Fluid Density Approximation for an Implicit Solvent Model

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch–Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Lukas Jager

aus

Niederkassel

Bonn 2007

Angefertigt mit Genehmigung der Mathematisch–Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich–Wilhelms–Universität Bonn

1. Referent: Prof. Dr. Michael Griebel

2. Referent: Prof. Dr. Rolf Krause

Tag der Promotion: 13.12.2007

Erscheinungsjahr: 2007

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Zusammenfassung

Wir beschäftigen uns in der vorliegenden Arbeit mit der mikroskopischen Simulation molekularer Systeme in Lösung. Die explizite Simulation des Gesamtsystems ist wegen der hohen Zahl der Lösungsmittelmoleküle zu aufwendig. Daher wird der Einfluss des Lösungsmittels implizit mittels des sogenannten Potential of Mean Force (PMF) berücksichtigt. Die effiziente Berechnung eines solchen impliziten Lösungsmittelmodells ermöglicht dann eine effizientere Simulation des Gesamtsystems von gelöstem Stoff und Lösungsmittel. Vielversprechende Ansätze, das PMF näherungsweise zu berechnen, bieten Integralgleichungsmethoden zur Approximation der mittleren Dichte von Flüssigkeiten. Existierende Verfahren basieren praktisch ausnahmslos auf der Ornstein-Zernike Gleichung, können allerdings aufgrund des immer noch hohen numerischen Aufwands und der benutzten Näherungen nicht als effizientes implizites Lösungsmittelmodell verwendet werden.

Wir leiten daher ausgehend von der YBG-Hierarchie der statistischen Physik ein neues Modell, das sogenannte BGY3d Modell, für die Approximation der Dichte atomarer Lösungsmittel um einen gelösten Stoff herum her. Wir verwenden dabei zum ersten mal in diesem Zusammenhang die Kirkwood Approximation. Mittels eines speziellen Produktansatzes ist es uns möglich, das BGY3d Modell numerisch sehr effizient zu lösen. Der Rechenaufwand erweist sich als deutlich geringer als der des 3d-HNC Verfahrens von Beglov und Roux, welches auf der Ornstein-Zernike Gleichung basiert. Bei diesem Vergleich stellt sich ebenfalls heraus, dass die Kirkwood Approximation zur gleichen Gesamtgenauigkeit der Ergebnisse führt, dabei allerdings dem 3d-HNC Verfahren in der Näherung der Höhe und der Position des ersten Maximums der Dichteverteilung überlegen ist.

Um auch die Dichteverteilung realistischer Lösungsmittel berechnen zu können, erweitern wir unser Modell so, dass auch molekulare Flüssigkeiten als Lösungsmittel betrachtet werden können. Neben der Kirkwood Approximation für die Interaktionen zwischen den Molekülen, verwenden wir nun die sogenannte Normalized Site-Site Superposition Approximation von Taylor und Lipson für die Interaktionen innerhalb der Lösungsmittelmoleküle. Außerdem ist das molekulare BGY3d Modell (BGY3dM) in der Lage, die Dichte von Lösungsmitteln, die sowohl durch kurzreichweitige Potentiale, wie z.B. das Lennard-Jones Potential, als auch durch das langreichweitige Coulomb Potential beschrieben werden, effizient zu berechnen. Der Vergleich von Ergebnissen des BGY3dM Modells mit Ergebnissen aus Moleküldynamiksimulationen zeigt, dass unsere Modellierung und die daraus resultierende Dichteverteilung eine gute Approximation liefern. Der Rechenaufwand für das BGY3dM Verfahren ist dabei zwei bis drei Größenordnungen kleiner als der einer Moleküldynamiksimulation. In numerischen Beispielrechnungen führt die Anwendung unseres BGY3dM Verfahrens auf die Berechnung der Dichteverteilung von Kohlenstoffdisulfid um verschiedene Moleküle herum zu realistischen Dichte- und Ladungsverteilungen.

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei allen Personen bedanken, die mir bei der Fertigstellung dieser Arbeit stets mit Rat und Tat zur Seite gestanden haben. Dieser Dank gilt zunächst meinem Betreuer, Prof. Dr. Michael Griebel, der mir in unzähligen Diskussionen neue Ideen und Denkanstöße gegeben und mich oft wieder auf den richtigen Weg gebracht hat. Bei Prof. Dr. Rolf Krause möchte ich mich herzlich für die Übernahme des Zweitgutachtens bedanken. Ebenfalls ein ganz besonderer Dank gebührt Dr. Marc Alexander Schweitzer, der stets ein offenes Ohr für jegliche Probleme hatte und es immer verstanden hat, mich neu zu motivieren, wenn es notwendig war. Bei Gabriela Constantinescu, Frederik Heber, Dr. Marc Alexander Schweitzer und Ralf Wildenhues möchte ich mich für das Korrekturlesen meiner Arbeit bedanken. Nicht zuletzt gilt mein Dank auch meinem ehemaligen Kollegen Dr. Daniel Oeltz, mit dem ich jede noch so absurde Idee besprechen konnte und ohne den der Arbeitsalltag viel an Freude und Enthusiasmus verloren hat. Abschließend möchte ich mich bei allen Kollegen und Mitarbeitern des Instituts für Numerische Simulation für die freundliche und inspirierende Arbeitsatmosphäre bedanken.

Bonn, im September 2007 Lukas Jager

Contents

Chapter 1
Introduction

Proteins are essential parts of any organism and participate in every process within cells. For instance, they work as catalysts for biochemical reactions (enzymes), as oxygen carriers in our blood cells (hemoglobin) and are responsible for our skin and eye color (pigments). All enzymes, pigments, hormones etc. are proteins. Hence, the understanding of their structure and function is an important step in order to understand life.

Twenty different amino acids constitute the basis of which the proteins are built as linear polymers. The linear composition of a protein is easy to obtain. But its biological functionality strongly depends on the three-dimensional configuration of the protein. Only in this functional form the protein can interact with other molecules e.g. by a key and lock mechanism. Its three-dimensional structure is fully governed by the linear sequence of its amino acids. In nature, after a protein is assembled from the amino acids as a linear chain, it folds up in a matter of seconds or minutes to take on its functional three-dimensional conformation. But how the specific amino acid sequence affects the folded structure of a protein is by far not understood, yet. Hence, the protein-folding problem remains one of the most fundamental unsolved problems of molecular biology and is a key-topic of current research.

Computer simulations are an effective tool to investigate the protein-folding problem. The ultimate goal is to compute the native state of a protein, i.e. its three-dimensional shape, from the knowledge of the amino acid sequence alone. For this, a detailed representation of the protein as well as its natural environment needs to be implemented in the computer. This environment is a liquid solution. It cannot be neglected since the interaction between the protein and the solution plays an important role for the protein's structure. But the efficient incorporation of the solvent effects into the computer simulation is a major challenge. In a naive approach, the solute, i.e. the protein, as well as the solvent, e.g. water, are considered by a fully atomistic representation. This clearly leads to a very detailed description of

the solute-solvent interaction. But the explicit simulation of the solvent then requires the major part of the computational effort. Hence, this approach is merely applicable to small solutes but is simply unfeasible if an extensive simulation of a protein is required as this would be the case in a protein-folding simulation.

Currently, the most promising approach to overcome these complexity issues is to include the solvent effects by a so-called implicit solvent model. In such a model, the solute-solvent interactions are approximated without introducing new degrees of freedom to the system. Existing implicit solvent models yield qualitatively good results, yet are not able to reproduce experimental data on a quantitative level of detail. First attempts to design accurate implicit solvent models were made by employing methods based on the liquid state integral equation theories. Hirata, Rossky and Pettitt [51] were the first to formulate a method applicable to solute-solvent systems in 1983. Hereafter, many authors considered the integral equation theories with respect to the approximation of solvent effects for simple solutes. In principle, these methods are able to approximate the solvent effects accurately. But they cannot be applied to the protein-folding problem, yet, since they are still computationally too expensive when employed in simulations of large molecules such as proteins. We will therefore formulate and investigate a new method also based on the integral equation theories yet involving an approximation never considered before concerning the application of accurate and effective approximation of solvent effects in solute-solvent simulations – the Kirkwood approximation.

The native configuration of a protein represents a stable state of the system and is therefore a macroscopic property. Such macroscopic properties are assumed to be constituted by all possible microscopic states of the system. In principle, it should be possible to derive all macroscopic properties if sufficient information about the molecular interaction is given. The transfer from a microscopic level to the macroscopic quantity, the observable, is accomplished by a so-called ensemble average, i.e. an average over all possible microscopic states. Unfortunately, these ensemble averages can be computed exactly only for very few systems, as e.g. the ideal gas or the two-dimensional Ising model for ferromagnets. Computer simulations are able to approximate such ensemble averages. The standard tools in computational biology are Monte Carlo and molecular dynamics methods. They sample the phase space, that is the space of all possible microscopic states of a system, by a stochastic or a deterministic procedure, respectively. The ensemble average can then be approximated by a direct sum of the microscopic quantity of interest evaluated at the sampling points in phase space. This way, any desired macroscopic property of the system can be computed, assuming that the exact molecular interaction potential is known. Both methods, Monte Carlo and molecular dynamics, have been applied to a wide variety of molecular systems with great success. Besides the application to molecular systems [72, 106] they are for instance employed in the simulation of liquids [2] or in the field of nano-technology [102].

