm eff}^{(2)}(r)$ to be measured. Using MC simulations, we show explicitly that using $u_{\rm eff}^{(2)}(r)$ in the pairwise-additive approximation yields a $g^{(2)}(r)$ which agrees well with that obtained using the full three-body description, but does not yield the correct $g^{(3)}(r,s,t)$. Thus while $u_{\rm eff}^{(2)}(r)$ may offer an indication of the sign and range of $u^{(3)}(r,s,t)$, it is unable to reproduce all of the properties of the original fluid. We therefore also show that the PC scheme can be extended to measure $u^{(3)}(r,s,t)$, by matching the insertion and distance-histogram measurements of $g^{(3)}(r,s,t)$. We implement this scheme to measure $u^{(3)}(r,s,t)$ in the case that $u^{(2)}(r)$ is known, finding good agreement with the simulation input for both attractive and repulsive three-body interactions, as well as in the pairwise-additive case where $u^{(3)}(r,s,t) = 0$. The measurements are, however, rather noisy, and require further development before they can be used in colloidal systems and achieve the same quantitative accuracy obtained when measuring $u^{(2)}(r)$.

7.1 Introduction

In the last three chapters, we have treated systems as pairwise-additive, with the total potential energy given by a sum of interactions between each pair of particles in the system:

$$U_N(\mathbf{r}^N) = \sum_i \sum_{j>i} u^{(2)}(\mathbf{r}_i, \mathbf{r}_j). \tag{2.16}$$

While this model is widely-used and qualitatively captures many essential features of simple fluids [1], it is approximate even for systems of charged elementary particles such as electrons [133]. In general, the interaction between two particles is altered by the presence of nearby particles. This gives rise to higher-order interaction terms which depend on the positions of several particles, such that

$$U_N(\mathbf{r}^N) = \sum_i \sum_{j>i} u^{(2)}(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j} u^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \cdots \tag{2.15}$$

for a uniform fluid.[1]

Higher-order interactions typically arise because a subset of the degrees of freedom of the system have been ignored or integrated out to give a coarse-grained pair potential [93,149]. Even for simple monatomic fluids, such as liquid argon, three-body Axilrod-Teller interactions must be included to give quantitative agreement with experimental data [134,230–232]; these interactions are associated with the electronic degrees of freedom [149], and arise because the instantaneous dipole of a third atom alters the dipoles of an interacting pair. Similar effects are present between colloidal particles. For example, when describing the interactions of superparamagnetic particles in Chapter 3, we neglected local variations of the magnetic field due to the induced magnetic dipoles on each particle: the presence of a third particle will slightly modify the induced dipoles of two interacting particles, leading to an altered interaction. The interactions of charged-stabilised colloidal particles (Section 3.1.1) are another example: they are pairwise-additive only within the assumptions of the linearised Poisson-Boltzmann theory, and appreciable three-body interactions have been predicted [149, 264] and measured [87, 225, 265] for systems with large Debye lengths ${\kappa_\mathrm{D}}^{-1}$. In this case, the three-body interactions are associated with the ionic

[1] Treating the many-body interactions as corrections to the sum of pairwise interactions is appropriate providing the pairwise interaction is only slightly altered by the presence of other particles [133].

degrees of freedom [149].

We consider two options for incorporating higher-order interactions into the framework developed in this thesis. The first simply ignores the explicit three-body term and uses the pair potential measurement of Chapter 5 to obtain an *effective pair potential* $u_{\text{eff}}^{(2)}(r)$. In general, this will depend on the state of the system [37]: at higher density, for example, the influence of many-body terms will be greater [93]. While they may offer information on, e.g. the sign and range of the many-body interactions and are widely used, they give rise to inconsistent thermodynamic results [37], and a simulation using $u_{\text{eff}}^{(2)}(r)$ cannot be expected to reproduce the correct higher-order distribution functions beyond $g^{(2)}(r)$ [127, 128]. The second option considered extends the predictor-corrector (PC) scheme of Chapter 5 to the many-body case. In principle, this involves simultaneously inverting both $g^{(2)}(r)$ and $g^{(3)}(r,s,t)$ to yield $u^{(2)}(r)$ and $u^{(3)}(r,s,t)$, however, due to numerous challenges in the implementation, we focus on the more limited goal of measuring $u^{(3)}(r,s,t)$ when $u^{(2)}(r)$ is known.

Measuring $u^{(3)}(r,s,t)$ in atomic and molecular systems is challenging since only pairwise structural data is typically available from scattering studies, while inversion of three-body distribution functions is required to find $u^{(3)}(r,s,t)$ [127, 265]. Effective pair potentials have been derived [91,218,266], and careful small-angle neutron scattering experiments permit the strength of the London dispersion (two-body) and Axilrod-Teller (three-body) interactions to be measured in noble-gas fluids [267–273]. Effective pair potentials have also been derived in colloidal systems [93, 274, 275]; in this case, however, the colloidal positions are available using optical microscopy, enabling distribution functions, and thus interactions, to be measured to arbitrarily high orders in principle (see Chapter 6). Two approaches have been taken: $u^{(3)}(r,s,t)$ has been measured directly, by using optical tweezers to manipulate an assembly of three particles [87, 265, 276], and also by using a simulation-based scheme to invert $g^{(2)}(r)$ and $g^{(3)}(r,s,t)$ measured in an equilibrium fluid [225]. Our approach is similar to the latter, but instead compares distribution functions measured using the distance-histogram and insertion methods from the same configurational data.

We first introduce a model three-body interaction which, although not physically motivated, has the advantages of being simple and short ranged. We then discuss the meaning of effective pair potentials derived using the PC scheme of Chapter 5, and outline the extension of the PC scheme, inverting $g^{(3)}(r,s,t)$ to find $u^{(3)}(r,s,t)$. We apply each method in simulations

with attractive, repulsive and no three-body interactions, showing explicitly the deficiency of the measured effective pair potentials in reproducing $g^{(3)}(r,s,t)$, and demonstrating the measurement of $u^{(3)}(r,s,t)$ in the case where $u^{(2)}(r)$ is known.

7.2 Theory

7.2.1 A Simple Three-Body Interaction

As in the pairwise-additive case (Chapter 5), we test the methods developed in this chapter using simulations where the interactions are known. We introduce the simple three-body interaction

$$u^{(3)}(r,s,t) = \begin{cases} \zeta\left[\left(\frac{\sigma}{q(r,s,t)}\right)^3 - \frac{1}{8}\right], & q(r,s,t) \leq 2\sigma, \\ 0, & q(r,s,t) > 2\sigma, \end{cases} \tag{7.1}$$

where ζ is a constant, specifying the strength and sign of the interaction, and $q(r,s,t)$ is the diameter of the circumscribed circle of the triangle with edges r, s and t, which again correspond to the three interparticle distances (see Chapter 6). While the form of this interaction is not physically motivated, it is short-ranged, which is convenient for efficient simulations and analysis, and only gives rise to strong interactions when all three particles are close together, in line with predictions for atomic and colloidal systems [134, 264]. We supplement this interaction with a purely-repulsive pairwise interaction, as used in Chapter 6.

7.2.2 Effective Pair Potentials

Consider the insertion route to $g^{(2)}(r)$ in a uniform fluid:

$$g^{(2)}(r) = \frac{\langle \exp\left[-\beta\Psi(r)\right]\rangle}{\langle \exp\left[-\beta\Psi\right]\rangle}, \tag{2.62}$$

where Ψ is the additional potential energy of the system due to the hypothetical test-particle. While in Chapter 5 we assumed pairwise-additive interactions and Ψ was simply a sum of pairwise interactions, (2.62) is general and also holds for systems with many-body interactions.

In this chapter, we also consider three-body interactions and so

$$\Psi = \sum_i u^{(2)}(\mathbf{r}_\mathrm{t}, \mathbf{r}_i) + \sum_i \sum_{j<i} u^{(3)}(\mathbf{r}_\mathrm{t}, \mathbf{r}_i, \mathbf{r}_j), \tag{7.2}$$

where $\mathbf{r}_\mathrm{t}$ is the position of the test particle and the sums run over the existing particles of the system. To find an effective pair potential $u^{(2)}_\mathrm{eff}$, we simply replace the full Ψ with a sum of pairwise terms

$$\Psi_\mathrm{eff} = \sum_i u^{(2)}_\mathrm{eff}(\mathbf{r}_\mathrm{t}, \mathbf{r}_i), \tag{7.3}$$

and use the PC scheme of Chapter 5.

Unlike the true pair potential $u^{(2)}$, this effective pair potential depends on the state of the system, i.e. its temperature and density. At higher densities [93], it is likely that more particles will surround an interacting pair, leading to a greater deviation of $u^{(2)}_\mathrm{eff}$ from $u^{(2)}$: in general a separate $u^{(2)}_\mathrm{eff}$ will need to be measured at each density. Moreover, even for a given density, no $u^{(2)}_\mathrm{eff}$ can reproduce all the properties of the fluid: e.g. Louis has shown that the effective pair potential which gives the correct $g^{(2)}(r)$ will not give the correct potential energy using the energy equation (2.64) [37], nor will it reproduce the correct $g^{(3)}(r, s, t)$ [127,128]. He points out that these results agree with the earlier conclusion of Barker et al. [231]:

> We record our opinion that the use of density-dependent effective pair potentials can be misleading unless it is recognized that these are mathematical constructs to be used in specified equations rather than physical quantities.

As noted in Section 5.1, Chayes and Chayes have shown that even in the presence of many-body interactions, the effective pair potential giving rise to a particular $g^{(2)}(r)$ at a given density is unique [217]. While it is tempting to identify this with $u^{(2)}_\mathrm{eff}$ found using the PC scheme of Chapter 5, this would miss a subtle point. To see this, we begin by writing (2.62) using the trace notation of (2.12)

$$g^{(2)}(r) = \frac{\mathrm{Tr_c}\left[z^N \exp\left[-\beta U_N(\mathbf{r}^N)\right] \exp\left[-\beta\Psi(r)\right]\right]}{\mathrm{Tr_c}\left[z^N \exp\left[-\beta U_N(\mathbf{r}^N)\right] \exp\left[-\beta\Psi\right]\right]}, \tag{7.4}$$

where we have used the configurational PDF of the grand ensemble given by (2.8).[2] When

[2]For the purposes of this discussion, $\Psi(r)$ in the numerator is taken to be the insertion energy at a

measuring $u_{\text{eff}}^{(2)}$ using the PC scheme, the higher-order interactions are included explicitly in U_N, since the sample of configurations is drawn from the full configurational PDF. The insertion energy Ψ, however, is taken to be a sum of pairwise interactions depending on $u_{\text{eff}}^{(2)}$ and is given by (7.3). This is contrary to the derivation of (2.62) in Section 2.4, where the potential energy U_N is split into $U_{N-(n-m)}$ and Ψ ($n = 2$ and $m = 1$ here), which are each expected to contain interactions of the same functional form.

Essentially, the test-particle interacts as though it were of a different component: while the existing particles interact with the full interactions (here $u^{(2)}$ and $u^{(3)}$), their interactions with the test-particle are only pairwise ($u_{\text{eff}}^{(2)}$). The insertion measurement of $g^{(2)}(r)$ actually measures the *partial* pair distribution function between the existing-particle component and the test-particle component, with the test-particle component in infinite dilution (see Section 4.2.5). The measured $u_{\text{eff}}^{(2)}$ is that which matches this partial pair distribution function with the distance-histogram $g^{(2)}(r)$, measured in the single-component fluid with many-body interactions. It is therefore not clear that a pairwise-additive fluid interacting according to this $u_{\text{eff}}^{(2)}$ would have the correct $g^{(2)}(r)$, since in this case $u_{\text{eff}}^{(2)}$ would also determine the configurational PDF. In this chapter, we will test this explicitly using simulations, and also examine whether $u_{\text{eff}}^{(2)}$ can reproduce the correct three-body distribution function $g^{(3)}(r, s, t)$.

7.2.3 Three-Body Interactions: Predictor-Corrector Scheme

In principle, extending the PC scheme of Chapter 5 to enforce consistency between the distance-histogram and insertion measurements of $g^{(2)}(r)$ and $g^{(3)}(r, s, t)$ would allow for a simultaneous measurement of $u^{(2)}(r)$ and $u^{(3)}(r, s, t)$, fully characterising the interactions of the systems studied in this chapter. Due to the difficulty of implementing such a scheme, however, we focus here on measuring $u^{(3)}(r, s, t)$ when $u^{(2)}(r)$ is already known—experimentally, this could be useful in the case where $u^{(2)}(r)$ has been measured at low density, where three-body interactions are less important [90, 93]. In this case, only $g^{(3)}(r, s, t)$ needs to be considered: in Chapter 6, we saw that in addition to the distance-histogram method, there are three possible insertion routes to

particular relative position from an existing particle (with separation r), and Ψ in the denominator is the insertion energy at a *particular* point in the fluid. While this does not reflect the full averaging used in implementing the insertion method, it does give equivalent results when averaged over all configurations due to the symmetries of the fluid. This subtlety is not important here.

$g^{(3)}(r,s,t)$. We use the single-insertion method based on

$$g^{(3)}(r,s,t) = \frac{P^{(1)}(r,s,t)}{P} g^{(2)}(r), \qquad m = 2, \tag{6.3c}$$

for the insertion measurement, since this was found to have the lowest noise in Chapter 6. The ensemble averages in (6.3c) depend on $u^{(2)}(r)$ and $u^{(3)}(r,s,t)$: our goal is to find $u^{(3)}(r,s,t)$ such that the insertion $g^{(3)}(r,s,t)$ agrees with the distance-histogram result.

To derive the PC scheme, we consider the three-body cavity distribution function [277]

$$y^{(3)}(r,s,t) = \frac{g^{(3)}(r,s,t)}{e^{(2)}(r)e^{(2)}(s)e^{(2)}(t)e^{(3)}(r,s,t)}, \tag{7.5}$$

where $e^{(2)}(r) = \exp[-\beta u^{(2)}(r)]$ and $e^{(3)}(r,s,t) = \exp[-\beta u^{(3)}(r,s,t)]$. As with the two-body case, $y^{(3)}(r,s,t)$ is analogous to $g^{(3)}(r,s,t)$, but excludes the direct interactions between the particles (see Section 4.2.4).[3] As in Section 5.2.1, we assume that $y^{(3)}(r,s,t)$ is comparatively insensitive to the form of the interactions, making the approximation that

$$y^{(3)}_{\mathrm{h}}(r,s,t) \approx y^{(3)}_{j}(r,s,t). \tag{7.6}$$

Here, $y^{(3)}_{\mathrm{h}}(r,s,t)$ is the distance-histogram measurement and $y^{(3)}_{j}(r,s,t)$ is the insertion measurement, which is obtained using the three-body interaction after j iterations, $u^{(3)}_{j}(r,s,t)$. Noting that $e^{(2)}(r)$ is the same for both measurements, we substitute (7.5) into (7.6) to obtain

$$e^{(3)}(r,s,t) \approx e^{(3)}_{j}(r,s,t) \frac{g^{(3)}_{\mathrm{h}}(r,s,t)}{g^{(3)}_{j}(r,s,t)}. \tag{7.7}$$

The right-hand side of this equation is taken to be the next iteration of $e^{(3)}(r,s,t)$, yielding the corrector

$$e^{(3)}_{j+1}(r,s,t) = e^{(3)}_{j}(r,s,t) \frac{g^{(3)}_{\mathrm{h}}(r,s,t)}{g^{(3)}_{j}(r,s,t)}, \tag{7.8}$$

which is analogous to that in the two-body case (5.8). Note that we begin the scheme with $e^{(3)}_{0}(r,s,t) = 1$, corresponding to $u^{(3)}_{0}(r,s,t) = 0$.