However, Monte Carlo and molecular dynamics methods do not always yield satisfactory results in acceptable time. This is due to the enormous computational effort necessary to reach convergence. In the protein-folding problem, these complexity problems are caused by two different properties of the system: First, the phase space of a large molecule like a protein is divided into regions that are only connected by unprobable transitions. Each region represents a so-called meta-stable configuration of the molecule. These unprobable transitions between meta-stable states make sampling of the entire phase space very challenging. The second issue is the simulation of the environment of the protein. Proteins always appear in solution which has to be implemented in the computer simulation as well. The solvent is a fluid, e.g. water, which consists of small molecules that have to be included explicitly in a large number in order to form the bulk solvent. This drastically increases the dimension of the phase space. The number of required solvent molecules depends on the size of the protein and the size of the domain but can easily reach tens of thousands of particles. Hence, the computational effort of Monte Carlo and molecular dynamics methods to reach convergence is also increased by several orders and is beyond the capability of today's computers.

A protein in water is a typical representative of what we call a solute-solvent system. In these systems the properties of the solute are at the center of interest, i.e., the ensemble average is to be computed of a microscopic quantity that does not explicitly depend on the solvent degrees of freedom. Nevertheless, the macroscopic property is strongly influenced by the specific solvent due to the microscopic solute-solvent interaction. In the case of a folded protein for example, the stable configuration indeed depends on the type of the solvent. But obviously, the configuration of the solute does not depend explicitly on the solvent degrees of freedom. It is therefore possible to derive the so-called potential of mean force (PMF) which incorporates the effects on the solute due to the solvent implicitly, i.e. without introducing new degrees of freedom to the system. With respect to the macroscopic properties, the systems with explicit and implicit solvent representation are identical. Formally, the PMF is derived by integrating the probability density of the system over all solvent degrees of freedom. This leads to an interaction potential that only depends on the solute degrees of freedom and contains the solvent effects implicitly. Figure 1.1 illustrates the difference between the two approaches. The left plot shows a typical configuration of a biomolecular system: A protein is to be simulated in aqueous solution. To this end, the water molecules are included explicitly in the simulation box. However, no explicit solvent molecules are necessary by the use of the PMF. Instead, a force field that implicitly incorporates the solvent effects on the solute is considered. This is indicated by the red field of Figure 1.1 (right).

The introduction of the PMF formally simplifies the simulation of solute-solvent systems since only the solute degrees of freedom have to be considered. However, the exact computation of the PMF is as challenging as the exact computation of the

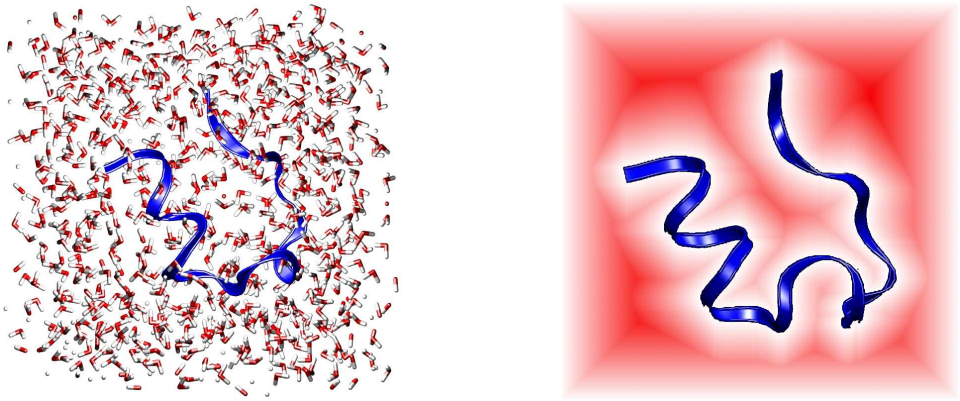

Figure 1.1. *Left: A protein (blue ribbon) surrounded by H_2O molecules (explicit solvent). Right: The same protein with implicit solvent indicated by the red field.*

ensemble average of the entire system. Again, the PMF can be computed exactly only for very few simple systems. Hence, we need to resort to efficient approximation approaches. In principle, Monte Carlo and molecular dynamics methods can be applied in order to approximate the PMF. But this has an essential drawback: Since no parametrization of the PMF is known, it has to be computed anew for any configuration of the solute. Hence, every step of a simulation of the solute-solvent system with an implicit solvent model would comprise the approximation of the PMF by a Monte Carlo or a molecular dynamics simulation. This is also computationally unfeasible. Therefore, other methods have to be considered that approximate the PMF more effectively.

Existing implicit solvent models employed for simulations of biomolecular systems are very simple approximations to the PMF [103, 104]. They are only able to include the most important solvent effects on a qualitative level of detail. The most popular model is the so-called GB/SA continuum model of solvation. It is a combination of a model that approximates the van der Waals interaction between solute and solvent by means of the solvent accessible surface area (SASA) and the generalized Born (GB) model for the approximation of the electrostatic interactions. Both parts can be evaluated with a numerical complexity that is linearly dependent on the number of solute particles. Note however that it is intended for qualitative not quantitative studies and therefore cannot give accurate predictions of solvation free energies or other properties of the solute-solvent interaction.

In recent years, the development of new implicit solvent models that reproduce the solvent effects more accurately and that are still numerically tractable has been advanced by many authors. The most promising approaches are the methods based

on the liquid state integral equation theories. These integral equations emerge from a hierarchy of equations for the reduced probability distributions which can be employed to compute the ensemble averages by lower-dimensional integrals. Likewise, the PMF can be computed by an integral over a reduced probability distribution. To be more precise, if one restricts the solute-solvent interaction to pair-potentials, the required reduced probability distribution is equivalent to the average solvent density around the solute. The assumption of pair-potentials is admissible since all popular force-fields for biomolecular simulation comprise only pair-potentials for the solute-solvent interaction. This provides the best compromise between accuracy and computational effort. The computation of the average solvent density again corresponds to the computation of an ensemble average. But the liquid state integral equation theories can be used as a starting point to develop methods that approximate the solvent density around an arbitrary solute.

The integral equation theories have been developed to compute macroscopic properties of fluids without explicitly performing the integration over the full phase space. They have been applied very successfully to simple pure fluids, to fluids at interfaces as well as to molecular fluids [45, 47]. The theories can be divided into two classes depending on the fundamental relation of classical statistical mechanics they are based on. On the one hand this is the Ornstein-Zernike equation and on the other hand the YBG-hierarchy (Yvon, Born, Green), see [47]. Since both equations are underdetermined they cannot be solved without additional assumptions. These assumptions then yield so-called closure relations, i.e. additional equations. The difference in the approaches based on the Ornstein-Zernike equation and the YBG-hierarchy lies in the employed closure relations. In the literature, the Ornstein-Zernike based methods are widely used for the investigation of atomic and molecular fluids. This may be due to the fact that they can be reduced to one-dimensional equations in the case of spherical symmetry. A wide variety of closure relations has been developed for the Ornstein-Zernike equation and their theoretical background is well understood [45,47]. On the other hand, methods based on the YBG-hierarchy are less popular. They share the drawback that their reduction to one-dimensional equations in the case of spherical symmetry is not trivial. Hence, their approximate solution by computer simulation requires a greater computational effort. Nevertheless, they have been applied to atomic fluids, to fluids at interfaces, and, in recent years, to polymers [32, 116].

Concerning the application to the solute-solvent systems, the literature is largely focused on Ornstein-Zernike based methods. Ikeguchi and Doi [53] and Beglov and Roux [10] have employed the Ornstein-Zernike equation together with the hypernetted-chain (HNC) and Percus-Yevick (PY) closure for the computation of the density of a simple monoatomic solvent around solutes of arbitrary shape. Kovalenko, Hirata et al. [61–67] and Beglov, Roux et al. [12, 28] have extended the methods in order to be able to cope with molecular solvents as well. The so-called

3d-RISM-PLHNC and 3d-RISM-HNC methods have been applied to several solute-solvent systems, as e.g. alkanes, alcohols, carboxylic acids and simple amides in water. In [28] solvation free energies of several solute-solvent systems are computed and the results are in acceptable agreement with experimental data. The errors are assumed to be the result of the approximation comprised in the closure relations. Therefore, the authors propose empirical corrections in order to improve the agreement between theory and experimental data.

To our knowledge, methods based on the YBG-hierarchy have never been considered for the computation of solvent densities in solute-solvent systems. They could enable the incorporation of new approximations that have not been employed for this application, yet. Furthermore, the numerical solution of a method based on the YBG-hierarchy could prove to be more efficient when full three-dimensional resolution of the solvent density is required as it is typically the case in solute-solvent systems. Developments that are related to the application of solute-solvent systems where made in the field of polymeric fluids. To this end, Eu and Gan [32], Taylor and Lipson [116] and Attard [4] have developed methods based on the YBG-hierarchy that have been quite successfully applied to several polymer models [40–43,117–119,121,122]. In these models, a polymer chain consists of either hard or soft spheres with rigid or flexible bonds. But neither chains with different types of particles nor more complex interaction potentials as e.g. the Coulomb potential have been considered.

We are going to present a new approach based on the YBG-hierarchy and investigate in full detail its usefulness for the computation of solvent density distributions around a solute of arbitrary shape. Therefore, we first derive our BGY3d model for monoatomic solvents directly from the YBG-hierarchy. We employ the Kirkwood superposition approximation as our closure relation. We show how the resulting BGY3d equations can be transformed such that an efficient numerical solution in three dimensions by means of Fourier transformations is possible. Application of our model to solvents interacting by the Lennard-Jones potential reveals that the results are of the same quality as those obtained by Beglov and Roux [10] with their Ornstein-Zernike based method. However, the computational effort is much smaller in our approach. When compared to a molecular dynamics simulation, the solution of our BGY3d model even performs about four orders of magnitude faster for the computation of the solvent density. Hence, our model yields a drastic improvement with respect to the computational efficiency.