[3] Here, the direct interactions between the three particles comprise the three pairwise interactions and the three-body interaction.

In our implementation, we use a damping factor $\eta < 1$ to reduce the magnitude of each update:

$$e_{j+1}^{(3)}(r,s,t) = e_j^{(3)}(r,s,t)\left(\frac{g_{\mathrm{h}}^{(3)}(r,s,t)}{g_j^{(3)}(r,s,t)}\right)^{\eta}; \tag{7.9}$$

this is necessary due to the increased noise of both measurements compared with the two-body case, which leads to instabilities in the iterative procedure.[4] Note also that the single-insertion method is unable to measure $g_j^{(3)}(r,s,t)$ where the side lengths all correspond with strong overlap (see Section 6.4.1), precluding a measurement of $u^{(3)}(r,s,t)$ in this region. This has no physical consequences, however, since at these separations the strongly repulsive pairwise interactions will dominate the three-body interactions anyway. Finally, the convergence of the algorithm can be followed using

$$\chi_j^2 = \sum_k \left(g_j^{(3)}(r_k,s_k,t_k) - g_{\mathrm{h}}^{(3)}(r_k,s_k,t_k)\right)^2, \tag{7.10}$$

where the sum extends over all bins, except those whose midpoints do not satisfy the triangle equalities, and those at small separations where $g_j^{(3)}(r,s,t)$ cannot be measured.

7.3 Methods

7.3.1 Implementation

Measuring Effective Pair Potentials

As discussed in Section 7.2.2, the effective pair potentials are measured by applying the PC scheme of Chapter 5 to configurational data obtained from the full simulation with both two-body and three-body interactions. Details of this implementation are found in Section 5.3.1.

[4] This is not without precedent: in measurements of $u^{(2)}(r)$ based on iteratively comparing with $g^{(2)}(r)$ from simulation, Soper implemented a scheme where the measured $u^{(2)}(r)$ is smoothed once a certain level of noise is attained, and also used a separation-dependent damping factor to control the derivative of $u^{(2)}(r)$ at the cut-off [221].

Predictor-Corrector Scheme for Three-Body Interactions

We first calculate the distance-histogram measurements $g_{\mathrm{h}}^{(2)}(r)$ and $g_{\mathrm{h}}^{(3)}(r,s,t)$ using the implementation of Section 6.3.1. The resolution of the measured $g^{(3)}(r,s,t)$ is an important trade-off: if the bins are too small, the noise will be greater and hinder comparison with the insertion measurement, whereas if the bins are too large, a poorer interpolation of $u^{(3)}(r,s,t)$ will be used to calculate the insertion energies for each iteration. For the insertion measurement we use the weighted-histogram approach of Section 6.3.1 (see page 96), although since this calculation needs to be performed for each iteration, the triangle side lengths needs to be efficiently stored in the memory for a fast implementation: we again make use of sparse matrices [204], and use a lookup table for $u_j^{(3)}(r,s,t)$.

We place a fixed number of insertion points at random positions in each of the analysed equilibrium snapshots. We take the cut-off for $u^{(3)}(r,s,t)$ to match the distance criteria used in measuring $g_{\mathrm{h}}^{(3)}(r,s,t)$, i.e. two side lengths, r and s, are less than or equal to a cut-off distance d; then the third side length $t \leq 2d$ by the triangle inequality. For triangles that do not obey these criteria, we take $u^{(3)}(r,s,t) = 0$. We find all pairs of particles around each insertion point which, together with the insertion point itself, satisfy these criteria. There are two possibilities: either t is the particle-particle distance (case A) or t is an insertion-particle distance, with s as the particle-particle distance (case B). Importantly, each neighbouring pair of particles should only be included once when counting the insertion energy, and so we exclude triangles with all three side lengths less than d from case B, since these are already accounted for in case A. We construct separate lookup matrices $\mathbf{L}_{\mathbf{A}}^{(3)}$ and $\mathbf{L}_{\mathbf{B}}^{(3)}$ for each case, with the columns corresponding to the bins of the lookup table and the rows to the insertion points. The matrix elements are the number of neighbouring pairs around each insertion point which form triangles falling into the corresponding lookup bin. We construct an additional lookup matrix $\mathbf{L}^{(2)}$ for the pairwise interactions, in the same way as in Section 5.3.1.

During the iterative procedure, we store the current $e_j^{(3)}(r,s,t)$ as a three-dimensional array of values for the midpoints (r_k, s_k, t_k) of the histogram bins. For each iteration, we calculate the lookup table from $e_j^{(3)}(r,s,t)$ by linear interpolation followed by $u^{(3)}(r,s,t) = -k_{\mathrm{B}}T\ln[e^{(3)}(r,s,t)]$—as in the two body case, we replace very small (or even negative) values of $e^{(3)}(r,s,t)$ after interpolation by a small non-zero value, in our case 10^{-20}. For a given point,

the interpolated value is based on the value of $e_j^{(3)}(r,s,t)$ at the eight nearest bin midpoints in (r,s,t)-space. Close to the limits given by the triangle inequality, $e_j^{(3)}(r,s,t)$ will not be defined at some of these midpoints, and instead we perform an extrapolation based only on the midpoints where $e_j^{(3)}(r,s,t)$ is defined.[5] To avoid needing similar extrapolations due to distance criteria, we choose these to be slightly larger ($2\times$ the bin width) than the range of $u^{(3)}(r,s,t)$, with the additional values of $u_j^{(3)}(r,s,t)$ set to zero before using the lookup table. We alter the shape of the array resulting from the interpolation to give a column vector $\mathbf{u}_j^{(3)}$. We calculate the insertion energy for each point as the sum of the contribution due to three body-interactions, given by the matrix product $(\mathbf{L_A}^{(3)} + \mathbf{L_B}^{(3)})\mathbf{u}_j^{(3)}$, and that due to pairwise interactions, given by $\mathbf{L}^{(2)}\mathbf{u}^{(2)}$ (see Section 5.3.1); here we calculate the lookup table $\mathbf{u}^{(2)}$ from the known $u^{(2)}(r)$. We then calculate $\exp[-\Psi/k_\mathrm{B}T]$ for each insertion point, stored as the column vector $\mathbf{e}_j$. The bulk ensemble average P is simply the average of the elements of this vector.

Instead of using separate comparison matrices, which would use additional memory, we use the lookup matrices to calculate the product $P^{(1)}(r,s,t)g^{(2)}(r)$ of (6.3c). We first calculate the matrix product $\widetilde{\mathbf{L_A}}^{(3)}\mathbf{e}_j$, where the tilde represents the matrix tranpose, and then multiply by the (interpolated) value of $g_\mathrm{h}^{(2)}(r)$ for the t-value of each bin, which corresponds to the particle-particle distance in case A. We repeat the process for case B, calculating $\widetilde{\mathbf{L_B}}^{(3)}\mathbf{e}_j$ and then multiplying by $g_\mathrm{h}^{(2)}(r)$ for the s-value of each bin. Adding these together and dividing by the number of triangles in each bin would give $P^{(1)}(r,s,t)g^{(2)}(r)$, but at the resolution of the lookup table rather than that of $g_\mathrm{h}^{(3)}(r,s,t)$, with correspondingly greater noise. Moreover, while it was important when calculating Ψ to consider each neighbouring pair of particles only once, here the symmetry of $g^{(3)}(r,s,t)$ with respect to interchange of the side lengths means that each such triangle can contribute multiple times. We therefore sum the values from each of the small bins which correspond to the same histogram bin, before reflecting the array in the planes $r=s$, $r=t$ and $s=t$, and adding the results together. We then divide by the number of triangles in each of these (larger) bins to give $P^{(1)}(r,s,t)g^{(2)}(r)$; dividing by P gives the insertion result $g_j^{(3)}(r,s,t)$ at the same resolution as $g_\mathrm{h}^{(3)}(r,s,t)$. We combine these results to calculate the next iteration $e_{j+1}^{(3)}(r,s,t)$ using (7.9), replacing zero values of $g_j^{(3)}(r,s,t)$, and then $e_{j+1}^{(3)}(r,s,t)$, by 10^{-20}. The corresponding three-body interaction is $u_{j+1}^{(3)}(r,s,t) = -k_\mathrm{B}T\ln[e_{j+1}^{(3)}(r,s,t)]$.

[5] We perform these extrapolations at the same time as the interpolations, using a customisation of the `RegularGridInterpolator` class from the `interpolate` sub-package of SciPy [204].

7.3.2 Protocols

Simulation: Full Interactions

In this chapter, we perform simulations in the canonical ensemble with a fixed number of particles. While this leads to small finite size effects ($\sim 1/N$) in the distribution functions [see note 5 of Section 4.3.2 (page 57)], it is necessary here since the measured $u_{\text{eff}}^{(2)}$ are state-dependent. By fixing the density, we can straightforwardly ensure that the subsequent simulations using the measured $u_{\text{eff}}^{(2)}$ are performed at the density for which they were derived.[6]

As in Chapter 6, for the pairwise interactions we used the purely repulsive WCA potential (6.13) [261] with the parameters $\sigma = 1$, $\beta = 1$ and $\epsilon = 1$, while for the three-body interactions we used (7.1) with $\zeta = -3$ (attractive three-body interactions), $\zeta = +3$ (repulsive three-body interactions) and $\zeta = 0$ (no three-body interactions). When constructing the cell scheme (see Section 3.3.4), we took care to ensure all particles with which the moving particle could interact were either in the same cell or neighbouring cells. The maximum distance between two particles on a circle is the diameter, and we therefore used $2\sigma + \delta$ for the cell length, where δ is the maximum displacement. Note that this also ensured that all the two-body interactions were captured.

We used a square simulation box of side length 50σ, corresponding to a reduced density $\rho\sigma^2 = 0.41$. We used the same $\delta = 0.45\sigma$ for each simulation, corresponding to acceptance probabilities of $\sim 52\%$ (attractive), $\sim 58\%$ (repulsive), and $\sim 60\%$ (no three-body interactions). We performed 10^7 equilibration steps, and then took 10^5 snapshots at intervals of 10^4 steps.

Analysis: Effective Pair Potentials

We used all snapshots from the full three-body simulation for the distance-histogram measurement of $g_{\text{h}}^{(2)}(r)$, and every tenth snapshot (10^4 snapshots) for the insertion analysis, with 1024 insertion points in each. We used a comparison interval of $10^{-2}\sigma$ and a lookup interval of $10^{-3}\sigma$, with a distance cut-off of 5σ. We ran the PC scheme for 250 iterations, achieving convergence in all three cases.

[6] If we were to use the grand ensemble, we would also have to optimise the chemical potential to reproduce the correct density.

Simulation: Effective Pairwise Interactions

We performed the pairwise-additive simulations with $u_{\text{eff}}^{(2)}$ at the same state as those with the full interactions, with 1024 particles in a square box of side length 50σ, and $\beta = 1$. In each case, we constructed a piecewise pair potential from the measured $u_{\text{eff}}^{(2)}$, performing the simulation with a distance cut-off of 5σ and a suitable cell scheme. We again used $\delta = 0.45\sigma$, corresponding to acceptance probabilities of $\sim 52\%$ (attractive), $\sim 59\%$ (repulsive), and $\sim 60\%$ (no three-body interactions). We again performed 10^7 equilibration steps, taking 10^5 snapshots at intervals of 10^4 steps.

We then measured the distribution functions $g^{(2)}(r)$ and $g^{(3)}(r, s, t)$ using the distance-histogram method implemented in the Chapter 6, and compared them with corresponding measurements for the simulation with the full interactions. We used all snapshots in the analyses, with a distance cut-off of 5σ for each measurement. We used a bin width of 0.01σ for $g^{(2)}(r)$ and 0.05σ for $g^{(3)}(r, s, t)$.

Analysis: Three-Body Interactions

We included alternate snapshots (total 5×10^4) in the distance-histogram measurements, using a bin width of 0.05σ for both $g_{\text{h}}^{(2)}(r)$ and $g_{\text{h}}^{(3)}(r, s, t)$ and distance criteria such that $d = 3\sigma$. As discussed in Section 7.3.1, we actually used a slightly larger d for $g^{(3)}(r, s, t)$ than that for $u^{(3)}(r, s, t)$ to avoid issues near the cut-off when interpolating. We measured $g_{\text{h}}^{(2)}(r)$ up to the maximum particle-particle distance, $2d$, considered in (6.3c). We used 10^3 insertion points in every eighth snapshot (total 1.25×10^4) to construct the lookup matrices, and the bin width for $\widetilde{\mathbf{L}_{\mathbf{A}}}^{(3)}$ and $\widetilde{\mathbf{L}_{\mathbf{B}}}^{(3)}$ was $4\times$ smaller than that used for the distance-histogram measurement; for $\widetilde{\mathbf{L}}^{(2)}$ it was $10\times$ smaller. We used a damping factor $\eta = 0.01$ and performed the iterative procedure for 2×10^3 iterations. While this was not enough for the scheme to converge, the agreement between the final $g_i^{(3)}(r, s, t)$ and $g_{\text{h}}^{(3)}(r, s, t)$ was excellent.

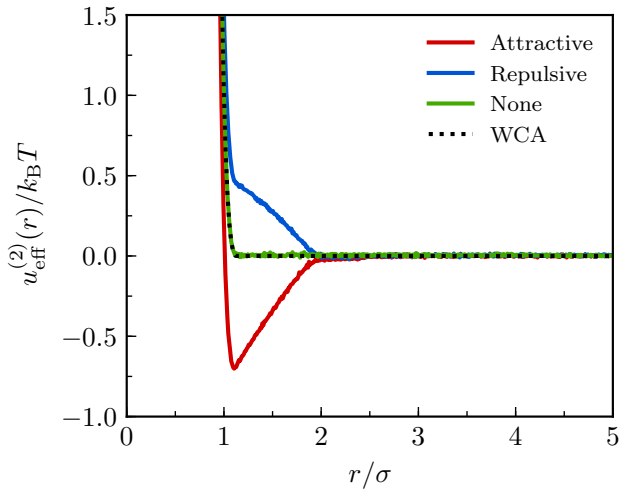


Figure 7.1: Effective pair potentials $u^{(2)}_{\mathrm{eff}}(r)$ measured from the many-body simulation data. The true two-body interaction $u^{(2)}_{\mathrm{WCA}}(r)$ is shown for comparison.

7.4 Results and Discussion

7.4.1 Effective Pair Potentials

Measured $u^{(2)}_{\mathrm{eff}}(r)$

We show the measured effective pair potentials $u^{(2)}_{\mathrm{eff}}(r)$ in Figure 7.1, where they are compared with the true two-body interaction $u^{(2)}_{\mathrm{WCA}}(r)$. As in Chapter 5, in the pairwise additive case ($\zeta = 0$) we recover the true pair potential up to small deviations resulting from the simulation noise. In the simulations with attractive or repulsive three-body interactions ($\zeta \pm 3$), however, the bare $u^{(2)}_{\mathrm{WCA}}(r)$ is augmented by attractive or repulsive features, respectively. The separations where these features are mainly found ($\sigma \lesssim r \lesssim 2\sigma$) are consistent with the underlying $u^{(3)}(r,s,t)$, where the cut-off $q(r,s,t) \leq 2\sigma$ implies that each pair of particles must be within 2σ for a non-zero interaction.

Simulation: $g^{(2)}(r)$

In the left-hand column of Figure 7.2, we compare the $g^{(2)}(r)$ measured in pairwise-additive simulations with these effective pair potentials with those from the full simulations with three-body interactions. The measured $u^{(2)}_{\mathrm{eff}}(r)$ are able to reproduce the pairwise structure either

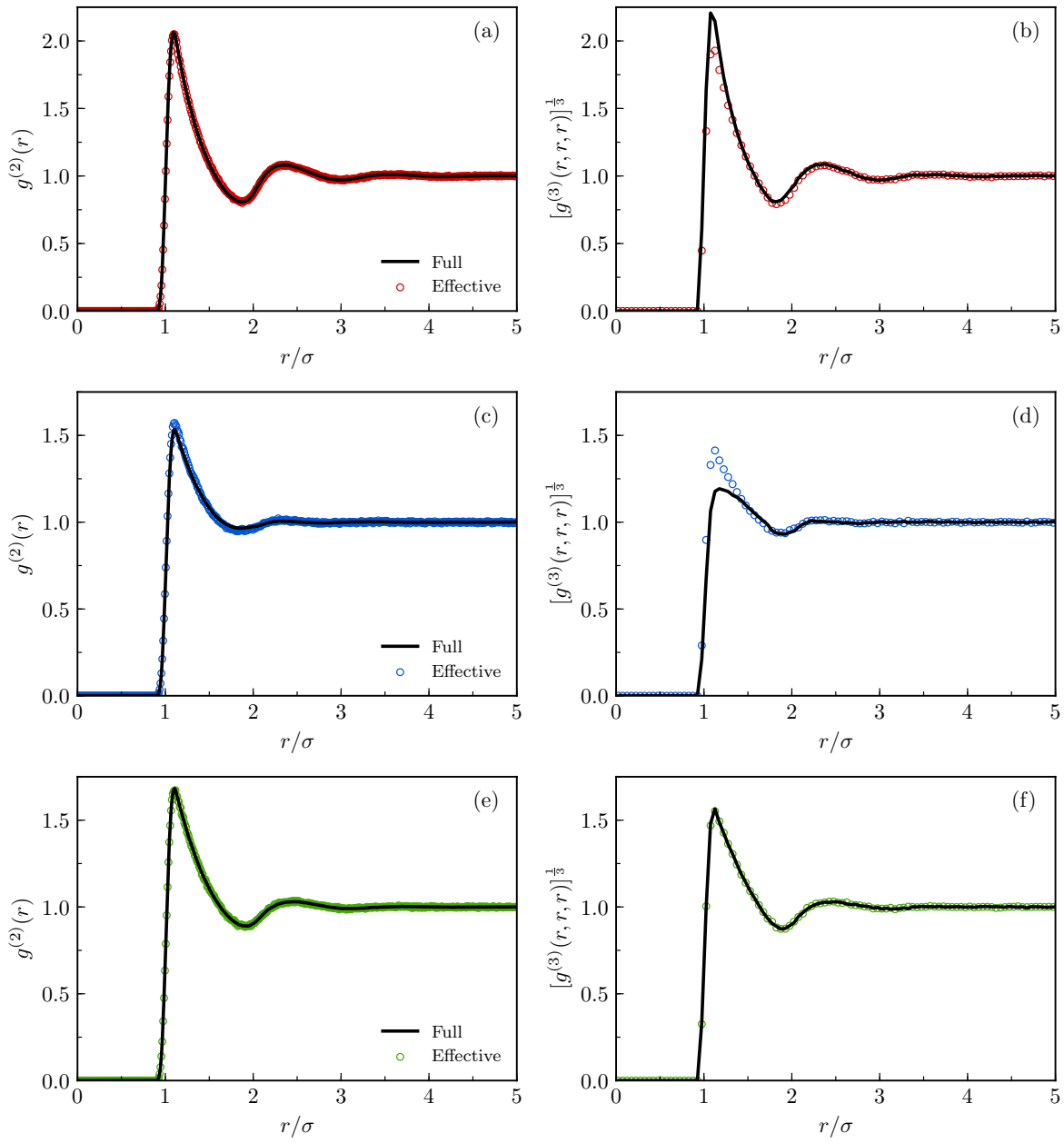


Figure 7.2: Measured $g^{(2)}(r)$ and $g^{(3)}(r,r,r)$ from the pairwise-additive simulations using the measured $u^{(2)}_{\text{eff}}(r)$ for attractive [(a), (b)], repulsive [(c), (d)] and no [(e), (f)] three-body interactions (coloured circles). The corresponding measurements from the full three-body simulations are also shown (black lines).

exactly or very closely, justifying this approach to deriving effective pair potentials. For the attractive $u^{(3)}(r,s,t)$, the agreement is essentially perfect, and small deviations can be attributed to simulation noise. The repulsive case is less conclusive, with the $g^{(2)}(r)$ from the pairwise simulation slightly accentuating the first maximum and first minimum of that from the three-body simulation. At this stage, it is not clear if this is a genuine deviation or an error introduced

by the canonical simulations, and this merits further theoretical and numerical investigation. As expected, in the case with no three-body interactions $g^{(2)}(r)$ is recovered extremely well, since the measured $u^{(2)}_{\text{eff}}(r)$ is the same as $u^{(2)}_{\text{WCA}}(r)$ within noise.

Simulation: $g^{(3)}(r,r,r)$

The right-hand column of Figure 7.2 shows the corresponding measurements of $g^{(3)}(r,r,r)$, for equilateral configurations. As anticipated in Section 7.2.2, the measured $u^{(2)}_{\text{eff}}(r)$, which are obtained by matching the distance-histogram and insertion results of $g^{(2)}(r)$, are not able to reproduce the correct three-body structure. For the attractive $u^{(3)}(r,s,t)$, the primary maximum is smaller in the pairwise-additive simulation, since the radially-symmetric $u^{(2)}_{\text{eff}}(r)$ is unable to fully capture the strongly favourable three-body interaction associated with triangles where all three particles approach closely, and the number of such triplets is underestimated [Figure 7.2(b)]. In the repulsive case, the situation is reversed and the number of triplets is overestimated by $u^{(2)}_{\text{eff}}(r)$ [Figure 7.2(d)]. At larger separations, $g^{(3)}(r,r,r)$ is well approximated by the Kirkwood Superposition Approximation (see Section 6.4.1), and so the excellent agreement of $g^{(2)}(r)$ carries over into $g^{(3)}(r,r,r)$. As expected, $g^{(3)}(r,r,r)$ is reproduced perfectly in the case with no three-body interactions, where the measured $u^{(2)}_{\text{eff}}(r)$ fully captures the interactions.

7.4.2 Three-Body Interactions

We compare the results of the $u^{(3)}(r,s,t)$ measurements to the simulation input in Figure 7.3 (attractive interaction), Figure 7.4 (repulsive interaction) and Figure 7.5 (no three-body interactions). In each case, panel (a) shows the measurement for equilateral configurations, while panel (b) shows the measurements for configurations where a pair of particles are at a fixed separation, which corresponds to the $g^{(3)}(r,s,t)$ bin containing the first maximum of $g^{(2)}(r)$, and the position of the third is allowed to vary. Although the measurements show considerably greater noise than the pair potential measurements of Chapter 5, which demonstrates the increased difficulty of the three-body measurement, the agreement with the input $u^{(3)}(r,s,t)$ is very promising. The sign, magnitude and range of the interaction are clear from each measurement, and the density plots in panel (b) show that the strongest three-body interactions occur when the three particles are all close to one another. The increasingly opaque lines in panel (a) represent the result after different numbers of iterations of the PC scheme, with 50 iterations between lines of increasing

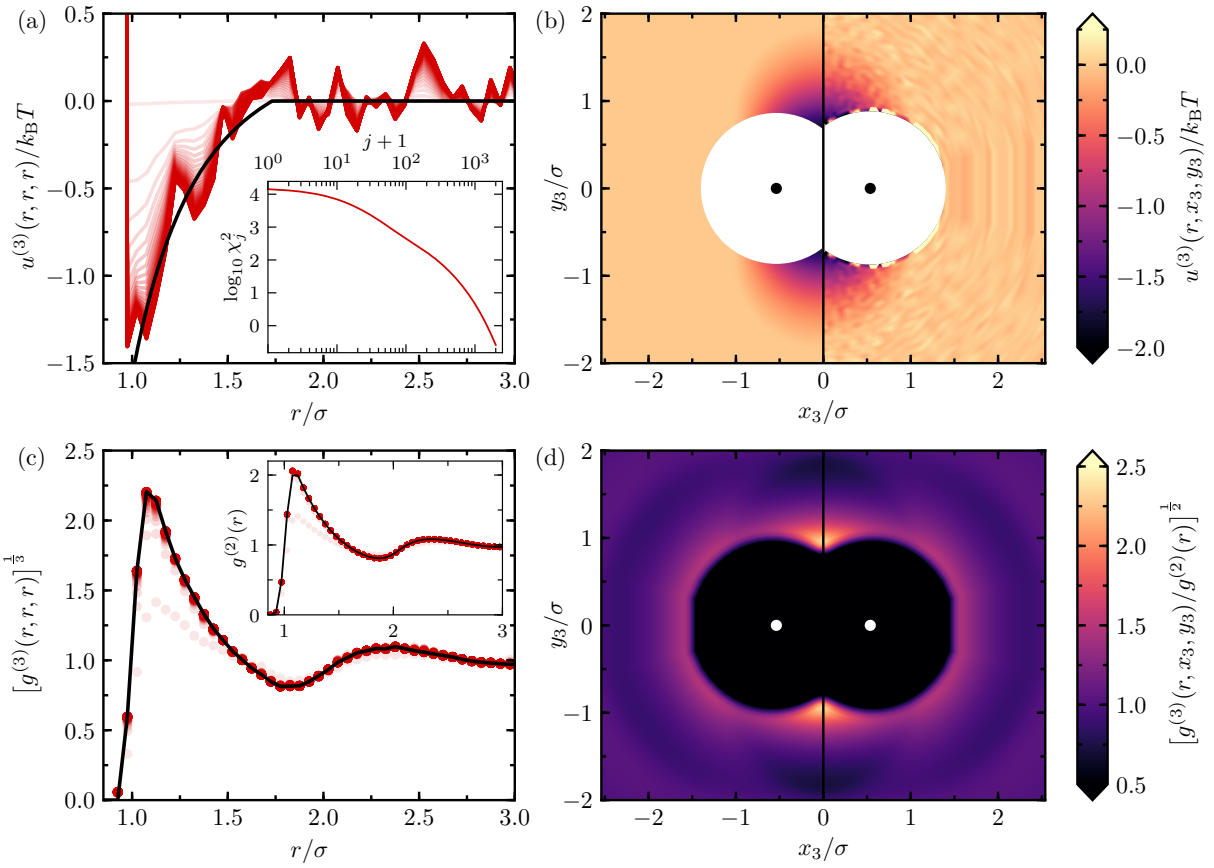


Figure 7.3: Results of the extended PC scheme for attractive three-body interactions ($\zeta = -3$). (a) Measured $u^{(3)}(r,r,r)$ for equilateral triangle configurations compared with the simulation input (black line): the increasingly opaque lines correspond to more iterations (intervals of 50 iterations). Inset: χ_j^2 as the algorithm proceeds. (b) Simulation input $u^{(3)}(r,s,t)$ (left) and the measured $u^{(3)}(r,s,t)$ (right), with two of the particles (black circles) held at a fixed separation [the maximum of $g^{(2)}(r)$]. The white region is where measurements were not possible. (c) Measured $g_j^{(3)}(r,r,r)$ (symbols) compared with the distance-histogram result (black line). Inset: same for $g^{(2)}(r)$. (d) Distance-histogram $g_{\mathrm{h}}^{(3)}(r,s,t)$ (left) and the final insertion $g_j^{(3)}(r,s,t)$ (right), with the same separation as in (b). The plots in (c) and (d) have been rescaled.

opacity: while the broad agreement with the simulation input improves as the algorithm proceeds, the noise also increases, confirming the need to use either more data or improved noise suppression procedures [221].

We are unable to measure $u^{(3)}(r,s,t)$ in cases where two of the particles are so close that $u^{(2)}(r) \sim r^{-12}$ dominates over $u^{(3)}(r,s,t) \sim q(r,s,t)^{-3}$, since here $g^{(3)}(r,s,t)$ is essentially 0 irrespective of $u^{(3)}(r,s,t)$. Instead, we observe an initial repulsion for $r \lesssim \sigma$ in all cases, and no value is measured at closer separations: to assist with comparison, we have drawn a white circle around each of the fixed particles to indicate this region. On close inspection, we observe that there is a thin ring around each fixed particle where the measurements are very

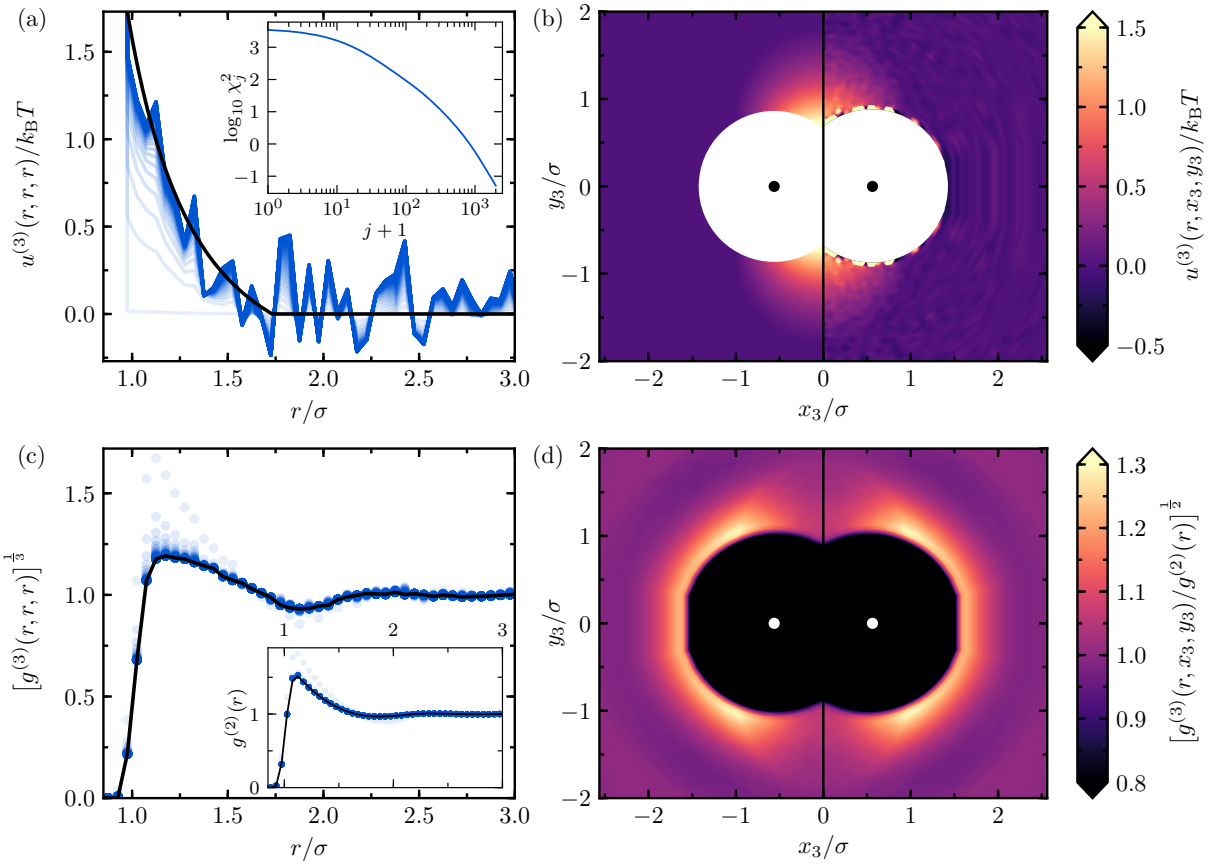


Figure 7.4: Results of the extended PC scheme for repulsive three-body interactions ($\zeta = +3$). The panel descriptions are the same as in Figure 7.3.

noisy: this results from the relatively large bin widths used for both measurements, which lead to inaccuracies in the measured $g^{(3)}(r,s,t)$ in regions where it varies rapidly (as discussed in Chapter 6). Since the PC scheme strives to match the two measurements, this inaccuracy is transferred into $u^{(3)}(r,s,t)$ and may partially explain the instability of the algorithm, which results in the need for a damping factor η.