In order to be able to consider more realistic solvents, we further extend our model to molecular solvents. To this end, the solvent molecules are modeled as rigid bodies. The intramolecular distribution functions are derived by taking the limit of an infinite restoring force between two bonded particles. The resulting molecular BGY3d (BGY3dM) equations can be used to compute the site densities, i.e. the densities of the atoms that constitute the solvent molecules, of a complex molecular

solvent around an arbitrary solute. These equations can efficiently be solved in three-dimensions by means of Fourier transformations as in the case of monoatomic solvents. We are further able to consider solvents with charged sites. The numerical treatment of the Coulomb interaction requires a special splitting of the potential in order to cope with the long-range part. A comparison of the results with those obtained by a molecular dynamics simulation shows a similar agreement as in the case of monoatomic solvents. The gain with respect to the computational effort still is close to three orders of magnitude for the solution of the BGY3dM model compared to a molecular dynamics simulation.

The computed results of our method based on the YBG-hierarchy clearly show that it performs at least as good as the Ornstein-Zernike based methods concerning the evaluation of the PMF of solute-solvent systems. The new BGY3d and BGY3dM models are superior with respect to computational effort and comparable with respect to accuracy. They can deal with the most important interaction potentials in the field of biomolecular simulation, namely with the short-range Lennard-Jones and with the long-range Coulomb potential. However, it is an approximative model. Hence, the accuracy could for example be improved by introducing empirical corrections to the approximations as it also necessary for the Ornstein-Zernike based methods.

The remainder of this monograph is organized as follows: Some basic concepts of classical and statistical mechanics are introduced in Chapter 2. We briefly present the two most popular computational methods for molecular simulation on a microscopic scale, namely the Monte Carlo and the molecular dynamics methods. As a motivation we explain why there is a need for more specialized methods for a very important class of applications, i.e., for so-called solute-solvent systems. We discuss the liquid state integral equation theories for simple pure fluids since they provide the fundamental concepts for the development of accurate implicit solvent models.

In Chapter 3, we present important characteristics of the solute-solvent systems. The potential of mean force (PMF) is formally introduced and some classical methods to approximate it are presented. Further, we show how the PMF can be computed by means of the solvent density.

We give an overview of existing methods for the computation of solvent densities around complex solutes based on the integral equation theories in Chapter 4. Here, we derive our BGY3d and BGY3dM models based on the YBG-hierarchy. Our special product approach for the solvent densities facilitates an efficient numerical treatment of the resulting integro-differential equations.

The numerical details for the solution of our BGY3d and BGY3dM equations are presented in Chapter 5. To this end, we discuss our implemented algorithm and the discretization. We further validate our methods for monoatomic and molecular solvents with respect to convergence and accuracy by comparing the results to molecular dynamics simulations. We assess their efficiency and compare it to

the 3d-HNC method of Beglov and Roux [10]. Here, it turns out that our BGY3d method for monoatomic solvents is superior to the 3d-HNC method with respect to computational effort.

We apply our BGY3dM model for molecular solvents to some realistic solute-solvent systems in Chapter 6. We compute the site densities of carbon disulfide around several solutes with different properties concerning their size and their partial charges. To this end, the BGY3dM model leads to reasonable density and charge distributions of the solvent. Finally, we summarize our findings in Chapter 7 and give an outlook on future developments of the BGY3d and BGY3dM models.

Chapter 2
Theory of Simple Liquids

We will first present the fundamental concepts of classical mechanics and statistical physics on the basis of simple liquid systems. The goal is to understand the connection between the microscopic description of a system and its macroscopic properties. We begin with the microscopic description, i.e. the description of the system on the atomic level which is fully described by the inter-atomic interaction. To this end, we assume that the system is described with appropriate accuracy by the laws of classical mechanics. Moreover, the interaction between the particles is described by an empirical potential function, since the incorporation of quantum mechanical effects is still unfeasible with today's computers in extensive microscopic simulations with thousands of particles.

The transfer from the microscopic description to the macroscopic level is accomplished by the concepts of statistical mechanics. Since the respective relations cannot be evaluated exactly, we will present methods that can approximately compute macroscopic quantities from the microscopic representation of a system. These are the Monte Carlo and molecular dynamics methods on the one hand and the liquid state integral equation theory on the other hand. The former methods are applicable quite generally whereas the latter is a more specialized theory that assumes certain properties of the atomic interaction and the macroscopic property under consideration. We will present the integral equation theories in more detail, since they build the basis for our derivations concerning the solute-solvent systems.

2.1 Classical Mechanics

Molecular systems can be described on the microscopic level by a discrete number N of particles. We assume that the dynamics of such a system is governed by the laws of classical mechanics. To describe the system, each particle has a label i, $i = 1, 2 \ldots, N$, a position $\mathbf{x}_i \in \mathbb{R}^d$ and a momentum $\mathbf{p}_i \in \mathbb{R}^d$. We only consider the case of three dimensions $d = 3$. In order to specify the complete dynamical state of a

system, knowledge of all positions $\mathbf{x}_1, \ldots, \mathbf{x}_N$ and the conjugate momenta $\mathbf{p}_1, \ldots, \mathbf{p}_N$ is necessary. Hence, the system has $6N$ degrees of freedom. It is convenient to define the phase space Γ_N which is the set of all possible states of a system consisting of N particles:

$$\Gamma_N(\Omega) = \left\{(\mathbf{p}, \mathbf{x}) : \mathbf{p} \in \mathbb{R}^{3N}, \mathbf{x} \in \Omega_N\right\} \tag{2.1}$$

with

$$\Omega_N = \underbrace{\Omega \times \cdots \times \Omega}_{N \text{ times}}, \qquad \Omega \subseteq \mathbb{R}^3.$$

We write short $(\mathbf{p}, \mathbf{x})$ for $(\mathbf{p}_1, \ldots, \mathbf{p}_N, \mathbf{x}_1, \ldots, \mathbf{x}_N)$. All dynamical functions emerge from the set of functions on phase space and are written as $a(\mathbf{p}, \mathbf{x})$. The total energy which is the sum of the kinetic energy and the potential energy of the system appears as a special function. The kinetic energy is due to the motion of the particles whereas the potential energy is a result of the interactions between the particles. We assume the absence of any external field. It is further assumed that the system is conservative, i.e., the total energy is conserved. The total energy is represented by the Hamiltonian

$$\begin{aligned} H(\mathbf{p}, \mathbf{x}) &= H_{kin} + H_{pot} \\ &= \frac{1}{2} \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{m_i} + V(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \end{aligned} \tag{2.2}$$

with H_{kin} the kinetic energy and H_{pot} the potential energy. The potential V describes the interaction of the particles.

The dynamics of a system is fully governed by its N position vectors and N momenta which are functions of time: $\mathbf{x}(t)$, $\mathbf{p}(t)$. Their time-dependence is described by Hamilton's equations of motion

$$\begin{aligned} \dot{\mathbf{x}}_i(t) &= \frac{\partial H(\mathbf{p}, \mathbf{x})}{\partial \mathbf{p}_i}, \quad i = 1, \ldots, N, \\ \dot{\mathbf{p}}_i(t) &= -\frac{\partial H(\mathbf{p}, \mathbf{x})}{\partial \mathbf{x}_i}, \quad i = 1, \ldots, N. \end{aligned} \tag{2.3}$$

The dot $\dot{}$ is short notation for the time derivative $\frac{\partial \mathbf{x}}{\partial t}$. A system with particle coordinates $\mathbf{x}^0$ and momenta $\mathbf{p}^0$ at time $t = 0$ evolves according to (2.3) and we write $\mathbf{x}(t) = \mathbf{x}(t; \mathbf{p}^0, \mathbf{x}^0)$ and $\mathbf{p}(t) = \mathbf{p}(t; \mathbf{p}^0, \mathbf{x}^0)$. The function $\mathbf{x}(t; \mathbf{p}^0, \mathbf{x}^0)$ for an interval $t \in [0, t_{end}]$ is called a trajectory. The set of points $(\mathbf{p}(t), \mathbf{x}(t))$ for $-\infty \leq t \leq \infty$ defines the orbit of the system in phase space, i.e. the set of all points the system can reach.

The equations of motion (2.3) describe the time evolution of the microscopic system. However, a specific representation of a trajectory does not have any significance

concerning the macroscopic properties of the system. The view of a microscopic system that is located at a specific point in phase space at a specific time has to be replaced by the concept that the macroscopic system is represented by all possible microscopic states. Hence, it has to be clarified how the microscopic description is exactly connected to the macroscopic properties of a system.

2.2 Statistical Mechanics

The dynamics of a microscopic system is fully governed by the Hamiltonian mechanics described above. Corresponding microscopic dynamical quantities are functions of the phase space variables $(\mathbf{p}, \mathbf{x})$ only. These may depend also on the parameters $\mathbf{r} \in \mathbb{R}^3$ and t, the position in physical space and the time, respectively. Properties of the system on the macroscopic level are described by fields, i.e. functions $A(\mathbf{r}, t)$ which depend on $\mathbf{r}$ and t only. The purpose of statistical mechanics is to bridge the scales by identifying for every microscopic quantity its unique macroscopic correspondence

$$a(\mathbf{p}, \mathbf{x}) \overset{\text{Stat. Mech.}}{\longrightarrow} A(\mathbf{r}, t).$$

In order to create such a correspondence we need a mapping from phase space to physical space. This mapping associates for a given point $(\mathbf{r}, t)$ in (physical) space and time and with each function $a(\mathbf{p}, \mathbf{x})$ on phase space a scalar. We write

$$A(\mathbf{r}, t) = \langle a(\mathbf{p}, \mathbf{x}) \rangle = \langle a \rangle. \tag{2.4}$$

The mapping is linear and maps scalars in phase space to scalars in physical space. It is realized by

$$\langle a \rangle = \int_{\Gamma_N} a(\mathbf{p}, \mathbf{x}) \pi(\mathbf{p}, \mathbf{x}) \; d\mathbf{p} d\mathbf{x} \tag{2.5}$$

with $\pi(\mathbf{p}, \mathbf{x})$ a function on phase space which is positive definite,

$$\pi(\mathbf{p}, \mathbf{x}) \geq 0, \qquad \forall \; \mathbf{p}, \mathbf{x}, \tag{2.6}$$

and satisfies the normalization condition

$$\int_{\Gamma_N} \pi(\mathbf{p}, \mathbf{x}) \; d\mathbf{p} d\mathbf{x} = 1. \tag{2.7}$$

Functions π, which satisfy conditions (2.6) and (2.7), are called distribution functions. The basic postulate of classical statistical mechanics is that the state of a system is completely determined by the specification of the distribution function. An observable A, i.e. a macroscopic quantity, is associated with a microscopic function a by relation (2.5), see e.g. [8] for a more detailed description of the postulate.