The inset of panel (a) shows the improved agreement of the insertion $g^{(3)}(r,s,t)$ with the distance-histogram measurement as the algorithm progresses; we also show this explicitly in panel (c), where the increased opacity of the symbols again indicates the progression of the PC scheme. In the attractive case, the first maximum of the insertion $g^{(3)}(r,r,r)$ is enhanced as the algorithm progresses, since the strong $u^{(3)}(r,r,r)$ for such configurations is increasingly detected [panel (a)]; in the repulsive case, the situation is reversed and the first maximum is reduced as the algorithm proceeds. The same trends are observed in the measured $g^{(2)}(r)$ [inset of panel (c)]. In the case with no three-body interaction, the agreement of the measured distribution functions is immediately excellent, since the initial condition $u^{(3)}(r,s,t) = 0$ is correct, but

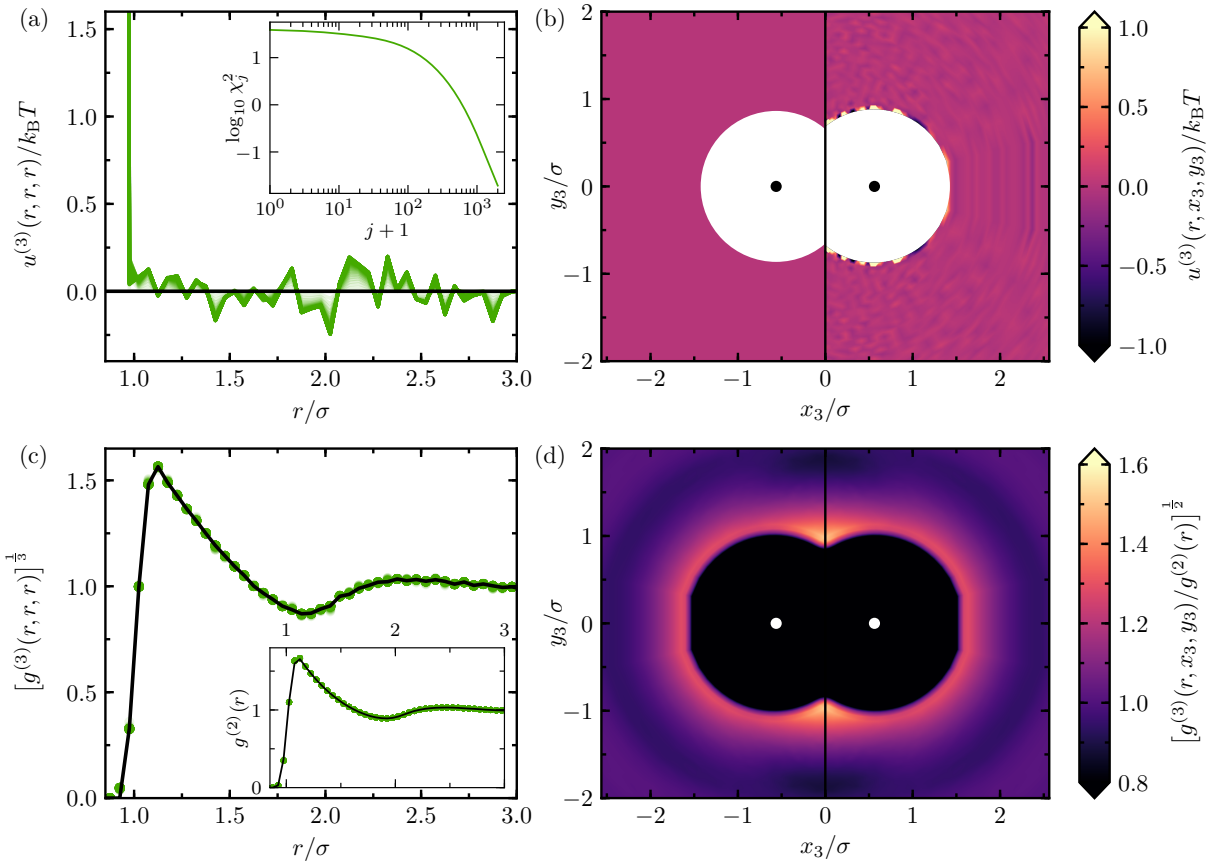


Figure 7.5: Results of the extended PC scheme for the pairwise-additive case ($\zeta = 0$). The panel descriptions are the same as in Figure 7.3.

there is a slight discrepancy due to noise which the algorithm seeks to rectify. The result is that the noise in $u^{(3)}(r,s,t)$ grows as the algorithm proceeds, although not to the same extent as in the attractive and repulsive cases. In panel (d), we compare the distance-histogram and final insertion measurements of $g^{(3)}(r,s,t)$ for the same fixed separation as in panel (b), and find excellent agreement, to a level where any discrepancies cannot be detected by eye. Note that these measurements also show the effect of $u^{(3)}(r,s,t)$ on the three body structure: in the attractive case, the number of triplets where the three particles are close together is increased, while in the repulsive case the number is reduced.

Taken together, these measurements illustrate the promise of this approach for measuring three-body interactions, but further work is required to reduce the noise and enable more quantitative measurements. This is likely to involve a combination of increased computational resources, allowing more data to be analysed, and algorithmic improvements such as those implemented by Soper in the two-body case [221]. These may include more sophisticated damping factors, depending on the side lengths, and smoothing the current $u_j^{(3)}(r,s,t)$ if the total noise

rises above a threshold.

7.5 Conclusion and Outlook

We have demonstrated two approaches for incorporating three-body interactions into the framework of this thesis. We have shown that naively applying the PC scheme of Chapter 5 in systems with three-body interactions allows a state-dependent *effective pair potential* $u^{(2)}_{\text{eff}}(r)$ to be measured, which reproduces the distance-histogram $g^{(2)}(r)$ when used in the insertion measurement. We have performed pairwise-additive simulations with the measured $u^{(2)}_{\text{eff}}(r)$ and have found the resulting $g^{(2)}(r)$ to agree very well with those of the original three-body simulations, showing that this result is either exact or a very good approximation. While the measured $u^{(2)}_{\text{eff}}(r)$ may be useful in estimating the sign and range of the three-body interactions, the pairwise-additive simulations do not result in the correct $g^{(3)}(r,s,t)$, and so $u^{(2)}_{\text{eff}}(r)$ does not fully reproduce the properties of the system, in agreement with expectations.

We have also extended the PC scheme to incorporate the three-body interaction $u^{(3)}(r,s,t)$. In principle, this involves matching the distance-histogram and insertion measurements of both $g^{(2)}(r)$ and $g^{(3)}(r,s,t)$, yielding both $u^{(2)}(r)$ and $u^{(3)}(r,s,t)$, but in this chapter we have focussed on the simpler task of measuring $u^{(3)}(r,s,t)$ when $u^{(2)}(r)$ is known. We have found broad agreement between the measurements and the simulation input, but the measured $u^{(3)}(r,s,t)$ are rather noisy compared with the pair potential measurements in Chapter 5. Future development of the algorithm will focus on suppressing this noise by smoothing, damping, or incorporating more data, and extending the measurement to obtain $u^{(2)}(r)$ and $u^{(3)}(r,s,t)$ simultaneously.

Overall, the results of this chapter illustrate the promise of the insertion approach to study systems with higher-order interactions. The effective pair potential measurement will be useful in deriving coarse-grained interactions for use in simulations of biological systems and more complex fluids [99,220,229]. The nature of the measured $u^{(2)}_{\text{eff}}(r)$ remains unclear, suggesting a theoretical study into whether it reproduces the exact $g^{(2)}(r)$ of the system with many-body interactions. Improving the implementation of the full $u^{(3)}(r,s,t)$ measurement will open the possibility of applying it in colloidal systems to support previous studies [87, 225, 265, 276]; as in the two-body case, this may become a routine measurement requiring only a simple microscope. In principle, the PC scheme can be further extended to measure higher-order interactions, however

the implementation difficulties encountered with the three-body scheme suggest a more realistic approach would be to use this scheme to derive effective two-body and three-body interactions in such systems.

Conclusion and Outlook

In this thesis, we have seen that the distribution functions $g^{(n)}$ of a classical fluid can be measured from configurational data using test-particle insertion methods, as an alternative to conventional distance-histogram approaches. The arbitrary resolution of the insertion methods is particularly useful in cases where the distribution functions vary rapidly and the distance-histogram approximation breaks down, which occurs when the particles are close together. In particular, we have used insertion methods to exactly measure $g^{(2)}(r)$ at contact in hard-disk systems (Chapter 4) and to address the poor resolution of the distance-histogram method in measuring $g^{(3)}$ (Chapter 6).

Importantly, the insertion methods require the precise interparticle interactions to be known, and so we might expect such methods to be of limited use in colloidal systems where the interactions are typically known only approximately. On the contrary, however, determining the interactions which match the results of the insertion method to those of the distance-histogram method provides an elegant route to measure the interactions in such systems. In Chapter 5, we showed that matching $g^{(2)}(r)$ allows for a model-free measurement of the pair potential $u^{(2)}(r)$ in pairwise-additive systems, successfully demonstrating the technique in the superparamagnetic colloidal model system. In systems with a three-body interaction $u^{(3)}$ (Chapter 7), matching $g^{(2)}(r)$ yields an *effective pair potential* which, while giving an indication of the sign and range of $u^{(3)}$, is unable to reproduce $g^{(3)}$ in a pairwise-additive simulation. To measure $u^{(3)}$ requires matching the insertion and distance-histogram results for $g^{(3)}$, and we have demonstrated the feasibility of this approach in the case that $u^{(2)}(r)$ is already known.

Obtaining the interactions of a fluid from equilibrium configurations solves a key inverse problem in liquid state theory. Its application to provide a routine measurement of the interactions between colloidal particles is expected to inform condensed matter studies using colloidal model systems, and to be useful in industry, where interparticle interactions play an important role in product formulation. Moreover, in the chapters of this thesis we suggest a multitude of avenues for further research. The insertion methods can be implemented to measure distribution

functions, and hence interactions and thermodynamic quantities, in more complex systems, such as inhomogeneous fluids and those with anisotropic particles. Effective pair potentials which correctly reproduce lower-order distribution functions may be useful in investigating more complex systems, such as those composed of active particles or encountered in biological materials. Finally, an improved implementation of the three-body interaction measurements will enable fundamental investigations into the nature of such interactions in colloidal systems.

Appendix

A Links between $g^{(2)}(r)$ and Thermodynamic Quantities

A.1 The Energy Equation

Consider a pairwise-additive fluid in the absence of an external field, with potential energy given by (2.16)

$$U(\mathbf{r}^N) = \sum_i \sum_{j>i} u^{(2)}(\mathbf{r}_i, \mathbf{r}_j). \tag{2.16}$$

The average potential energy $\langle U \rangle$ of this system is found, using (2.12), to be

$$\langle U \rangle = \mathrm{Tr}_\mathrm{c}\left[P_\mathrm{c} U_N\right] \tag{A.1}$$

$$= \sum_{N=0}^{\infty} \frac{1}{N!} \int \mathrm{d}\mathbf{r}^N P_\mathrm{c}(\mathbf{r}^N; N) \sum_{i=1}^{N} \sum_{j=i+1}^{N} u^{(2)}(\mathbf{r}_i, \mathbf{r}_j) \tag{A.2}$$

$$= \sum_{N=0}^{\infty} \frac{1}{N!} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \int \mathrm{d}\mathbf{r}^N P_\mathrm{c}(\mathbf{r}^N; N) u^{(2)}(\mathbf{r}_i, \mathbf{r}_j), \tag{A.3}$$

where on the final line we have switched the order of the pairwise sum and the integral. Since both P_c and $u^{(2)}$ are symmetric with respect to interchange of particles, the integral is the same for each term in the pairwise sum. We can therefore replace the pairwise sum by its first term

multiplied by the number of pairs $N(N-1)/2$, yielding

$$\langle U \rangle = \frac{1}{2} \sum_{N=0}^{\infty} \frac{N(N-1)}{N!} \int \mathrm{d}\mathbf{r}^N P_\mathrm{c}(\mathbf{r}^N; N) u^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \tag{A.4}$$

$$= \frac{1}{2} \int \mathrm{d}\mathbf{r}_1 \int \mathrm{d}\mathbf{r}_2 u^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \sum_{N=2}^{\infty} \frac{1}{(N-2)!} \int \mathrm{d}\mathbf{r}^{(N-2)} P_\mathrm{c}(\mathbf{r}^N; N) \tag{A.5}$$

$$= \frac{1}{2} \int \mathrm{d}\mathbf{r}_1 \int \mathrm{d}\mathbf{r}_2 u^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2). \tag{A.6}$$

On the second line we have noted that the first two terms of the sum are zero, that the factor $u^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ can be taken out of the integrals over $\mathbf{r}^{(N-2)}$, and that the order of the sum and the integrals over $\mathbf{r}_1$ and $\mathbf{r}_2$ may be changed. On the final line, we have used (2.36) to replace the sum by $\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$. Since the system is uniform, $u^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ and $\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ depend only on the relative position $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$, and the integration variables may be changed to first integrate over $\mathbf{r}_{12}$, and then over all space:

$$\langle U \rangle = \frac{1}{2} \int \mathrm{d}\mathbf{r} \int \mathrm{d}\mathbf{r}_{12} u^{(2)}(\mathbf{r}_{12}) \rho^{(2)}(\mathbf{r}_{12}), \tag{A.7}$$

$$\frac{\langle U \rangle}{N} = \frac{\rho}{2} \int \mathrm{d}\mathbf{r}_{12} u^{(2)}(\mathbf{r}_{12}) g^{(2)}(\mathbf{r}_{12}), \tag{2.63}$$

where we have noted that $\rho^{(2)}(\mathbf{r}_{12}) = \rho^2 g^{(2)}(\mathbf{r}_{12})$ for a uniform system, according to (2.43). For the two-dimensional systems of isotropic particles used in this thesis, we may also integrate out the angular dependence to obtain

$$\frac{\langle U \rangle}{N} = \rho\pi \int_0^{\infty} r u^{(2)}(r) g^{(2)}(r) \mathrm{d}r, \tag{2.64}$$

with $r = |\mathbf{r}_{12}|$. A corresponding derivation in the canonical ensemble is found in Reference [1]. Similar principles can be used to derive analogous equations for systems with many-body interactions [135].