The properties of the distribution function π lead to the interpretation as phase space probability density. To be more precise, $\pi \; d\mathbf{p} d\mathbf{x}$ is the probability to find the

system within the infinitesimal domain $\mathbf{p} + d\mathbf{p}$, $\mathbf{x} + d\mathbf{x}$ in phase space. The integral in (2.5) can therefore be understood as weighted average of the microscopic function a and is sometimes called the phase space average of a. In classical mechanics the state of a system is represented by a single point in phase space. This view is replaced by the knowledge of the distribution function in statistical mechanics, where all points in phase space are considered weighted with their probability. This totality of possible single point states endowed with the probability density is called the statistical ensemble.

The specific realization of the probability density specifies some important properties of the macroscopic system. It defines e.g. whether a system has constant energy, temperature or mass. The respective microscopic systems then differ in their ability to exchange energy or particles with their environment. Examples are the micro canonical, the canonical and the grand canonical ensemble. In the micro canonical ensemble the system is completely isolated and cannot exchange energy or mass with its environment. In the canonical ensemble the system is coupled to a so-called heat reservoir such that it has a constant temperature. If the system can also exchange mass with its environment, this is called the grand canonical ensemble. We will now discuss the properties of the canonical ensemble in more detail since it is the ensemble which we will employ in the following.

The Canonical Ensemble

As already noted above, a system in the canonical ensemble can exchange energy with the environment which is called the heat reservoir. Since the heat reservoir is assumed to be very large, this leads to a constant temperature of the system under consideration. The volume $|\Omega|$ and the number of particles in the system N are also constant. We now define the set Γ_c of all possible states $(\mathbf{p}, \mathbf{x})$ the system in the canonical ensemble can assume

$$\Gamma_c = \Gamma_N(\Omega). \tag{2.8}$$

It equals the entire phase space (2.1). However, the probability of any state in Γ_c depends on its energy. The probability distribution is given by

$$\pi_c(\mathbf{p}, \mathbf{x}) = \frac{Z^{-1}(N, \Omega, T)}{N! h^{3N}} e^{-\beta H(\mathbf{p}, \mathbf{x})} \tag{2.9}$$

with $\beta = \frac{1}{k_B T}$ and k_B the Boltzmann constant, Planck's constant h and the factor $N!$, which accounts for the fact that the particles are indistinguishable. The function $Z(N, \Omega, T)$ is called the partition function of the canonical ensemble and is given by

$$Z(N, \Omega, T) = \frac{1}{N! h^{3N}} \int_{\Gamma_c} e^{-\beta H(\mathbf{p}, \mathbf{x})} \, d\mathbf{p} d\mathbf{x}. \tag{2.10}$$

The integration over momenta can be carried out exactly, which leads to

$$Z(N,\Omega,T) = \frac{1}{N!h^{3N}} \int_{\Omega_N} \int_{\mathbb{R}^{3N}} e^{-\beta H(\mathbf{p},\mathbf{x})}\, d\mathbf{p}d\mathbf{x} \tag{2.11}$$

$$= \frac{1}{N!h^{3N}} \int_{\Omega_N} \int_{\mathbb{R}^{3N}} e^{-\beta\left(\frac{1}{2}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i}+V(\mathbf{x})\right)}\, d\mathbf{p}d\mathbf{x} \tag{2.12}$$

$$= \frac{1}{N!h^{3N}} \int_{\mathbb{R}^{3N}} e^{-\beta\left(\frac{1}{2}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i}\right)}\, d\mathbf{p} \int_{\Omega_N} e^{-\beta V(\mathbf{x})}\, d\mathbf{x} \tag{2.13}$$

$$= \frac{1}{N!h^{3N}} \left(\int_{-\infty}^{+\infty} e^{-\beta\frac{p^2}{2m}} dp\right)^{3N} \int_{\Omega_N} e^{-\beta V(\mathbf{x})}\, d\mathbf{x} \tag{2.14}$$

$$= \frac{\lambda_c^N}{N!} \int_{\Omega_N} e^{-\beta V(\mathbf{x})}\, d\mathbf{x} \tag{2.15}$$

with $\lambda_c = \left(\frac{2m\pi}{\beta h^2}\right)^{3/2}$. The configurational part of the partition function can be computed exactly only for very few simple systems as e.g. the ideal gas. In general, it has to be approximated. Other thermodynamic properties of the canonical ensemble are related to the partition function by the relation

$$F(N,\Omega,T) = -\frac{1}{\beta}\ln Z(N,\Omega,T) \tag{2.16}$$

with $F(N,\Omega,T)$ the free energy of the system.

The Thermodynamic Limit

The probability distribution of the canonical ensemble (2.9) can now be used to compute the ensemble average (2.5). By this, we can compute macroscopic properties from the microscopic description of the system. However, these averages formally depend on the domain and the number of particles of the system. Observables, however, should be independent of these quantities. The fact that for very small systems, the ensemble averages lead to macroscopic properties that depend on the size of the system, is called the finite size effect. If we increase the system size, the ensemble average converges in the limit of infinite volume and particle number, i.e.

$$A(\mathbf{r},t;\rho,T) = \lim_{\substack{|\Omega|\to\infty, N\to\infty \\ \frac{N}{|\Omega|}=\text{const}}} \int_{\Gamma_N} a(\mathbf{p},\mathbf{x})\pi(\mathbf{p},\mathbf{x})\, d\mathbf{p}d\mathbf{x}. \tag{2.17}$$

The number density $\rho = \frac{N}{|\Omega|}$ has to be constant in this limit process. The resulting observable A is a function of this density and the temperature. The transition to the infinite system size is called the thermodynamic limit. In this limit all ensembles are equivalent.

Practically, it is sufficient to consider systems that are appropriately big instead of taking the limit to infinity. This works very well for many applications of interest. The above considerations lead however to an important aspect when we try to approximate ensemble averages by computer simulations. When we employ small microscopic systems, we always have to struggle with the the finite size effect, i.e., the ensemble averages still depend on the system size. But the system size can only be increased to a certain extend due to the limited computational power of today's computers. Under certain conditions, other approaches can reduce the finite size effect. The introduction of periodic boundary conditions implements for example a periodic system of infinite size. Yet, the finite size effect is one of the main reasons why computer simulations on a microscopic scale, even for homogeneous systems, are not (yet) able to reproduce all effects which are observed macroscopically.

Time Evolution of the Phase Space Probability Distribution

The probability distribution π completely describes the state of a system. So far, we only considered systems at equilibrium, i.e., the distribution function is a constant function over time. As an example for a system not at equilibrium, we consider a gas confined in a box that is divided by a wall into two parts. At times $t < t_0$ the gas is located only in the left half of the box. At $t = t_0$ the wall is removed and the gas can expand into the right half. Surely, after a very short time $t = t_0 + \Delta t$ the probability to find any gas at the right wall of the box still is zero. Obviously, the system is not at equilibrium. But how does the function $\pi(\mathbf{p}, \mathbf{x}, t)$ evolve for $t > t_0$? This evolution is described by the Liouville equation

$$\frac{\partial \pi}{\partial t} = \{H, \pi\} \tag{2.18}$$

with the Hamiltonian H (2.2) and the Poisson-Brackets $\{.\}$ defined by

$$\{A, B\} = \sum_{i=1}^{N} \left(\frac{\partial A}{\partial \mathbf{x}_i} \cdot \frac{\partial B}{\partial \mathbf{p}_i} - \frac{\partial A}{\partial \mathbf{p}_i} \cdot \frac{\partial B}{\partial \mathbf{x}_i} \right). \tag{2.19}$$

Here, the dot $\cdot$ denotes the scalar product of two vectors in $\mathbb{R}^3$. The Liouville equation can be seen as the $6N$ dimensional analogue of the continuity equation of a classical fluid. Probability can neither be destroyed nor created as time evolves, i.e., the normalization condition (2.7) holds for any t.

If we use Hamilton's equations of motion (2.3), we can write (2.18) as

$$\frac{\partial \pi}{\partial t} = -\sum_{i=1}^{N} \left(\frac{\partial \pi}{\partial \mathbf{x}_i} \cdot \dot{\mathbf{x}}_i + \frac{\partial \pi}{\partial \mathbf{p}_i} \cdot \dot{\mathbf{p}}_i \right). \tag{2.20}$$

The time dependence of any observable A can be described in a similar manner as

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \sum_{i=1}^{N} \left(\frac{\partial A}{\partial \mathbf{x}_i} \cdot \dot{\mathbf{x}}_i + \frac{\partial A}{\partial \mathbf{p}_i} \cdot \dot{\mathbf{p}}_i \right). \tag{2.21}$$

Exact solutions of equation (2.18) can only be found for very simple examples. For realistic systems it is not immediately useful due to its high complexity. Yet, it can be formally simplified by integration over $N - n$ particle degrees of freedom. This results in a hierarchy of equations which can be used as the starting point of the liquid state integral equation theories. They will be presented in Section 2.4.2.