A.2 The Pressure Equation

While the energy equation is not used directly in this thesis, its derivation shares several essential features with the more involved derivation of the *pressure equation*. Equations (2.13) and (2.14)

from Section 2.1:

$$\Omega = -k_{\mathrm{B}}T \ln \Xi, \tag{2.13}$$

$$\mathrm{d}\Omega = -p\mathrm{d}V - S\mathrm{d}T - \langle N \rangle \, \mathrm{d}\mu, \tag{2.14}$$

can be combined to yield

$$\beta p = \frac{1}{\Xi} \left(\frac{\partial \Xi}{\partial V} \right)_{T,\mu}. \tag{A.8}$$

We now suppose that the volume $V = \alpha R^s$, where R is some length, s is the dimensionality and α is a factor depending on the shape of the container. Using the chain rule and substituting Ξ from (2.6), we obtain

$$\beta p = \frac{1}{\Xi \alpha s R^{s-1}} \left(\frac{\partial \Xi}{\partial R} \right)_{T,\mu} \tag{A.9}$$

$$= \frac{1}{\Xi \alpha s R^{s-1}} \sum_{N=0}^{\infty} \frac{z^N}{N!} \frac{\partial}{\partial R} \int \mathrm{d}\mathbf{r}^N \exp \left[-\beta \sum_{i=1}^{N} \sum_{j=i+1}^{N} u^{(2)}(r_{ij}) \right], \tag{A.10}$$

where we have again considered a uniform pairwise-additive fluid and have omitted the variables held constant for notational convenience. To perform the derivative, we introduce the scaled coordinates [130] $\mathbf{q} = \mathbf{r}/R$ to give

$$\beta p = \frac{1}{\Xi \alpha s R^{s-1}} \sum_{N=0}^{\infty} \frac{z^N}{N!} \frac{\partial}{\partial R} \int \mathrm{d}\mathbf{q}^N R^{sN} \exp \left[-\beta \sum_{i=1}^{N} \sum_{j=i+1}^{N} u^{(2)}(q_{ij} R) \right] \tag{A.11}$$

$$= \frac{1}{\Xi \alpha s R^{s-1}} \sum_{N=0}^{\infty} \frac{z^N}{N!} \int \mathrm{d}\mathbf{q}^N \Bigg\{ sN R^{sN-1} \exp \left[-\beta U_N \right] - \beta R^{sN} \exp \left[-\beta U_N \right] \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\mathrm{d}u^{(2)}(q_{ij}R)}{\mathrm{d}(q_{ij}R)} q_{ij} \Bigg\} \tag{A.12}$$

$$= \frac{1}{sV} \left\{ \sum_{N=0}^{\infty} \frac{sN}{N!} \int \mathrm{d}\mathbf{r}^N P_{\mathrm{c}}(\mathbf{r}^N; N) - \sum_{N=0}^{\infty} \frac{\beta}{N!} \int \mathrm{d}\mathbf{r}^N P_{\mathrm{c}}(\mathbf{r}^N; N) \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\mathrm{d}u^{(2)}(r_{ij})}{\mathrm{d}r_{ij}} r_{ij} \right\}. \tag{A.13}$$

The first term in brackets on the final line is simply $s \langle N \rangle$, while the second term can be

manipulated as in the derivation of (2.63) to yield

$$\beta p = \frac{\langle N \rangle}{V} - \frac{\beta}{s} \int \mathrm{d}\mathbf{r} r \frac{\mathrm{d}u^{(2)}(r)}{\mathrm{d}r} \rho^{(2)}(r) \tag{A.14}$$

$$= \rho - \frac{\beta \rho^2}{s} \int \mathrm{d}\mathbf{r} r \frac{\mathrm{d}u^{(2)}(r)}{\mathrm{d}r} g^{(2)}(r). \tag{2.65}$$

In the two-dimensional case ($s = 2$), and after integrating out the angular degree of freedom, we obtain

$$\frac{\beta p}{\rho} = 1 - \frac{\beta \rho \pi}{2} \int_0^\infty r^2 \frac{\mathrm{d}u^{(2)}(r)}{\mathrm{d}r} g^{(2)}(r) \mathrm{d}r. \tag{2.66}$$

A similar derivation, but in the canonical ensemble, is found in Reference [130]. Again, analogous equations can be derived for systems with many-body interactions [135].

B Statistical Geometry of Hard-Disk Systems

This appendix essentially reproduces the corresponding section of Reference [115] (with the permission of AIP Publishing), on which Chapter 4 is based.

The expression (4.22) can be used to derive an elegant link between the thermodynamics of hard disks and certain geometric quantities: in the single-component case, these are the average volume available for test-particle insertion, V_s, and the surface area of this volume, A_s [200]. Similar results have been derived for multicomponent systems [278] and can be derived, in similar spirit to Speedy's original single-component derivation [200], by substituting (4.22) into (4.20), which yields

$$\beta P = \sum_\alpha \rho_\alpha + \frac{\pi}{2} \sum_{\alpha,\beta} \rho_\alpha \rho_\beta \sigma_{\alpha\beta}^2 \frac{P_\alpha^{\star(\beta)}(\sigma_{\alpha\beta})}{P_\alpha}, \tag{B.15}$$

where we have noted that that $g_{\alpha\beta}(\sigma_{\alpha\beta}^+) = y_{\alpha\beta}(\sigma_{\alpha\beta})$ for hard disks.

The insertion probabilities are related to geometric quantities (see Figure B.1) by first noting that $P_\alpha = \langle V_{\mathrm{s},\alpha} \rangle / V$, where $\langle V_{\mathrm{s},\alpha} \rangle$ is the average volume available for insertion of a test particle of species α. The local insertion probability $P_\alpha^{\star(\beta)}(\sigma_{\alpha\beta}) = \langle A_{\mathrm{s},\alpha}^{(\beta)} \rangle / 2\pi\sigma_{\alpha\beta} N_\beta$, where $\langle A_{\mathrm{s},\alpha}^{(\beta)} \rangle$ is the average of the component of the surface area of $V_{\mathrm{s},\alpha}$ which corresponds to contact with

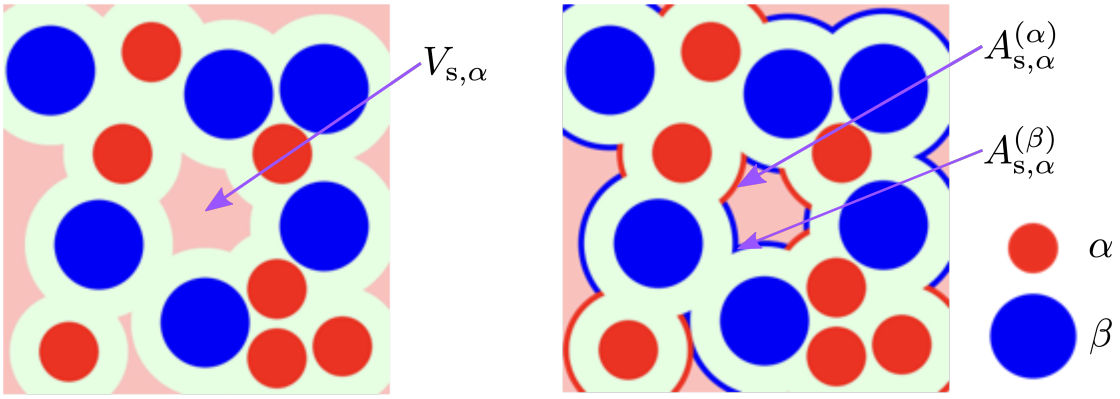


Figure B.1: The statistical geometric quantities for a binary 2D hard disk system with species α (red) and β (blue). Left: around each disk is an exclusion zone (light green) which is unavailable to the centre of an α particle. Unlike the disks themselves, the exclusion zones may overlap. The remaining volume (orange) is the volume available for insertion of an α particle, $V_{\mathrm{s},\alpha}$. Right: the surface area of $V_{\mathrm{s},\alpha}$ is divided into components due to α particles (red lines), $A^{(\alpha)}_{\mathrm{s},\alpha}$, and due to β particles (blue lines) $A^{(\beta)}_{\mathrm{s},\alpha}$.

β particles, and N_β is the total number of β particles. This ratio is thus the average fraction of the circle of radius $\sigma_{\alpha\beta}$ around each β particle which is available for α particle insertion. Substituting these into (B.15) gives

$$\beta P = \sum_\alpha \rho_\alpha + \frac{1}{4}\sum_{\alpha,\beta} \rho_\alpha \sigma_{\alpha\beta} \frac{\langle A^{(\beta)}_{\mathrm{s},\alpha}\rangle}{\langle V_{\mathrm{s},\alpha}\rangle}, \tag{B.16}$$

which is the 2D version of the result derived by Corti and Bowles [278]. The result (B.16) extends straightforwardly to fluids in higher dimensions [200, 278] and the 3D version of the single-component result has even been used in experiments on colloidal hard spheres [13].

Bibliography

[1] J.-P. Hansen and I. R. McDonald. *Theory of Simple Liquids, Fourth Edition: With Applications to Soft Matter* (Academic Press, New York, 2013).

[2] The Editors of Encyclopaedia Britannica (Last Updated: 8 April 2020). *colloid | Definition & Facts | Britannica*. [online] Available at: ⟨https://www.britannica.com/science/colloid⟩ [Accessed 29 April 2020].

[3] M. Doi. *Soft Matter Physics* (Oxford University Press, Oxford, 2013).

[4] P. N. Pusey, W. van Megen, P. Bartlett, B. J. Ackerson, J. G. Rarity, and S. M. Underwood. Structure of crystals of hard colloidal spheres. Phys. Rev. Lett. **63**, 2753 (1989).

[5] U. Gasser, E. R. Weeks, A. Schofield, P. N. Pusey, and D. A. Weitz. Real-Space Imaging of Nucleation and Growth in Colloidal Crystallization. Science **292**, 258 (2001).

[6] J.-M. Meijer, A. Pal, S. Ouhajji, H. N. W. Lekkerkerker, A. P. Philipse, and A. V. Petukhov. Observation of solid–solid transitions in 3D crystals of colloidal superballs. Nat. Commun. **8** (2017).

[7] P. N. Pusey and W. van Megen. Observation of a glass transition in suspensions of spherical colloidal particles. Phys. Rev. Lett. **59**, 2083 (1987).

[8] A. van Blaaderen and P. Wiltzius. Real-Space Structure of Colloidal Hard-Sphere Glasses. Science **270**, 1177 (1995).

[9] G. L. Hunter and E. R. Weeks. The physics of the colloidal glass transition. Rep. Prog. Phys. **75**, 066501 (2012).

[10] J. E. Hallett, F. Turci, and C. P. Royall. Local structure in deeply supercooled liquids exhibits growing lengthscales and dynamical correlations. Nat. Commun. **9** (2018).

[11] W. V. Megen and I. Snook. Equilibrium properties of suspensions. Adv. Colloid Interface Sci. **21**, 119 (1984).

[12] A. Moussaïd, W. C. K. Poon, P. N. Pusey, and M. F. Soliva. Structure of Marginal and Fully Developed Colloidal Liquids. Phys. Rev. Lett. **82**, 225 (1999).

[13] R. P. A. Dullens, D. G. A. L. Aarts, and W. K. Kegel. Direct measurement of the free energy by optical microscopy. Proc. Natl. Acad. Sci. U.S.A. **103**, 529 (2006).

[14] C. P. Royall, D. G. A. L. Aarts, and H. Tanaka. Bridging length scales in colloidal liquids and interfaces from near-critical divergence to single particles. Nat. Phys. **3**, 636 (2007).

[15] A. L. Thorneywork, R. Roth, D. G. A. L. Aarts, and R. P. A. Dullens. Communication: Radial distribution functions in a two-dimensional binary colloidal hard sphere system. J. Chem. Phys. **140**, 161106 (2014).

[16] E. Grelet. Hexagonal Order in Crystalline and Columnar Phases of Hard Rods. Phys. Rev. Lett. **100**, 168301 (2008).

[17] O. J. Dammone, I. Zacharoudiou, R. P. A. Dullens, J. M. Yeomans, M. P. Lettinga, and D. G. A. L. Aarts. Confinement Induced Splay-to-Bend Transition of Colloidal Rods. Phys. Rev. Lett. **109**, 108303 (2012).

[18] C. Fernández-Rico, T. Yanagishima, A. Curran, D. G. A. L. Aarts, and R. P. A. Dullens. Synthesis of Colloidal SU-8 Polymer Rods Using Sonication. Adv. Mater. **31**, 1807514 (2019).

[19] E. Zaccarelli. Colloidal gels: equilibrium and non-equilibrium routes. J. Phys. Condens. Matter **19**, 323101 (2007).

[20] P. J. Lu, E. Zaccarelli, F. Ciulla, A. B. Schofield, F. Sciortino, and D. A. Weitz. Gelation of particles with short-range attraction. Nature (London) **453**, 499 (2008).

[21] D. G. A. L. Aarts, M. Schmidt, and H. N. W. Lekkerkerker. Direct visual observation of thermal capillary waves. Science **304**, 847 (2004).

[22] D. G. A. L. Aarts. The interface in demixed colloid-polymer systems: wetting, waves and droplets. Soft Matter **3**, 19 (2007).

[23] A. A. Verhoeff, F. A. Lavergne, D. Bartolo, D. G. A. L. Aarts, and R. P. A. Dullens. Optical trapping of interfaces at ultra-low interfacial tension. Soft Matter **11**, 3100 (2015).

[24] J. C. Crocker and D. G. Grier. Methods of Digital Video Microscopy for Colloidal Studies. J. Colloid Interface Sci. **179**, 298 (1996).

[25] A. D. Dinsmore, E. R. Weeks, V. Prasad, A. C. Levitt, and D. A. Weitz. Three-dimensional confocal microscopy of colloids. Appl. Opt. **40**, 4152 (2001).

[26] V. W. A. de Villeneuve, R. P. A. Dullens, D. G. A. L. Aarts, E. Groeneveld, J. H. Scherff, W. K. Kegel, and H. N. W. Lekkerkerker. Colloidal Hard-Sphere Crystal Growth Frustrated by Large Spherical Impurities. Science **309**, 1231 (2005).

[27] S. Gokhale, K. H. Nagamanasa, R. Ganapathy, and A. K. Sood. Grain growth and grain boundary dynamics in colloidal polycrystals. Soft Matter **9**, 6634 (2013).

[28] F. A. Lavergne, D. G. A. L. Aarts, and R. P. A. Dullens. Anomalous Grain Growth in a Polycrystalline Monolayer of Colloidal Hard Spheres. Phys. Rev. X **7**, 041064 (2017).

[29] W. T. M. Irvine, A. D. Hollingsworth, D. G. Grier, and P. M. Chaikin. Dislocation reactions, grain boundaries, and irreversibility in two-dimensional lattices using topological tweezers. Proc. Natl. Acad. Sci. U.S.A. **110**, 15544 (2013).

[30] F. A. Lavergne, A. Curran, D. G. A. L. Aarts, and R. P. A. Dullens. Dislocation-controlled formation and kinetics of grain boundary loops in two-dimensional crystals. Proc. Natl. Acad. Sci. U.S.A. **115**, 6922 (2018).

[31] F. A. Lavergne, A. Curran, D. G. A. L. Aarts, and R. P. A. Dullens. Shrinkage mechanisms of grain boundary loops in two-dimensional colloidal crystals. Eur. Phys. J. B **92** (2019).

[32] E. R. Weeks, J. C. Crocker, A. C. Levitt, A. Schofield, and D. A. Weitz. Three-Dimensional Direct Imaging of Structural Relaxation Near the Colloidal Glass Transition. Science **287**, 627 (2000).

[33] J. M. Lynch, G. C. Cianci, and E. R. Weeks. Dynamics and structure of an aging binary colloidal glass. Phys. Rev. E **78**, 031410 (2008).

[34] P. M. Chaikin and T. C. Lubensky. *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 1995).