2.3 Molecular Simulation

As we have learned in the preceding Section, statistical mechanics teaches us how to compute observables, which are macroscopic functions of space and time, from microscopic quantities, which are functions of the momenta and the positions of the particles. In principle, it is then possible to compute macroscopic properties of a system if the Hamiltonian, i.e. the interaction potential V in (2.2), is known. By this, computer simulations can replace real world experiments to a certain extend, assuming that enough information about the microscopic composition of the material under consideration is available.

We will consider some practical aspects of computing the partition function or ensemble averages (2.5). In general, this is a very challenging task. Except for some simple choices of potential functions in the Hamiltonian (2.2), it is not possible to compute the integral exactly. Hence, computational methods are used to approximate (2.5). However, standard integration techniques cannot be adapted easily to this problem because of the high dimension of the phase space Γ_N. Numerical quadrature methods run into complexity problems already for small particle numbers. This is also known as the curse of dimensionality, which simply describes the fact, that the cost of computing integrals like (2.5) depends exponentially on the number of particles, i.e. the dimension of the problem [13]. A discretization of the phase space Γ_N by a full grid would require $\mathcal{O}(n^{6N})$ points if n is the number of grid points in one dimension and N the total number of particles. This is unfeasible except for very small numbers of particles.

To overcome the curse of dimensionality one can further investigate the structure of the phase space. Depending on the probability measure π_c of the canonical ensemble, not all possible configurations $\mathbf{q} = (\mathbf{p}, \mathbf{x})$ are equally important for the integral (2.5). Hence, most computational methods in molecular simulation sample the phase space taking into account the known (relative) probability of the points. This way, the use of a full grid is avoided and the phase space is explored according to the probability distribution. These methods lead to the exact solution for an infinite number of sampling points. For a finite number of sampling points, statements about

the error are statistically in nature, i.e., if important parts of the phase space where not sufficiently sampled, the true error can be much greater than its statistical estimate. Nevertheless, these methods often yield reasonable results for numbers of computational steps that are feasible with today's computers.

In the next sections, we will shortly present the ideas of Monte Carlo and molecular dynamics methods. These methods belong to the standard tools in the field of biomolecular simulation and can generally be applied in order to compute ensemble averages like

$$\langle a \rangle = \int_{\Gamma_N} a(\mathbf{p}, \mathbf{x}) \pi(\mathbf{p}, \mathbf{x}) \; d\mathbf{p} d\mathbf{x}.$$

2.3.1 Monte Carlo

The Monte Carlo method is a stochastic method. A sequence of points is to be constructed that obeys the given probability distribution π, which depends on the application. For this, any point in the integration domain is computed by a stochastic procedure. Since the Monte Carlo concept can generally be adopted to any integration problem, a lot of algorithms exist that implement the generation of point sequences differently. We want to simulate the canonical ensemble of a molecular system. To this end, the most popular method is the Metropolis algorithm [79].

We assume that the microscopic quantity a does not depend on the momenta

$$a(\mathbf{p}, \mathbf{x}) = a(\mathbf{x}).$$

Hence, the integration over the momenta can be carried out exactly, see Section 2.2, and we only have to sample the configurational part of the phase space. This is done in the Metropolis algorithm by adding a small random displacement $\Delta\mathbf{x}_k$ to the old configuration of the system $\mathbf{x}_k$ in each iteration step k. The new configuration is accepted or rejected with probability

$$\min\left(1, \frac{e^{-\beta V(\mathbf{x}_k + \Delta\mathbf{x}_k)}}{e^{-\beta V(\mathbf{x}_k)}}\right), \tag{2.22}$$

where V is the potential function. If the new step is accepted, we set $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}_k$, otherwise $\mathbf{x}_{k+1} = \mathbf{x}_k$. After K computational steps, a sequence of points $(\mathbf{x}_k)$ is generated that obeys the probability distribution of the configurational part of the canonical ensemble. The Monte Carlo approximate of the observable can be computed from this sequence of points by

$$\langle a \rangle \approx \frac{1}{K} \sum_{i=1}^{K} a(\mathbf{x}_k). \tag{2.23}$$

In other words, the Metropolis algorithm samples the phase space according to the probability distribution of the canonical ensemble. Then, a simple average of the

microscopic quantity evaluated at the sampling points gives an approximation of the observable, i.e. the integral (2.5). In the limit of $K \to \infty$ steps the sum (2.23) should equal the ensemble average (2.5). For this, the ergodicity of the Monte Carlo algorithm is required, i.e., any possible $\mathbf{x}$ can be assumed by the algorithm from any initial configuration $\mathbf{x}^0$ in a finite number of steps. The Metropolis algorithm itself is not ergodic, since it cannot overcome an infinite energy barrier. A combination of the Metropolis algorithm with another Monte Carlo scheme can however produce an ergodic algorithm. One could e.g. start the Metropolis algorithm from several randomized initial configurations $\mathbf{x}^0$ that are sampled by an ergodic algorithm.

The advantage of the Metropolis algorithm, as opposed to quadrature methods that would uniformly sample the phase space, is that the knowledge of the relative probability of two configurations is sufficient for the approximation of the integral. Otherwise, it would be necessary to compute also the partition function of the canonical ensemble, i.e. the normalization of the probability distribution. Similar to the Monte Carlo method a sequence of points according to the probability distribution is also constructed during a molecular dynamics simulation.

2.3.2 Molecular Dynamics

Molecular dynamics deals with the solution of Newton's equations of Motion

$$\dot{\mathbf{x}}_i = \mathbf{v}_i, \qquad \dot{\mathbf{p}}_i = -\nabla_{\mathbf{x}_i} V(\mathbf{x}), \qquad i = 1, 2, \cdots, N \tag{2.24}$$

with $\mathbf{x}_i$ the positions, $\mathbf{p}_i$ the momenta and $\mathbf{v}_i = \frac{\mathbf{p}_i}{m_i}$ the velocity of the ith particle. The dot in $\dot{\mathbf{x}}_i$ denotes the partial time derivative. These equations can be derived from the classical Hamiltonian H (2.2), see Section 2.1. The equations describe the time evolution of a molecular system, i.e., the solution is a trajectory $\mathbf{x}(t) = \mathbf{x}(t; \mathbf{x}^0, \mathbf{p}^0)$ with initial configuration $\mathbf{x}^0$ and momenta $\mathbf{p}^0$.

The connection to our ensemble average (2.5) is made by the ergodic hypothesis. The hypothesis basically states that the ensemble average can be replaced by a time average as

$$\int_{\Gamma_N} a(\mathbf{p}, \mathbf{x}) \pi(\mathbf{p}, \mathbf{x}) \; d\mathbf{p} d\mathbf{x} = \lim_{t_{end} \to \infty} \frac{1}{t_{end}} \int_0^{t_{end}} a(\mathbf{p}(t), \mathbf{x}(t)) \; dt \tag{2.25}$$

with $\mathbf{x}(t)$ the trajectory and $\mathbf{p}(t)$ the time evolution of the momenta of a molecular dynamics simulation. By this, it is assumed that it is just as good to observe a system over a long time as it is to consider many independent realizations of the system. We can easily to imagine a system, where the ergodic hypothesis is erroneous. If for example two configurations are separated by a very high (infinite) energy barrier, it is impossible for the system to reach one configuration from the other. Hence, time average and ensemble average differ in this case. Nevertheless,

in many practical applications of classical mechanics the hypothesis is assumed to make sense, see e.g. [46]. As in the case of the Metropolis algorithm one could also combine molecular dynamics with a simple Monte Carlo scheme in order to produce an ergodic molecular dynamics/Monte Carlo hybrid scheme.

An additional problem arises if we compute the trajectory approximately by a computer simulation. Even if we assume that (2.25) holds, it is not immediately clear whether it also applies for the discrete trajectory, i.e.

$$\int_{\Gamma_N} a(\mathbf{p},\mathbf{x})\pi(\mathbf{p},\mathbf{x})\; d\mathbf{p}d\mathbf{x} = \lim_{N_t\to\infty} \frac{1}{N_t}\sum_{i=1}^{N_t} a(\mathbf{p}(t_i),\mathbf{x}(t_i)) \tag{2.26}$$

with N_t the number of discrete time steps. Mathematically, Newton's equations of motion constitute a system of ODEs that can be solved by a time integration scheme. It can be shown that in order for (2.26) to be true, the integration scheme has to conserve the volume in phase space over time, i.e., the volume of any set of points in phase space that is moved according to the equations of motions is conserved. Such integration schemes are called symplectic, see [46] for details. As a consequence, symplectic integrators have the property to approximate well the ensemble averages (2.5) instead of the trajectory $\mathbf{x}(t)$ itself.

The Hamiltonian in (2.2) describes a micro canonical system. The simulation of the canonical ensemble is realized by so-called thermostats. To this end, the coupling to the heat reservoir can be realized by introducing a frictional term into Newton's equations of motion which then read as

$$\dot{\mathbf{x}}_i = \mathbf{v}_i, \qquad \dot{\mathbf{p}}_i = -\nabla_{\mathbf{x}_i} V(\mathbf{x}) - \xi(t) m_i \mathbf{v}_i, \qquad i = 1, 2, \cdots, N. \tag{2.27}$$

Depending on the sign of $\xi(t)$ the system gains or looses energy. The function $\xi(t)$ could be computed such that the kinetic energy and the temperature would be constant. Instead, it is more convenient to use the Nosé-Hoover thermostat, see [46]. Here, the heat reservoir is simulated as an additional degree of freedom which determines the strength of the coupling. This way, the temperature is allowed to fluctuate around its desired value. The size of the fluctuation is determined by the coupling parameter.