[35] A. V. Blaaderen, A. Imhof, W. Hage, and A. Vrij. Three-dimensional imaging of submicrometer colloidal particles in concentrated suspensions using confocal scanning laser microscopy. Langmuir **8**, 1514 (1992).

[36] M. S. Elliot and W. C. K. Poon. Conventional optical microscopy of colloidal suspensions. Adv. Colloid Interface Sci. **92**, 133 (2001).

[37] A. A. Louis. Beware of density dependent pair potentials. J. Phys. Condens. Matter **14**, 9187 (2002).

[38] D. F. Evans and H. Wennerström. *The Colloidal Domain: Where Physics, Chemistry, Biology, and Technology Meet, Second Edition* (Wiley-VCH, New York, 1999).

[39] J. N. Israelachvili. *Intermolecular and Surface Forces, Second Edition* (Academic Press, New York, 1991).

[40] S. Sacanna and D. J. Pine. Shape-anisotropic colloids: Building blocks for complex assemblies. Curr. Opin. Colloid Interface Sci. **16**, 96 (2011).

[41] S. Sacanna, D. J. Pine, and G.-R. Yi. Engineering shape: the novel geometries of colloidal self-assembly. Soft Matter **9**, 8096 (2013).

[42] microParticles GmbH (2020). *Product Palette | microParticles GmbH*. [online] Available at: ⟨https://microparticles.de/en/product-palette.html⟩ [Accessed 23 April 2020].

[43] A. Yethiraj and A. van Blaaderen. A colloidal model system with an interaction tunable from hard sphere to soft and dipolar. Nature (London) **421**, 513 (2003).

[44] C. P. Royall, A. A. Louis, and H. Tanaka. Measuring colloidal interactions with confocal microscopy. J. Chem. Phys. **127**, 044507 (2007).

[45] H. N. Lekkerkerker and R. Tuinier. *Colloids and the Depletion Interaction* (Springer, Dordrecht, The Netherlands, 2011).

[46] D. Napper. Steric stabilization. J. Colloid Interface Sci. **58**, 390 (1977).

[47] L. Antl, J. W. Goodwin, R. D. Hill, R. H. Ottewill, S. M. Owens, S. Papworth, and J. A. Waters. The preparation of poly(methyl methacrylate) latices in non-aqueous media. Colloids Surf. **17**, 67 (1986).

[48] D. van der Beek and H. N. W. Lekkerkerker. Liquid Crystal Phases of Charged Colloidal Platelets. Langmuir **20**, 8582 (2004).

[49] A. Kuijk, D. V. Byelov, A. V. Petukhov, A. van Blaaderen, and A. Imhof. Phase behavior of colloidal silica rods. Faraday Discuss. **159**, 181 (2012).

[50] W. C. K. Poon. The physics of a model colloid-polymer mixture. J. Phys. Condens. Matter **14**, R859 (2002).

[51] P. Bartlett, R. H. Ottewill, and P. N. Pusey. Superlattice formation in binary mixtures of hard-sphere colloids. Phys. Rev. Lett. **68**, 3801 (1992).

[52] K. P. Velikov, C. G. Christova, R. P. A. Dullens, and A. van Blaaderen. Layer-by-Layer Growth of Binary Colloidal Crystals. Science **296**, 106 (2002).

[53] Q. Chen, S. C. Bae, and S. Granick. Directed self-assembly of a colloidal kagome lattice. Nature (London) **469**, 381 (2011).

[54] J. Gong, R. S. Newman, M. Engel, M. Zhao, F. Bian, S. C. Glotzer, and Z. Tang. Shape-dependent ordering of gold nanocrystals into large-scale superlattices. Nat. Commun. **8** (2017).

[55] H. Lin, S. Lee, L. Sun, M. Spellings, M. Engel, S. C. Glotzer, and C. A. Mirkin. Clathrate colloidal crystals. Science **355**, 931 (2017).

[56] A. Stradner, H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf, and P. Schurtenberger. Equilibrium cluster formation in concentrated protein solutions and colloids. Nature (London) **432**, 492 (2004).

[57] D. G. Grier. A revolution in optical manipulation. Nature (London) **424**, 810 (2003).

[58] P. Jones, O. Marago, and G. Volpe. *Optical Tweezers* (Cambridge University Press, Cambridge, 2015).

[59] A. Kuijk, T. Troppenz, L. Filion, A. Imhof, R. van Roij, M. Dijkstra, and A. van Blaaderen. Effect of external electric fields on the phase behavior of colloidal silica rods. Soft Matter **10**, 6249 (2014).

[60] K. Zahn, J. M. Méndez-Alcaraz, and G. Maret. Hydrodynamic Interactions May Enhance the Self-Diffusion of Colloidal Particles. Phys. Rev. Lett. **79**, 175 (1997).

[61] K. Zahn, R. Lenke, and G. Maret. Two-Stage Melting of Paramagnetic Colloidal Crystals in Two Dimensions. Phys. Rev. Lett. **82**, 2721 (1999).

[62] K. Zahn, A. Wille, G. Maret, S. Sengupta, and P. Nielaba. Elastic Properties of 2D Colloidal Crystals from Video Microscopy. Phys. Rev. Lett. **90**, 155506 (2003).

[63] D. Frenkel and B. Smit. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science)* (Academic Press, New York, 2001).

[64] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids, Second Edition* (Oxford University Press, Oxford, 2017).

[65] B. J. Alder and T. E. Wainwright. Phase Transition for a Hard Sphere System. J. Chem. Phys. **27**, 1208 (1957).

[66] A. Rahman. Correlations in the Motion of Atoms in Liquid Argon. Phys. Rev. **136**, A405 (1964).

[67] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087 (1953).

[68] E. P. Bernard, W. Krauth, and D. B. Wilson. Event-chain Monte Carlo algorithms for hard-sphere systems. Phys. Rev. E **80**, 056704 (2009).

[69] D. Frenkel. Simulations: The dark side. Eur. Phys. J. Plus **128**, 10 (2013).

[70] M. J. Cliffe, M. T. Dove, D. A. Drabold, and A. L. Goodwin. Structure Determination of Disordered Materials from Diffraction Data. Phys. Rev. Lett. **104**, 125501 (2010).

[71] K. Binder. Finite size effects on phase transitions. Ferroelectrics **73**, 43 (1987).

[72] M. Dijkstra, R. van Roij, and R. Evans. Phase diagram of highly asymmetric binary hard-sphere mixtures. Phys. Rev. E **59**, 5744 (1999).

[73] D. J. Ashton, N. B. Wilding, R. Roth, and R. Evans. Depletion potentials in highly size-asymmetric binary hard-sphere mixtures: Comparison of simulation results with theory. Phys. Rev. E **84**, 061136 (2011).

[74] W. C. K. Poon. Colloids as big atoms: the genesis of a paradigm. J. Phys. A Math. Theor. **49**, 401001 (2016).

[75] A. L. Thorneywork, J. L. Abbott, D. G. A. L. Aarts, and R. P. A. Dullens. Two-Dimensional Melting of Colloidal Hard Spheres. Phys. Rev. Lett. **118**, 158001 (2017).

[76] B. Comiskey, J. D. Albert, H. Yoshizawa, and J. Jacobson. An electrophoretic ink for all-printed reflective electronic displays. Nature (London) **394**, 253 (1998).

[77] A. C. Arsenault, D. P. Puzzo, I. Manners, and G. A. Ozin. Photonic-crystal full-colour displays. Nat. Photonics **1**, 468 (2007).

[78] M. Beija, R. Salvayre, N. Lauth-de Viguerie, and J.-D. Marty. Colloidal systems for drug delivery: from design to therapy. Trends Biotechnol. **30**, 485 (2012).

[79] J. de Vicente, D. J. Klingenberg, and R. Hidalgo-Alvarez. Magnetorheological fluids: a review. Soft Matter **7**, 3701 (2011).

[80] S. Klumpp (2011). *Small Things Considered: Of Terms in Biology: Colloids*. [online] Available at: ⟨https://schaechter.asmblog.org/schaechter/2011/02/of-terms-in-biology-colloids.html⟩ [Accessed 13 March 2020].

[81] W. W. Wood and J. D. Jacobson. Preliminary Results from a Recalculation of the Monte Carlo Equation of State of Hard Spheres. J. Chem. Phys. **27**, 1207 (1957).

[82] L. Onsager. The Effects of Shape on the Interaction of Colloidal Particles. Ann. N. Y. Acad. Sci. **51**, 627 (1949).

[83] P. Bolhuis and D. Frenkel. Tracing the phase boundaries of hard spherocylinders. J. Chem. Phys. **106**, 666 (1997).

[84] W. Poon. Colloids as Big Atoms. Science **304**, 830 (2004).

[85] M. T. Dove. *Structure and Dynamics: An Atomic View of Materials (Oxford Master Series in Physics)* (Oxford University Press, New York, 2003).

[86] J. C. Crocker and D. G. Grier. Microscopic measurement of the pair interaction potential of charge-stabilized colloid. Phys. Rev. Lett. **73**, 352 (1994).

[87] M. Brunner, J. Dobnikar, H.-H. von Grünberg, and C. Bechinger. Direct Measurement of Three-Body Interactions amongst Charged Colloids. Phys. Rev. Lett. **92**, 078301 (2004).

[88] M. Piech and J. Y. Walz. Direct measurement of depletion and structural forces in polydisperse, charged systems. J. Colloid Interface Sci. **253**, 117 (2002).

[89] C. Hertlein, L. Helden, A. Gambassi, S. Dietrich, and C. Bechinger. Direct measurement of critical Casimir forces. Nature (London) **451**, 172 (2008).

[90] K. Vondermassen, J. Bongers, A. Mueller, and H. Versmold. Brownian Motion: A Tool To Determine the Pair Potential between Colloid Particles. Langmuir **10**, 1351 (1994).

[91] W. Schommers. A pair potential for liquid rubidium from the pair correlation function. Phys. Lett. A **43**, 157 (1973).

[92] D. Levesque, J. J. Weis, and L. Reatto. Pair interaction from structural data for dense classical liquids. Phys. Rev. Lett. **54**, 451 (1985).

[93] M. Brunner, C. Bechinger, W. Strepp, V. Lobaskin, and H. H. von Grünberg. Density-dependent pair interactions in 2D. Europhys. Lett. **58**, 926 (2002).

[94] C. P. Royall, M. E. Leunissen, and A. Van Blaaderen. A new colloidal model system to study long-range interactions quantitatively in real space. J. Phys. Condens. Matter **15**, S3581 (2003).

[95] D. Reith, M. Pütz, and F. Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. J. Comput. Chem. **24**, 1624 (2003).

[96] T. C. Moore, C. R. Iacovella, and C. McCabe. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. J. Chem. Phys. **140**, 224104 (2014).

[97] A. E. Stones, R. P. A. Dullens, and D. G. A. L. Aarts. Model-Free Measurement of the Pair Potential in Colloidal Fluids Using Optical Microscopy. Phys. Rev. Lett. **123**, 098002 (2019).

[98] D. Lerche and T. Sobisch. Direct and Accelerated Characterization of Formulation Stability. J. Dispers. Sci. Technol. **32**, 1799 (2011).

[99] H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole, and S. J. Marrink. The power of coarse graining in biomolecular simulations. Wiley Interdiscip. Rev. Comput. Mol. Sci. **4**, 225 (2014).

[100] B. Widom. Some Topics in the Theory of Fluids. J. Chem. Phys. **39**, 2808 (1963).

[101] J. L. Jackson and L. S. Klein. Potential Distribution Method in Equilibrium Statistical Mechanics. Phys. Fluids **7**, 228 (1964).

[102] B. Widom. Laboring in the Vineyard of Physical Chemistry. Annu. Rev. Phys. Chem. **62**, 1 (2011).

[103] L. L. Lee and K. S. Shing. A test particle approach to the zero separation theorems of molecular distribution functions. J. Chem. Phys. **91**, 477 (1989).

[104] M. Llano-Restrepo and W. G. Chapman. Bridge function and cavity correlation function for the Lennard-Jones fluid from simulation. J. Chem. Phys. **97**, 2046 (1992).

[105] D. Ghonasgi, M. Llano-Restrepo, and W. G. Chapman. Henryâs law constant for diatomic and polyatomic Lennard-Jones molecules. J. Chem. Phys. **98**, 5662 (1993).

[106] M. Llano-Restrepo and W. G. Chapman. Bridge function and cavity correlation function for the soft sphere fluid from simulation: Implications on closure relations. J. Chem. Phys. **100**, 5139 (1994).

[107] R. Fantoni and G. Pastore. Computer simulation study of the closure relations in hard sphere fluids. J. Chem. Phys. **120**, 10681 (2004).

[108] D. L. Cheung, L. Anton, M. P. Allen, and A. J. Masters. Structure of molecular liquids: Cavity and bridge functions of the hard spheroid fluid. Phys. Rev. E **73**, 061204 (2006).

[109] T. R. Kunor and S. Taraphder. Bridge functions near the liquid-vapor coexistence curve in binary Lennard-Jones mixtures. Phys. Rev. E **74**, 011201 (2006).

[110] D. L. Cheung, L. Anton, M. P. Allen, and A. J. Masters. Computer simulation of liquids and liquid crystals. Comput. Phys. Commun. **179**, 61 (2008).

[111] J. Puibasset and L. Belloni. Bridge function for the dipolar fluid from simulation. J. Chem. Phys. **136**, 154503 (2012).

[112] I. Vyalov, G. Chuev, and N. Georgi. Solute-solvent cavity and bridge functions. I. Varying size of the solute. J. Chem. Phys. **141**, 074505 (2014).

[113] A. R. Saeger, J. K. Johnson, W. G. Chapman, and D. Henderson. Cavity correlation and bridge functions at high density and near the critical point: a test of second-order Percus-Yevick theory. Mol. Phys. **114**, 2516 (2016).

[114] J. K. Johnson, D. Henderson, S. Labík, and A. Malijevský. A comparison of the correlation functions of the Lennard-Jones fluid for the first-order Duh-Haymet-Henderson closure with molecular simulations. Mol. Phys. **115**, 1335 (2017).

[115] A. E. Stones, R. P. A. Dullens, and D. G. A. L. Aarts. Communication: Contact values of pair distribution functions in colloidal hard disks by test-particle insertion. J. Chem. Phys. **148**, 241102 (2018).

[116] J. A. Barker and D. Henderson. What is "liquid"? Understanding the states of matter. Rev. Mod. Phys. **48**, 587 (1976).

[117] G. M. Kepler and S. Fraden. Attractive potential between confined colloids at low ionic strength. Phys. Rev. Lett. **73**, 356 (1994).

[118] H. M. Ho, B. Lin, and S. A. Rice. Three-particle correlation functions of quasi-two-dimensional one-component and binary colloid suspensions. J. Chem. Phys. **125**, 184715 (2006).

[119] D. Henderson. Structure of the Triplet Distribution Function. J. Chem. Phys. **46**, 4306 (1967).

[120] Y.-T. Lee, F. H. Ree, and T. Ree. Distribution Function of Classical Fluids of Hard Spheres. I. J. Chem. Phys. **48**, 3506 (1968).

[121] F. H. Ree, Y.-T. Lee, and T. Ree. Distribution Function of Classical Fluids of Hard Spheres. II. J. Chem. Phys. **55**, 234 (1971).

[122] Y. Uehara, T. Ree, and F. H. Ree. Radial distribution function for hard disks from the BGY2 theory. J. Chem. Phys. **70**, 1876 (1979).

[123] Y. Uehara, Y.-T. Lee, T. Ree, and F. H. Ree. Triplet distribution functions for hard spheres and hard disks. J. Chem. Phys. **70**, 1884 (1979).