A molecular dynamics simulation with the Nosé-Hoover thermostat can be used to compute ensemble averages in the canonical ensemble by (2.25) and (2.26). If we again assume that the microscopic quantity does not depend on the momenta, the molecular dynamics approximate of the observable is

$$\langle a \rangle \approx \frac{1}{N_t}\sum_{i=1}^{N_t} a(\mathbf{x}(t_i)). \tag{2.28}$$

It is quite striking that the Monte Carlo and the molecular dynamics methods are very similar in the way the ensemble averages are computed. In both cases, a

sequence of points in configurational space is generated that obeys the probability distribution of the canonical ensemble. Then, the approximation of the phase space integral (2.5) can be computed by a simple average over the microscopic quantity evaluated at the points, compare equations (2.23) and (2.28). The only difference is the generation of the sequence of points, which is a stochastic procedure in the case of Monte Carlo. To this end, we have to ensure that the algorithm is ergodic. That can be guaranteed at least for a combination of the Metropolis algorithm with an ergodic Monte Carlo scheme. The procedure of generating the sequence of points is deterministic in the case of molecular dynamics. Here, ergodicity can also be guaranteed only for a combination of molecular dynamics and Monte Carlo. But it is plausible for many systems of classical mechanics. In summary, both methods are applicable and popular in the application field which is relevant for this thesis.

2.3.3 Limitations of Microscopic Simulation

As already discussed above, ergodicity is required to guarantee that the sums in (2.23) and (2.28) converge to the ensemble average. Numerically, the methods are limited by the number of steps K_{max} that are feasible with today's computers. Generally, one can assume the methods to yield reasonable results if the computed average does not change with more than a given error ϵ for $K > K_{max}$. For a large number of applications this is the case. On the other hand, there also exist a lot of applications which take the computational methods to their limits. Large biomolecules as e.g. proteins are a good example. These molecules can exhibit several distinct meta-stable configurations, i.e., a transition between these meta-stable states is very unlikely. Since the probability for a transition is very small, the Monte Carlo as well as the molecular dynamics methods require a tremendous amount of computational steps in order to sufficiently sample the phase space. In this case, convergence of the averages (2.23) and (2.28) is very slow.

Another negative example are system, where the microscopic quantity of interest depends explicitly only on a small fraction of all degrees of freedom of the entire system. We call such a system a solute-solvent system, where only the solute degrees of freedom are of interest. A protein in water is a typical representation of a solute-solvent system. Properties of the solute, as e.g. the native configuration of the protein, are to be computed. But the solute is influenced by the solvent and all microscopic quantities depend at least implicitly also on the solvent. The number of solvent degrees of freedom, however, can considerably exceed the number of solute degrees of freedom. Hence, the dimension of phase space becomes very large and sampling very slow. In spite of these problems, Monte Carlo and molecular dynamics are often applied to solute-solvent systems, see e.g. [72].

Nevertheless, it is of great interest to develop methods that are able to compute ensemble averages without sampling the entire phase space explicitly. The liquid

state integral equation theories provide such methods for simple liquids.

2.4 Liquid State Integral Equation Theory

The probability density $\pi(\mathbf{p}, \mathbf{x})$ describes the probability of finding the system in a small volume in phase space around $(\mathbf{p}, \mathbf{x})$. Observables are linked to microscopic quantities by a weighted average (2.5) with π as weighting function. Hence, we are now able, in principle, to compute any desired macroscopic property of the system by performing the ensemble average. But the integration domain, i.e. the phase space, is very high dimensional. Hence, Monte Carlo and molecular dynamics methods do not converge in acceptable time for many applications of interest. The situation could be improved by reducing the dimension of the integration domain. Thus, we investigate how the ensemble average can be transformed into an integral over a domain with lower dimension.

2.4.1 Reduced Distribution Functions

We consider a microscopic quantity depending only on $n < N$ particles, i.e.

$$a = a(\mathbf{q}_1, \ldots, \mathbf{q}_n) = a(\mathbf{q}_{(n)}),$$

where we write short $\mathbf{q} = (\mathbf{p}, \mathbf{x})$ and use

$$\mathbf{q}_{(n)} = \mathbf{q}_1, \ldots, \mathbf{q}_n \quad \text{and} \quad \mathbf{q}_{(N-n)} = \mathbf{q}_{n+1}, \ldots, \mathbf{q}_N.$$

Of course, the choice of the first n particles is arbitrary. Since the particles are indistinguishable, there exist $\frac{N!}{(N-n)!}$ different choices of n particles. Thus, we can write the ensemble average (2.5) as

$$\begin{aligned} \langle a \rangle &= \frac{N!}{(N-n)!} \int_{\Gamma_N} a(\mathbf{q}_{(n)}) \pi(\mathbf{q}_{(N)}) \, d\mathbf{q}_{(N)} \\ &= \int_{\Gamma_n} a(\mathbf{q}_{(n)}) \pi^{(n)}(\mathbf{q}_{(n)}) \, d\mathbf{q}_{(n)} \end{aligned} \tag{2.29}$$

with

$$\pi^{(n)}(\mathbf{q}_{(n)}) = \frac{N!}{(N-n)!} \int_{\Gamma_{N-n}} \pi(\mathbf{q}_{(N)}) \, d\mathbf{q}_{(N-n)}. \tag{2.30}$$

The function $\pi^{(n)}$ is called reduced probability density and gives the probability of finding n of the N particles in a small volume around $\mathbf{q}_{(n)}$. For two positive numbers $n < m \leq N$, it holds

$$\pi^{(n)}(\mathbf{q}_{(n)}) = \frac{(N-m)!}{(N-n)!} \int_{\Gamma_{m-n}} \pi^{(m)}(\mathbf{q}_{(m)}) \, d\mathbf{q}_{(m-n)}. \tag{2.31}$$

In the canonical ensemble (2.9) the (reduced) probability density can be further simplified. It can be factorized into a function of the momenta and a function of the coordinates

$$\pi^{(n)}(\mathbf{p}_{(n)}, \mathbf{x}_{(n)}) = \mathcal{P}^{(n)}(\mathbf{p}_{(n)})\rho^{(n)}(\mathbf{x}_{(n)}) \tag{2.32}$$

with

$$\mathcal{P}^{(n)}(\mathbf{p}_{(n)}) = \prod_{i=1}^{n} \left(\frac{\beta}{2\pi m_i}\right)^{\frac{3}{2}} e^{-\beta\frac{|\mathbf{p}_i|^2}{2m_i}}, \tag{2.33}$$

and we can write

$$\rho^{(n)}(\mathbf{x}_{(n)}) = \frac{N!}{(N-n)!} Z_\Omega^{-1} \int_{\Omega_{N-n}} e^{-\beta V(\mathbf{x}_{(N)})}\, d\mathbf{x}_{(N-n)} \tag{2.34}$$

with

$$Z_\Omega = \int_{\Omega_N} e^{-\beta V(\mathbf{x}_{(N)})}\, d\mathbf{x}_{(N)}. \tag{2.35}$$

Here, $\Omega_{N-n} \subset \mathbb{R}^{3(N-n)}$ and $\Omega_N \subset \mathbb{R}^{3N}$ represent the configurational parts of the phase spaces Γ_{N-n} and Γ_N, respectively. We further split the n-particle density $\rho^{(n)}$ into

$$\rho^{(n)}(\mathbf{x}_{(n)}) = \rho^n g^{(n)}(\mathbf{x}_{(n)}) \tag{2.36}$$

with the average density ρ and the n-particle distribution function $g^{(n)}$. Two successive distribution functions are related through

$$g^n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{\rho}{(N-n)} \int_\Omega g^{(n+1)}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+1})\, d\mathbf{x}_{n+1}. \tag{2.37}$$

By the formal introduction of the reduced probability density the ensemble average can be performed by an integration with reduced dimension $3n$ assuming that the microscopic quantity depends only on n particles. For very small numbers n this leads to an integral which can be approximately solved by numerical quadrature methods. But we shifted the problem of high-dimensional integration to the computation of the reduced distribution functions, which are also defined by a high-dimensional integral. Their exact computation is again restricted to very simple systems, as e.g. the ideal gas. In more realistic cases they have to be approximated as well. This can be done by means of the so-called YBG-hierarchy.

2.4.2 The YBG-Hierarchy

In order to understand the concepts of the liquid state integral equation theories it is most suitable to start the considerations with the Liouville equation. For this, we recall that the time evolution of the phase space probability distribution

$\pi(\mathbf{p}, \mathbf{x})$ obeys the Liouville equation (2.18). We restrict the potential function to be composed by a sum of pairwise terms

$$V(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} v(\mathbf{x}_i, \mathbf{x}_j) \tag{2.38}$$

and note that the forces F_{ij} between particle i and j are defined as

$$\mathbf{F}_{ij} = \mathbf{F}(\mathbf{x}_i, \mathbf{x}_j) = -\nabla_{\mathbf{x}_i} v(\mathbf{x}_i, \mathbf{x}_j). \tag{2.39}$$

Then, we can use the definition of the Hamiltonian (2.2) and write the Liouville equation as

$$\frac{\partial \pi}{\partial t} = -\sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial \pi}{\partial \mathbf{x}_i} - \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \mathbf{F}_{ij} \cdot \frac{\partial \pi}{\partial \mathbf{p}_i} - \sum_{i=1}^{N} \mathbf{F}_i^{ext} \cdot \frac{\partial \pi}{\partial \mathbf{p}_i}, \tag{2.40}$$

where $\mathbf{F}_i^{ext}$ are external forces acting on particle i. We are going to transform equation (2.40) so that it becomes a relation for the reduced distribution functions. For this, it is integrated over $N - n$ positions and momenta and multiplied by the factor $\frac{N!}{(N-n)!}$. We can use the definition of the reduced probability density $\pi^{(n)}$ (2.30) and the fact that π is symmetric under exchange of particles. Then we find that

$$\begin{aligned}
&\frac{\partial \pi^{(n)}}{\partial t} + \sum_{i=1}^{n} \frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial \pi^{(n)}}{\partial \mathbf{x}_i} + \sum_{i=1}^{N} \mathbf{F}_i^{ext} \cdot \frac{\partial \pi^{(n)}}{\partial \mathbf{p}_i} \\
&= -\frac{N!}{(N-n)!} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \int_{\Gamma_{(N-n)}(V)} \mathbf{F}_{ij} \cdot \frac{\partial \pi}{\partial \mathbf{p}_i} \, d\mathbf{x}_{(N-1)} d\mathbf{p}_{(N-n)} \\
&= -\sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbf{F}_{ij} \cdot \frac{\partial \pi^{(n)}}{\partial \mathbf{p}_i} - \frac{N!}{(N-n)!} \sum_{i=1}^{n} \sum_{j=n+1}^{N} \int_{\Gamma_{(N-n)}(V)} \mathbf{F}_{ij} \cdot \frac{\partial \pi}{\partial \mathbf{p}_i} \, d\mathbf{x}_{(N-n)} d\mathbf{p}_{(N-n)} \\
&= -\sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbf{F}_{ij} \cdot \frac{\partial \pi^{(n)}}{\partial \mathbf{p}_i} - \sum_{i=1}^{n} \int_{\Gamma_1(V)} \mathbf{F}_{in+1} \cdot \frac{\partial \pi^{(n+1)}}{\partial \mathbf{p}_i} \, d\mathbf{x}_{n+1} d\mathbf{p}_{n+1},
\end{aligned} \tag{2.41}$$

where we wrote $d\mathbf{p}_{(N-n)}$ for $d\mathbf{p}_{n+1} \cdots d\mathbf{p}_N$ and $d\mathbf{x}_{(N-n)}$ for $d\mathbf{x}_{n+1} \cdots d\mathbf{x}_N$. This equation links the reduced probability density $\pi^{(n)}$ to the reduced probability density $\pi^{(n+1)}$,

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} + \sum_{i=1}^{n} \left[\frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial}{\partial \mathbf{x}_i} + \left\{ \mathbf{F}_i^{ext} + \sum_{j=1, j\neq i}^{n} \mathbf{F}_{ij} \right\} \cdot \frac{\partial}{\partial \mathbf{p}_i} \right] \right) \pi^{(n)} \\
&\qquad = -\sum_{i=1}^{n} \int_{\Gamma_1(V)} \mathbf{F}_{in+1} \cdot \frac{\partial \pi^{(n+1)}}{\partial \mathbf{p}_i} \, d\mathbf{x}_{n+1} d\mathbf{p}_{n+1}.
\end{aligned} \tag{2.42}$$

The set of equations for $n = 1, \dots, N-1$ is called the BBGKY-hierarchy after Bogolyubov, Born, Green, Kirkwood and Yvon.

For the Hamiltonian (2.2) in the canonical ensemble the (reduced) probability densities can be factorized as

$$\pi^{(n)}(\mathbf{p}_{(n)}, \mathbf{x}_{(n)}) = \mathcal{P}^{(n)}(\mathbf{p}_{(n)})\rho^{(n)}(\mathbf{x}_{(n)}) \tag{2.43}$$

with

$$\mathcal{P}^{(n)}(\mathbf{p}_{(n)}) = \prod_{i=1}^{n} \left(\frac{\beta}{2\pi m_i}\right)^{\frac{d}{2}} e^{-\beta\frac{|\mathbf{p}_i|^2}{2m_i}} \tag{2.44}$$

only depending on the momenta. Inserting into equation (2.42) at equilibrium, i.e. $\frac{\partial}{\partial t}\pi^{(n)} = 0$, and noting that

$$\frac{\partial}{\partial \mathbf{p}_i}\mathcal{P}^{(n)}(\mathbf{p}_{(n)}) = -\frac{\beta}{m_i}\mathbf{p}_i\mathcal{P}^{(n)}(\mathbf{p}_{(n)}) \tag{2.45}$$

and

$$\int_{\mathbb{R}^3} \mathcal{P}^{(n+1)}(\mathbf{p}_{(n+1)})\ d\mathbf{p}_{n+1} = \mathcal{P}^{(n)}(\mathbf{p}_{(n)}), \tag{2.46}$$

this gives

$$\begin{aligned}\sum_{i=1}^{n} \mathbf{p}_i \cdot \left(\frac{\partial}{\partial \mathbf{x}_i} - \beta\left[\mathbf{F}_i^{ext} + \sum_{j=1, j\neq i}^{n} \mathbf{F}_{ij}\right]\right)\rho^{(n)}(\mathbf{x}_{(n)}) \\ = \beta\sum_{i=1}^{n} \mathbf{p}_i \cdot \int_{\Omega} \mathbf{F}_{in+1}\rho^{(n+1)}(\mathbf{x}_{(n+1)})\ d\mathbf{x}_{n+1}.\end{aligned} \tag{2.47}$$

Here, $\Omega \subseteq \mathbb{R}^3$ is the spatial domain of the system. The relation must be independent of the choice of the momenta $\mathbf{p}_i$. Hence, it must hold term by term, which leads us to the YBG-hierarchy (Yvon, Born, Green),

$$\begin{aligned}k_B T\nabla_{\mathbf{x}_1} g^{(n)}(\mathbf{x}_{(n)}) = \sum_{i=2}^{n} \mathbf{F}_{1i} g^{(n)}(\mathbf{x}_{(n)}) \\ + \rho\int_{\Omega} \mathbf{F}_{1n+1} g^{(n+1)}(\mathbf{x}_{(n+1)})\ d\mathbf{x}_{n+1},\end{aligned} \tag{2.48}$$

where we used $\rho^{(n)} = \rho^n g^{(n)}$ and set the external forces to $\mathbf{F}_i^{ext} = 0$.

2.4.3 The Born-Green Equation

The YBG- and the BBGKY-hierarchy are not immediately useful, since they relate one unknown function to another. In order to solve (2.48) a closure relation between

$g^{(n+1)}$ and $g^{(n)}$ is required. The case $n = 2$ is the best investigated one, see [47]. Since for isotropic fluids we have

$$\int_\Omega \mathbf{F}_{ij} g^{(2)}(\mathbf{x}_i, \mathbf{x}_j)\, d\mathbf{x}_j = 0, \tag{2.49}$$

the relation can be transformed to

$$\begin{aligned} &k_B T \nabla_{\mathbf{x}_1} \left(\ln(g^{(2)}(\mathbf{x}_1, \mathbf{x}_2)) + \beta v(\mathbf{x}_1, \mathbf{x}_2) \right) \\ &\quad = \rho \int_\Omega \mathbf{F}_{13} \left(\frac{g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{g^{(2)}(\mathbf{x}_1, \mathbf{x}_2)} - g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) \right) d\mathbf{x}_3. \end{aligned} \tag{2.50}$$

Together with the Kirkwood superposition approximation [59]

$$g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}(\mathbf{x}_2, \mathbf{x}_3), \tag{2.51}$$

this yields the Born-Green equation

$$\begin{aligned} &k_B T \nabla_{\mathbf{x}_1} \left(\ln(g^{(2)}(\mathbf{x}_1, \mathbf{x}_2)) + \beta v(\mathbf{x}_1, \mathbf{x}_2) \right) \\ &\quad = \rho \int_\Omega \mathbf{F}_{13}\; g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) \left(g^{(2)}(\mathbf{x}_2, \mathbf{x}_3) - 1 \right) d\mathbf{x}_3. \end{aligned} \tag{2.52}$$

For a given pair potential $v(\mathbf{x}_1, \mathbf{x}_2)$ the Born-Green equation can be solved to give $g^{(2)}$. For low densities ρ the results are in good agreement with those obtained by Monte Carlo or molecular dynamics methods or analytical results in the case of a hard sphere fluid [47]. This is quite astonishing if one considers the fact that all correlations of order three and higher are neglected by the Kirkwood approximation. The effects due to the higher order correlations become more definite at higher densities where this closure results in less satisfactory pair distributions.

Finding better closures for (2.50) is a very challenging task. According to Meeron [77] and Salpeter [105] the triplet correlation function can formally exact be expressed as

$$g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}(\mathbf{x}_2, \mathbf{x}_3)\; e^{\tau(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \rho)} \tag{2.53}$$

with $\tau(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \rho) = \sum_{n=1}^{\infty} \rho^n \delta_{n+3}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$. The coefficients $\delta_{n+3}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ consist of certain terms of the Mayer cluster expansion, the so-called simple 123-irreducible diagrams, see [77, 105] or [47] for details. The coefficients δ_4 and δ_5 were computed for a Lennard-Jones fluid [97, 98], but the computation of higher order terms still is not feasible with today's computers.

Better results for dense fluids can be obtained by using the Fisher-Kopeliovich closure [35] for the quadruplet distribution function

$$\begin{aligned} &g^{(4)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) \\ &\quad \approx \frac{g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4) g^{(3)}(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4) g^{(3)}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)}{g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}(\mathbf{x}_1, \mathbf{x}_4) g^{(2)}(\mathbf{x}_2, \mathbf{x}_3) g^{(2)}(\mathbf{x}_2, \mathbf{x}_4) g^{(2)}(\mathbf{x}_3, \mathbf{x}_4)}. \end{aligned} \tag{2.54}$$

Inserting into (2.48) for $n = 3$ this gives a relation for the triplet distribution function $g^{(3)}$, called BGY2 equation [70,94]. Lee et al. [70,94] computed $g^{(3)}$ for a hard sphere fluid. The results were significantly better than those obtained with the Kirkwood approximation and with the Percus-Yevick model (2.59) which we will introduce later. They proved that the BGY2 theory is superior to the closure (2.53) truncated after δ_5. But to our knowledge the BGY2 theory has never been applied to other potential functions than the hard sphere potential.