[124] A. D. J. Haymet, S. A. Rice, and W. G. Madden. An accurate integral equation for the pair and triplet distribution functions of a simple liquid. J. Chem. Phys. **74**, 3033 (1981).

[125] A. D. J. Haymet, S. A. Rice, and W. G. Madden. Triplet correlations in the Lennard-Jones fluid. J. Chem. Phys. **75**, 4696 (1981).

[126] M. P. Taylor and J. E. G. Lipson. On the Born-Green-Yvon equation and triplet distributions for hard spheres. J. Chem. Phys. **97**, 4301 (1992).

[127] R. Evans. Comment on Reverse Monte Carlo Simulation. Mol. Simul. **4**, 409 (1990).

[128] P. G. Bolhuis, A. A. Louis, and J. P. Hansen. Many-body interactions and correlations in coarse-grained descriptions of polymer solutions. Phys. Rev. E **64**, 021801 (2001).

[129] R. Evans. The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids. Adv. Phys. **28**, 143 (1979).

[130] J. S. Rowlinson and B. Widom. *Molecular Theory of Capillarity (Dover Books on Chemistry)* (Dover Publications, Mineola, NY, 2003).

[131] A. L. Thorneywork. *Structure and dynamics of two-dimensional colloidal hard spheres* (DPhil Thesis, University of Oxford, 2015).

[132] J. P. Sethna. *Statistical Mechanics: Entropy, Order Parameters and Complexity (Oxford Master Series in Physics)* (Oxford University Press, New York, 2006).

[133] R. D. Present. Non-additive interactions. Contemp. Phys. **12**, 595 (1971).

[134] B. M. Axilrod and E. Teller. Interaction of the van der Waals Type Between Three Atoms. J. Chem. Phys. **11**, 299 (1943).

[135] H. J. Raveché and R. D. Mountain. Triplet correlations. In *Progress in Liquid Physics*, edited by C. A. Croxton (J. Wiley & Sons. Ltd., Chichester, 1978).

[136] A. A. Schekochihin (Compiled on 13 April 2020). *Lectures on Kinetic Theory of Gases and Statistical Physics (Oxford Physics Paper A1). University of Oxford.* [online] Available at: ⟨http://www-thphys.physics.ox.ac.uk/people/AlexanderSchekochihin/A1/2014/A1LectureNotes.pdf⟩ [Accessed 29 April 2020].

[137] J. Rowlinson. Intermolecular potentials that are functions of thermodynamic variables. Mol. Phys. **52**, 567 (1984).

[138] J. K. Percus. Approximation Methods in Classical Statistical Mechanics. Phys. Rev. Lett. **8**, 462 (1962).

[139] J. K. Percus. The pair distribution function in classical statistical mechanics. In *Equilibrium Theory of Classical Fluids*, edited by H. L. Frisch and J. L. Lebowitz (W A Benjamin, Inc., New York, 1964).

[140] M. Puoskari. Exact Self-consistent Integral Equations for the Distribution Functions of Classical Fluids. Phys. Chem. Liq. **39**, 201 (2001).

[141] O. A. Pizio. On the simple derivation of approximate integral equations for triplet and higher-order distribution functions of homogeneous fluids. Mol. Phys. **77**, 1011 (1992).

[142] M. F. Holovko, O. A. Pizio, A. D. J. Haymet, and D. Henderson. On the application of the theory of inhomogeneous fluids for the triplet and higher order distribution functions. J. Mol. Liq. **58**, 101 (1993).

[143] L. L. Lee. A potential distribution approach to fused heterochain molecules. I. Mixtures of hard dumbbells and spheres. J. Chem. Phys. **103**, 4221 (1995).

[144] J. F. Robinson, F. Turci, R. Roth, and C. P. Royall. Morphometric Approach to Many-Body Correlations in Hard Spheres. Phys. Rev. Lett. **122**, 068004 (2019).

[145] B. Widom. Structure of interfaces from uniformity of the chemical potential. J. Stat. Phys. **19**, 563 (1978).

[146] J. C. Brown, P. N. Pusey, J. W. Goodwin, and R. H. Ottewill. Light scattering study of dynamic and time-averaged correlations in dispersions of charged particles. J. Phys. A Math. Gen. **8**, 664 (1975).

[147] P. N. Pusey. The dynamics of interacting Brownian particles. J. Phys. A Math. Gen. **8**, 1433 (1975).

[148] M. D. Haw. Colloidal suspensions, Brownian motion, molecular reality: a short history. J. Phys. Condens. Matter **14**, 7769 (2002).

[149] H. Löwen and E. Allahyarov. The role of effective triplet interactions in charged colloidal suspensions. J. Phys. Condens. Matter **10**, 4147 (1998).

[150] M. Dijkstra, R. van Roij, R. Roth, and A. Fortini. Effect of many-body interactions on the bulk and interfacial phase behavior of a model colloid-polymer mixture. Phys. Rev. E **73**, 041404 (2006).

[151] B. V. Derjaguin and L. Landau. Theory of the Stability of Strongly Charged Lyophobic Sols and of the Adhesion of Strongly Charged Particles in Solutions of Electrolytes. Acta Physicochim. URSS **14**, 633 (1941). Cited in [39].

[152] E. J. W. Verwey and J. Th. G. Overbeek. *Theory of Stability of Lyophobic Colloids* (Elsevier, Amsterdam, 1948). Cited in [39].

[153] R. van Roij and J.-P. Hansen. Van der Waals–Like Instability in Suspensions of Mutually Repelling Charged Colloids. Phys. Rev. Lett. **79**, 3082 (1997).

[154] R. van Roij, M. Dijkstra, and J.-P. Hansen. Phase diagram of charge-stabilized colloidal suspensions: van der Waals instability without attractive forces. Phys. Rev. E **59**, 2010 (1999).

[155] R. van Roij and R. Evans. Phase equilibria in a model of low-salt suspensions of charged colloids. J. Phys. Condens. Matter **11**, 10047 (1999).

[156] J.-M. Victor and J.-P. Hansen. Spinodal decomposition and the liquid-vapour equilibrium in charged colloidal dispersions. J. Chem. Soc., Faraday Trans. 2 **81**, 43 (1985).

[157] J. Ge, L. He, Y. Hu, and Y. Yin. Magnetically induced colloidal assembly into field-responsive photonic structures. Nanoscale **3**, 177 (2011).

[158] B. F. Edwards, D. M. Riffe, J.-Y. Ji, and W. A. Booth. Interactions between uniformly magnetized spheres. Am. J. Phys. **85**, 130 (2017).

[159] R. Piazza, T. Bellini, and V. Degiorgio. Equilibrium sedimentation profiles of screened charged colloids: A test of the hard-sphere equation of state. Phys. Rev. Lett. **71**, 4267 (1993).

[160] J. Zhu, M. Li, R. Rogers, W. Meyer, R. H. Ottewill, STS-73 Space Shuttle Crew, W. B. Russel, and P. M. Chaikin. Crystallization of hard-sphere colloids in microgravity. Nature (London) **387**, 883 (1997).

[161] N. J. B. Green. MChem Chemistry Tutorial, Oriel College, University of Oxford. [Private Communication, December 2014].

[162] V. N. Michailidou, G. Petekidis, J. W. Swan, and J. F. Brady. Dynamics of Concentrated Hard-Sphere Colloids Near a Wall. Phys. Rev. Lett. **102**, 068302 (2009).

[163] T. O. E. Skinner. *Two-dimensional colloidal systems: grain boundaries and confinement* (DPhil Thesis, University of Oxford, 2012).

[164] M. P. N. Juniper. *Dynamics of driven colloidal systems in one-dimensional potential energy landscapes* (DPhil Thesis, University of Oxford, 2014).

[165] F. A. Lavergne. *Dynamics of grain boundaries in two-dimensional colloidal crystals* (DPhil Thesis, University of Oxford, 2016).

[166] I. S. Khattab, F. Bandarkar, M. A. A. Fakhree, and A. Jouyban. Density, viscosity, and surface tension of water+ethanol mixtures from 293 to 323K. Korean J. Chem. Eng. **29**, 812 (2012).

[167] R. Roth, K. Mecke, and M. Oettel. Communication: Fundamental measure theory for hard disks: Fluid and solid. J. Chem. Phys. **136**, 081101 (2012).

[168] Thermo Fisher Scientific (2007). *Dynabeads® M-270 Carboxylic Acid*. [online] Available at: ⟨https://www.thermofisher.com/us/en/home/references/protocols/proteins-expression-isolation-and-analysis/protein-isolation-protocol/m-270-carboxylic-acid.html⟩ [Accessed 28 January 2020].

[169] D. Allan, T. Caswell, N. Keim, C. van der Wel, and Trackpy v0.3.0 and Trackpy v0.3.2 Contributors. trackpy: Trackpy v0.3.0 and Trackpy v0.3.2 (2015, 2016). [software] Available at: ⟨https://github.com/soft-matter/trackpy⟩ [Accessed 28 April 2020].

[170] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors. scikit-image: Image processing in Python. PeerJ **2**, e453 (2014).

[171] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids (Oxford Science Publications)* (Clarendon Press, Oxford, 1989).

[172] W. Krauth. *Statistical Mechanics: Algorithms and Computations (Oxford Master Series in Physics)* (Oxford University Press, New York, 2006).

[173] M. Michel, S. C. Kapfer, and W. Krauth. Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. J. Chem. Phys. **140**, 054116 (2014).

[174] M. Engel, J. A. Anderson, S. C. Glotzer, M. Isobe, E. P. Bernard, and W. Krauth. Hard-disk equation of state: First-order liquid-hexatic transition in two dimensions with three simulation methods. Phys. Rev. E **87**, 042134 (2013).

[175] M. Isobe and W. Krauth. Hard-sphere melting and crystallization with event-chain Monte Carlo. J. Chem. Phys. **143**, 084509 (2015).

[176] E. P. Bernard and W. Krauth. Two-Step Melting in Two Dimensions: First-Order Liquid-Hexatic Transition. Phys. Rev. Lett. **107**, 155704 (2011).

[177] E. P. Bernard. *Algorithms and applications of the Monte Carlo method: Two-dimensional melting and perfect sampling* (Ph.D. Thesis, École Normale Supérieure, 2011).

[178] W. Krauth (2011). *Bernard Krauth Wilson 2009 - Werner KRAUTH. Python Implementation*. [online] Available at: ⟨http://www.lps.ens.fr/∼krauth/index.php/Bernard_Krauth_Wilson_2009⟩ [Accessed 29 April 2020].

[179] B. J. Alder and T. E. Wainwright. Phase Transition in Elastic Disks. Phys. Rev. **127**, 359 (1962).

[180] P. N. Pusey and W. Van Megen. Phase behaviour of concentrated suspensions of nearly hard colloidal spheres. Nature (London) **320**, 340 (1986).

[181] E. G. Teich, G. van Anders, and S. C. Glotzer. Identity crisis in alchemical space drives the entropic colloidal glass transition. Nat. Commun. **10** (2019).

[182] D. Frenkel. Order through entropy. Nat. Mater. **14**, 9 (2014).

[183] D. Chandler, J. D. Weeks, and H. C. Andersen. Van der Waals Picture of Liquids, Solids, and Phase Transformations. Science **220**, 787 (1983).

[184] J. D. Weeks, D. Chandler, and H. C. Andersen. Perturbation Theory of the Thermodynamic Properties of Simple Liquids. J. Chem. Phys. **55**, 5422 (1971).

[185] H. C. Andersen, J. D. Weeks, and D. Chandler. Relationship between the Hard-Sphere Fluid and Fluids with Realistic Repulsive Forces. Phys. Rev. A **4**, 1597 (1971).

[186] A. J. Archer, B. Chacko, and R. Evans. The standard mean-field treatment of inter-particle attraction in classical DFT is better than one might expect. J. Chem. Phys. **147**, 034501 (2017).

[187] Y. Rosenfeld. Free-energy model for the inhomogeneous hard-sphere fluid mixture and density-functional theory of freezing. Phys. Rev. Lett. **63**, 980 (1989).

[188] Y. Rosenfeld, D. Levesque, and J.-J. Weis. Free-energy model for the inhomogeneous hard-sphere fluid mixture: Triplet and higher-order direct correlation functions in dense fluids. J. Chem. Phys. **92**, 6818 (1990).

[189] H. Hansen-Goos and R. Roth. Density functional theory for hard-sphere mixtures: the White Bear version mark II. J. Phys. Condens. Matter **18**, 8413 (2006).

[190] M. López de Haro, E. G. D. Cohen, and J. M. Kincaid. The Enskog theory for multicomponent mixtures. I. Linear transport theory. J. Chem. Phys. **78**, 2746 (1983).

[191] B. C. Freasier. Equation of state of fused hard spheres revisited. Mol. Phys. **39**, 1273 (1980).

[192] J. Tobochnik and P. M. Chapin. Monte Carlo simulation of hard spheres near random closest packing using spherical boundary conditions. J. Chem. Phys. **88**, 5824 (1988).

[193] M. Barošová, M. Malijevský, S. Labík, and W. R. Smith. Computer simulation of the chemical potentials of binary hard-sphere mixtures. Mol. Phys. **87**, 423 (1996).

[194] D. Cao, K. Chan, D. Henderson, and W. Wang. Monte Carlo data of dilute solutions of large spheres in binary hard sphere mixtures. Mol. Phys. **98**, 619 (2000).

[195] M. González-Melchor, J. Alejandre, and M. López de Haro. Equation of state and structure of binary mixtures of hard d-dimensional hyperspheres. J. Chem. Phys. **114**, 4905 (2001).

[196] J. R. Henderson. On the test particle approach to the statistical mechanics of fluids. Mol. Phys. **48**, 389 (1983).

[197] H. Reiss, H. L. Frisch, and J. L. Lebowitz. Statistical Mechanics of Rigid Spheres. J. Chem. Phys. **31**, 369 (1959).

[198] E. Helfand, H. L. Frisch, and J. L. Lebowitz. Theory of the Two- and One-Dimensional Rigid Sphere Fluids. J. Chem. Phys. **34**, 1037 (1961).

[199] D. A. Kofke and P. T. Cummings. Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation. Mol. Phys. **92**, 973 (1997).

[200] R. J. Speedy. Statistical geometry of hard-sphere systems. J. Chem. Soc., Faraday Trans. 2 **76**, 693 (1980).

[201] A. Santos, S. B. Yuste, and M. López de Haro. Contact values of the radial distribution functions of additive hard-sphere mixtures in d dimensions: A new proposal. J. Chem. Phys. **117**, 5785 (2002).

[202] J. L. Bentley. Multidimensional Binary Search Trees Used for Associative Searching. Commun. ACM **18**, 509 (1975).

[203] J. H. Friedman, J. L. Bentley, and R. A. Finkel. An Algorithm for Finding Best Matches in Logarithmic Expected Time. ACM Trans. Math. Softw. **3**, 209–226 (1977).

[204] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods **17**, 261 (2020).

[205] S. Maneewongvatana and D. M. Mount. Analysis of approximate nearest neighbor searching with clustered point sets. In *Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges, Proceedings of a DIMACS Workshop, USA, 1999*, edited by M. H. Goldwasser, D. S. Johnson, and C. C. McGeoch, volume 59 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 105–123 (American Mathematical Society, Providence, RI, 2002). Cited in [279].

[206] G. Torrie and G. N. Patey. Monte Carlo calculation of $y(r)$ for the hard-sphere fluid. Mol. Phys. **34**, 1623 (1977).

[207] G. P. Brenan and R. Evans. Free energy functionals and the structure of the uniform hard-sphere fluid. Mol. Phys. **73**, 789 (1991).

[208] W. G. Hoover and J. C. Poirier. Determination of Virial Coefficients from the Potential of Mean Force. J. Chem. Phys. **37**, 1041 (1962).

[209] E. Meeron and A. J. F. Siegert. Statistical Mechanics of Hard-Particle Systems. J. Chem. Phys. **48**, 3139 (1968).

[210] A. Santos, M. López de Haro, and S. B. Yuste. An accurate and simple equation of state for hard disks. J. Chem. Phys. **103**, 4622 (1995).

[211] E. W. Grundke and D. Henderson. Distribution functions of multi-component fluid mixtures of hard spheres. Mol. Phys. **24**, 269 (1972).

[212] M. D. Carbajal-Tinoco, F. Castro-Román, and J. L. Arauz-Lara. Static properties of confined colloidal suspensions. Phys. Rev. E **53**, 3745 (1996).

[213] R. Rajagopalan and K. Srinivasa Rao. Interaction forces in charged colloids: Inversion of static structure factors. Phys. Rev. E **55**, 4423 (1997).

[214] S. H. Behrens and D. G. Grier. Pair interaction of charged colloidal spheres near a charged wall. Phys. Rev. E **64**, 050401 (2001).

[215] M. Quesada-Pérez, A. Moncho-Jordá, F. Martínez-López, and R. Hidalgo-Álvarez. Probing interaction forces in colloidal monolayers: Inversion of structural data. J. Chem. Phys. **115**, 10897 (2001).

[216] R. L. Henderson. A uniqueness theorem for fluid pair correlation functions. Phys. Lett. A **49**, 197 (1974).

[217] J. T. Chayes and L. Chayes. On the validity of the inverse conjecture in classical density functional theory. J. Stat. Phys. **36**, 471 (1984).

[218] W. Schommers. Pair potentials in disordered many-particle systems: A study for liquid gallium. Phys. Rev. A **28**, 3599 (1983).

[219] L. Reatto, D. Levesque, and J. J. Weis. Iterative predictor-corrector method for extraction of the pair interaction from structural data for dense classical liquids. Phys. Rev. A **33**, 3451 (1986).

[220] A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. Phys. Rev. E **52**, 3730 (1995).

[221] A. Soper. Empirical potential Monte Carlo simulation of fluid structure. Chem. Phys. **202**, 295 (1996).

[222] A. P. Lyubartsev. Inverse Monte Carlo Methods. In *Coarse-Grained Modeling of Biomolecules*, edited by G. A. Papoian (CRC Press, Boca Raton, FL, 2017).

[223] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko. Versatile Object-Oriented Toolkit for Coarse-Graining Applications. J. Chem. Theory Comput. **5**, 3211 (2009).

[224] N. G. Almarza and E. Lomba. Determination of the interaction potential from the pair distribution function: An inverse Monte Carlo technique. Phys. Rev. E **68**, 011202 (2003).

[225] C. Russ, M. Brunner, C. Bechinger, and H. H. von Grünberg. Three-body forces at work: Three-body potentials derived from triplet correlations in colloidal suspensions. Europhys. Lett. **69**, 468 (2005).

[226] J. Philip (2007). *The Probability Distribution of the Distance Between Two Random Points in a Box*. [online] Available at: ⟨https://people.kth.se/∼johanph/habc.pdf⟩ [Accessed 24 April 2020].

[227] E. N. Rudisill and P. T. Cummings. Gibbs ensemble simulation of phase equilibrium in the hard core two-Yukawa fluid model for the Lennard-Jones fluid. Mol. Phys. **68**, 629 (1989).

[228] R. H. Swendsen. How the maximum step size in Monte Carlo simulations should be adjusted. Phys. Procedia **15**, 81 (2011).

[229] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. A. van der Vegt. Systematic coarse-graining methods for soft matter simulations – a review. Soft Matter **9**, 2108 (2013).

[230] J. A. Barker, D. Henderson, and W. R. Smith. Three-Body Forces in Dense Systems. Phys. Rev. Lett. **21**, 134 (1968).

[231] J. A. Barker, D. Henderson, and W. R. Smith. Pair and triplet interactions in argon. Mol. Phys. **17**, 579 (1969).

[232] J. A. Barker, R. A. Fisher, and R. O. Watts. Liquid argon: Monte carlo and molecular dynamics calculations. Mol. Phys. **21**, 657 (1971).

[233] J. G. Kirkwood. Statistical Mechanics of Fluid Mixtures. J. Chem. Phys. **3**, 300 (1935).

[234] P. Schofield. Wavelength-dependent fluctuations in classical fluids: I. The long wavelength limit. Proc. Phys. Soc. **88**, 149 (1966).

[235] R. E. Nettleton and M. S. Green. Expression in Terms of Molecular Distribution Functions for the Entropy Density in an Infinite System. J. Chem. Phys. **29**, 1365 (1958).

[236] H. J. Raveché. Entropy and Molecular Correlation Functions in Open Systems. I. Derivation. J. Chem. Phys. **55**, 2242 (1971).

[237] A. Baranyai and D. J. Evans. Direct entropy calculation from computer simulation of liquids. Phys. Rev. A **40**, 3817 (1989).

[238] H. J. Raveché, R. D. Mountain, and W. B. Streett. Freezing and melting properties of the Lennard-Jones system. J. Chem. Phys. **61**, 1970 (1974).

[239] C. Ruß, K. Zahn, and H.-H. von Grünberg. Triplet correlations in two-dimensional colloidal model liquids. J. Phys. Condens. Matter **15**, S3509 (2003).

[240] F. C. Frank. Supercooling of liquids. Proc. R. Soc. London, Ser. A **215**, 43 (1952).

[241] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti. Bond-orientational order in liquids and glasses. Phys. Rev. B **28**, 784 (1983).

[242] F. Spaepen. Five-fold symmetry in liquids. Nature (London) **408**, 781 (2000).

[243] M. Leocmach and H. Tanaka. Roles of icosahedral and crystal-like order in the hard spheres glass transition. Nat. Commun. **3** (2012).

[244] J. Taffs and C. P. Royall. The role of fivefold symmetry in suppressing crystallization. Nat. Commun. **7** (2016).

[245] B. J. Alder. Triplet Correlations in Hard Spheres. Phys. Rev. Lett. **12**, 317 (1964).

[246] A. Rahman. Triplet Correlations in Liquids. Phys. Rev. Lett. **12**, 575 (1964).

[247] J. A. Krumhansl and S.-s. Wang. Triplet Correlation in Liquid Argon by Monte Carlo Method: Low Densities. J. Chem. Phys. **56**, 2034 (1972).

[248] S.-s. Wang and J. A. Krumhansl. Superposition Assumption. II. High Density Fluid Argon. J. Chem. Phys. **56**, 4287 (1972).

[249] H. J. Raveché, R. D. Mountain, and W. B. Streett. Three Atom Correlations in the Lennard-Jones Fluid. J. Chem. Phys. **57**, 4999 (1972).

[250] R. Block and W. Schommers. Triplet correlations in disordered systems: A study for liquid rubidium. J. Phys. C **8**, 1997 (1975).

[251] M. Tanaka and Y. Fukui. Simulation of the Three-Particle Distribution Function in a Long-Range Oscillatory Potential Liquid. Prog. Theor. Phys. **53**, 1547 (1975).

[252] W. J. McNeil, W. G. Madden, A. D. J. Haymet, and S. A. Rice. Triplet correlation functions in the Lennard-Jones fluid: Tests against molecular dynamics simulations. J. Chem. Phys. **78**, 388 (1983).

[253] H. J. Raveché and R. D. Mountain. Three Atom Correlations in Liquid Neon. J. Chem. Phys. **57**, 3987 (1972).

[254] M. C. Abramo and M. P. Tosi. Triplet correlations in liquid argon. Lett. Nuovo Cimento **5**, 1044 (1972).

[255] P. Attard. An improved Kirkwood superposition approximation for three atoms in rolling contact. Mol. Phys. **74**, 547 (1991).

[256] P. Attard. Spherically inhomogeneous fluids. I. Percus-Yevick hard spheres: Osmotic coefficients and triplet correlations. J. Chem. Phys. **91**, 3072 (1989).

[257] P. Walker, U. Krohn, and D. Carty. ARBTools: A Tricubic Spline Interpolator for Three-Dimensional Scalar or Vector Fields. J. Open Res. Softw. **7**, 12 (2019).

[258] P. Attard and G. Stell. Three-particle correlations in a hard-sphere fluid. Chem. Phys. Lett. **189**, 128 (1992).

[259] E. A. Müller and K. E. Gubbins. Triplet correlation function for hard sphere systems. Mol. Phys. **80**, 91 (1993).

[260] Y. Kalyuzhnyi, J. Škvára, and I. Nezbeda. Analytic results for the three- and four-particle correlation functions of the fluid of hard disks. J. Chem. Phys. **150**, 034502 (2019).

[261] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. J. Chem. Phys. **54**, 5237 (1971).

[262] A. E. Stones and D. G. A. L. Aarts. Measuring Many-Body Distribution Functions in Fluids using Test-Particle Insertion, arXiv:2007.04016v1.

[263] J. D. Hunter. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. **9**, 90 (2007).

[264] C. Russ, H. H. von Grünberg, M. Dijkstra, and R. van Roij. Three-body forces between charged colloidal particles. Phys. Rev. E **66**, 011402 (2002).

[265] J. Dobnikar, M. Brunner, H.-H. von Grünberg, and C. Bechinger. Three-body interactions in colloidal systems. Phys. Rev. E **69**, 031402 (2004).

[266] M. C. Bellissent-Funel, P. Chieux, D. Levesque, and J. J. Weis. Structure factor and effective two-body potential for liquid gallium. Phys. Rev. A **39**, 6310 (1989).

[267] L. Reatto and M. Tau. Interatomic dispersion forces and small angle scattering from rare gas fluids. J. Phys. Condens. Matter **4**, 1 (1992).

[268] R. Magli, F. Barocchi, P. Chieux, and R. Fontana. Experimental Determination of the Long-Range Potential of Argon Pairs by Means of Small-Angle Neutron Diffraction. Phys. Rev. Lett. **77**, 846 (1996).

[269] F. Formisano, C. J. Benmore, U. Bafile, F. Barocchi, P. A. Egelstaff, R. Magli, and P. Verkerk. Long Range Potential Effects in Low Density Krypton Gas. Phys. Rev. Lett. **79**, 221 (1997).

[270] F. Formisano, F. Barocchi, and R. Magli. Long-range interactions in xenon. Phys. Rev. E **58**, 2648 (1998).

[271] E. Guarini, G. Casanova, U. Bafile, and F. Barocchi. Structure of gaseous Kr in the low-q region by neutron scattering and interaction potentials. Phys. Rev. E **60**, 6682 (1999).

[272] C. J. Benmore, F. Formisano, R. Magli, U. Bafile, P. Verkerk, P. A. Egelstaff, and F. Barocchi. Neutron diffraction study of long-range interactions in gaseous krypton. J. Phys. Condens. Matter **11**, 3091 (1999).

[273] E. Guarini, G. Casanova, U. Bafile, F. Formisano, R. Magli, and F. Barocchi. Direct determination of long-range dipolar interactions in Kr by neutron diffraction. Europhys. Lett. **49**, 62 (2000).

[274] R. Klein, H. H. von Grünberg, C. Bechinger, M. Brunner, and V. Lobaskin. Macroion shielding and state-dependent pair potentials in colloidal suspensions. J. Phys. Condens. Matter **14**, 7631 (2002).

[275] V. Lobaskin, M. Brunner, C. Bechinger, and H. H. von Grünberg. On the nature of long-range contributions to pair interactions between charged colloids in two dimensions. J. Phys. Condens. Matter **15**, 6693 (2003).

[276] S. Paladugu, A. Callegari, Y. Tuna, L. Barth, S. Dietrich, A. Gambassi, and G. Volpe. Nonadditivity of critical Casimir forces. Nat. Commun. **7** (2016).

[277] L. L. Lee. The test-particle induced inhomogeneous direct correlation functions and extensions of Widom's theorem: Impacts on the incremental chemical potentials and high-order correlation functions. J. Chem. Phys. **139**, 154501 (2013).

[278] D. S. Corti and R. K. Bowles. Statistical geometry of hard sphere systems: exact relations for additive and non-additive mixtures. Mol. Phys. **96**, 1623 (1999).

[279] The SciPy community (2019). SciPy Reference Guide: Release 1.4.1. [online] Available at: ⟨https://docs.scipy.org/doc/scipy-1.4.1/scipy-ref-1.4.1.pdf⟩ [Accessed 25th April 2020].

List of Publications

To date the following publications have been published based on the work in this thesis:

- A. E. Stones, R. P. A. Dullens and D. G. A. L. Aarts. Communication: Contact values of pair distribution functions in colloidal hard disks by test-particle insertion. J. Chem. Phys. **148**, 241102 (2018).
- A. E. Stones, R. P. A. Dullens and D. G. A. L. Aarts. Model-free measurement of the pair potential in colloidal fluids using optical microscopy. Phys. Rev. Lett. **123**, 098002 (2019).

The following publication is in preparation:

- A. E. Stones and D. G. A. L. Aarts. Measuring many-body distribution functions in fluids using test-particle insertion.

Acknowledgements

Firstly, I would like to say thanks to Dirk and Roel for their support as my supervisors over the last few years. In particular, I would like to thank Dirk for his ceaseless enthusiasm towards this work throughout, and especially for his support when I have been writing things up. We share the habit of getting carried away sometimes with exciting new ideas, which can be quite infectious and has benefitted this thesis tremendously. He has also taught me, however, to identify when it is time to come back to reality, write things down and move on. I would also like to thank Roel for his guidance and for always being available to chat. His ability to look at things from a different perspective has been tremendously useful in helping me to refine and communicate my ideas, and the group trip to the *Liquid Matter Conference* in Ljubljana will live long in my memory. I would also like to thank some other inspirational role models: Bob Evans, who has been very encouraging and invited me to give my first external research seminar, and my undergraduate tutors Nick Green and Andrew Goodwin, who are always the source of great chats and advice, scientific or otherwise.

I am incredibly grateful to the members of the Oxford Colloid Group, past and present, who have not only supported this work with scientific discussions, demonstrations, suggestions and advice, but have also provided a friendly and enjoyable environment in which to work. I have been fortunate to supervise two fantastic Master's degree students: Alice and Sacha, and I would like to thank Arran in particular for his help and support with Alice's project. Thanks also to those with whom I have shared an office over the years, who have given me daily support and motivation and got me interested in many different problems—especially Lachlan, who was there during the writing-up period. I've particularly enjoyed working on joint ideas with Carla—we complement each other's interests well which makes for very engaging projects and discussions.

I also give thanks to everyone who has helped me to develop personally during my DPhil.

I count myself extremely lucky to have truly fantastic friends, many of whom I have met during this period. These include housemates, friends from college (Hertford and Oriel), and lots of other people who I have met along the way. In the first two years, I taught at Oriel and St Anne's Colleges, and both SCRs were great communities to be in: I'd like to thank Sean and Juliane for their help in settling in, and David and Andrew for their support as organising tutors. I want to mention two friends in particular. Carla, you have been there through thick and thin. Nothing I write here can come close to expressing how grateful I am, and I just want to say thanks for everything. Mark, I've known you since my first day in Oxford, and it's been great to have you as a housemate. Thanks especially for keeping me motivated when I need it, and for putting up with me when I've been writing up.

Finally, I want to say thanks to all the members of my family, who are only ever a phone call or a train ride away, and have supported and encouraged me all along. I couldn't have done any of it without you.