In summary, there is still no other closure than the Kirkwood approximation (2.51) which is physically reasonable on the one hand and computationally tractable on the other hand. Yet, as we pointed out before, the accuracy of the superposition approximation for low densities is satisfactory and comparable to the accuracy of the most popular closures for the Ornstein-Zernike equation, that we will introduce in the following section.

2.4.4 The Ornstein-Zernike Equation

Ornstein and Zernike [82] first introduced a new concept, where the pair distribution function $g^{(2)}(\mathbf{x}_1, \mathbf{x}_2)$ is calculated by means of so called correlation functions, see e.g. [47]. It describes the fact that the total correlation $h(\mathbf{x}_1, \mathbf{x}_2) = g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) - 1$ of two particles is due to the direct correlation of these particles $c(\mathbf{x}_1, \mathbf{x}_2)$ and the indirect correlation mediated through all other particles. Mathematically, this is written as

$$\begin{aligned} h(\mathbf{x}_1, \mathbf{x}_2) = c(\mathbf{x}_1, \mathbf{x}_2) &+ \rho \int_\Omega c(\mathbf{x}_1, \mathbf{x}_3) c(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3 \\ &+ \rho^2 \int_\Omega c(\mathbf{x}_1, \mathbf{x}_3) c(\mathbf{x}_3, \mathbf{x}_4) c(\mathbf{x}_2, \mathbf{x}_4)\ d\mathbf{x}_3 d\mathbf{x}_4 \\ &+ \dots, \end{aligned} \tag{2.55}$$

with ρ the density of the fluid. By substituting the definition for $h(\mathbf{x}_2, \mathbf{x}_3)$ back into (2.55), this can also be represented as the famous Ornstein-Zernike equation

$$h(\mathbf{x}_1, \mathbf{x}_2) = c(\mathbf{x}_1, \mathbf{x}_2) + \rho \int_\Omega c(\mathbf{x}_1, \mathbf{x}_3) h(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3. \tag{2.56}$$

An important fact to note is that the Ornstein-Zernike equation is exact; there is no approximation involved. It simply relates the total correlation function $h(\mathbf{x}_1, \mathbf{x}_2)$ to the direct correlation function $c(\mathbf{x}_1, \mathbf{x}_2)$, which is defined by (2.56). However, both functions are not known in advance. It is plausible to suppose that the range of the function $c(\mathbf{x}_1, \mathbf{x}_2)$ is comparable to that of the pair potential $v(\mathbf{x}_1, \mathbf{x}_2)$, whereas the range of $h(\mathbf{x}_1, \mathbf{x}_2)$ can be much longer due to the effects of all other particles.

In order to solve (2.56) one needs a second equation, the closure relation. This also relates the direct and the total correlation function and includes additionally

the pair potential v. Such closure relations can be obtained by functional expansions of the density or of the pair potential where one particle is kept fixed, see [47]. The most popular example of a closure relation is the hypernetted chain closure (HNC)

$$h(r) = e^{-\beta v(r)+h(r)-c(r)} - 1. \tag{2.57}$$

Since all functions in (2.57) depend only on the distance of the particles, it can be written in terms of this distance directly, i.e. $r = r_{12} = |\mathbf{x}_1 - \mathbf{x}_2|$ in this case. As mentioned above, relation (2.57) is not exact but approximated. In order to account for the error, a so-called bridge function $b(r)$ is introduced,

$$h(r) = e^{-\beta v(r)+h(r)-c(r)+b(r)} - 1. \tag{2.58}$$

Formally, equation (2.58) is now exact. There is however no exact expression for $b(r)$ which is computable. Hence, the approximation of the bridge function is the key point of today's integral equation theories, see [47]. Popular bridge functions for simple liquids are the hypernetted chain (HNC) closure (2.57), $b(r) = 0$, or the Percus-Yevick approximation

$$b(r) = \ln\left(1 + h(r) - c(r)\right) - h(r) + c(r). \tag{2.59}$$

Together with either form of (2.58), the system of equations (2.56) can be solved for a given pair potential $v(r)$.

Extensions to the HNC and PY model have been considered in the literature, see [47] for an overview. Since the HNC and PY models can be derived by functional Taylor expansions truncated at first order, the question naturally arises whether a truncation at second order would significantly improve the model. These second order models have been tested for Lennard-Jones fluids and show a clear improvement compared to the first order models, with the drawback of being numerically awkward to handle. In the case of the HNC closure the approximation can be improved by known bridge functions of a simple reference system. This is called the reference HNC approximation (RHNC). To this end, the bridge function $b(r)$ in (2.58) is computed for the reference system $b_0(r)$. One usually chooses a hard sphere fluid as reference fluid since it is the only fluid for which the bridge function is known with sufficient accuracy. The diameter of the spheres appears as a free parameter of the model and can be adjusted to reproduce known results. The overall agreement with the results of Monte Carlo simulations is then very good. Other closure relations were proposed that introduce an adjustable parameter or function. These closures contain the HNC and PY closures as special cases and allow to adjust the approximation in order to yield improved results for a reference system such as a fluid of hard spheres, see [76] for an overview. Even though these adjustable closure relations can be in very good agreement with known results, they do not lead to a self consistent theory of liquid state, since the determination of the free parameters is an empiric procedure.

2.4.5 Comparison of Different Approximations

We will now shortly discuss the differences between the different equations and approximations. At a first sight, it seems that the Born-Green equation (2.52) and the Ornstein-Zernike equation (2.56) together with the HNC or PY closure are not linked together. But it is not surprising that they can be transformed into each other. For example, it can be shown that the Ornstein-Zernike equation with a special closure relation recovers the Born-Green equation, see [47]. Therefore, this is really a comparison of different approximations of the same model. However, we will stay with the strict separation of the models based on the YBG-hierarchy and the Ornstein-Zernike based models in our discussion, since we want to underline their difference with respect to the numerical solution. Moreover, their respective form is predestinated for the usage of certain approximations, even though the models can be transformed into each other. This will be investigated in chapter 5 in more detail. Nevertheless, we will now present some results for pure fluids computed with the Born-Green equation (BG) and the HNC- and the PY-closure together with the Ornstein-Zernike equation in order to illustrate their differences without discussing any numerical procedure. A more detailed investigation of the differences between the integral equation methods can be found e.g. in [58].

The computed pair distribution functions and the direct correlation functions for a simple monoatomic fluid at reduced temperature $T = 1.65$ and number densities $\rho = 0.3, 0.5, 0.8$ are shown in Figures 2.1, 2.2 and 2.3. The interaction is described by the Lennard-Jones potential (with $\epsilon = 1$ and $\sigma = 1$)

$$v^{LJ}(r) = 4\epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right). \tag{2.60}$$

In the case of the Born-Green equation, the direct correlation function is computed by inserting the total correlation function $h = g - 1$ into the Ornstein-Zernike equation (2.56). Here, we omit the subscript (2) od the pair distribution function.

An important observation is that the resulting functions in Figures 2.1 - 2.3 become more similar with decreasing density. This is substantiated by the fact that all three approximations give the correct result of

$$g(r) = e^{-\beta v^{LJ}(r)} \tag{2.61}$$

in the limit $\rho \to 0$. It can be shown that they all yield the correct expression of order ρ of the density expansion of g, see [47]. But they differ for all higher order terms of the expansion. Hence, the differences become larger with increasing ρ. Here, the BG equation plays a special role. As can be seen in Figure 2.3 (right), the BG model leads to an unphysical direct correlation function at density $\rho = 0.8$. This looks like an obvious deficiency of this model, but shows how sensitive the total and direct

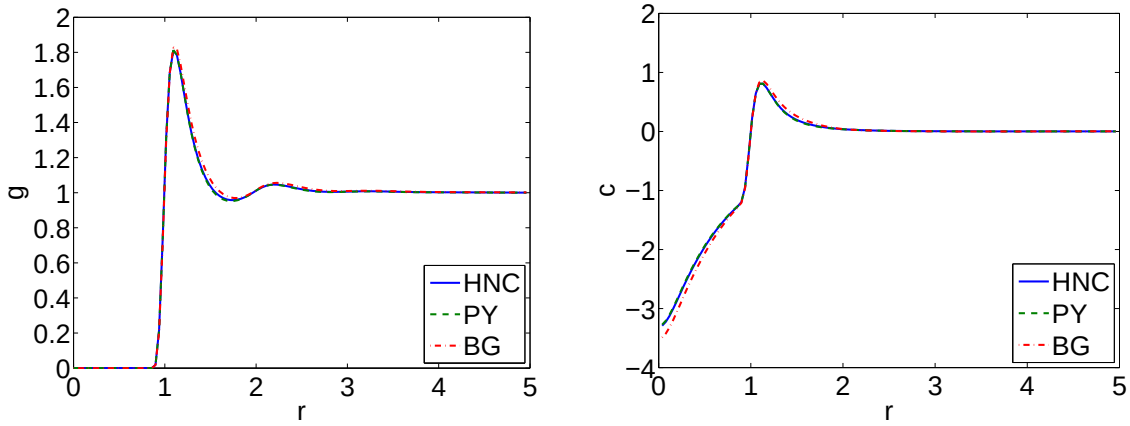

Figure 2.1. *Left: Pair distribution function for $\rho = 0.3$. Right: Corresponding direct correlation function.*

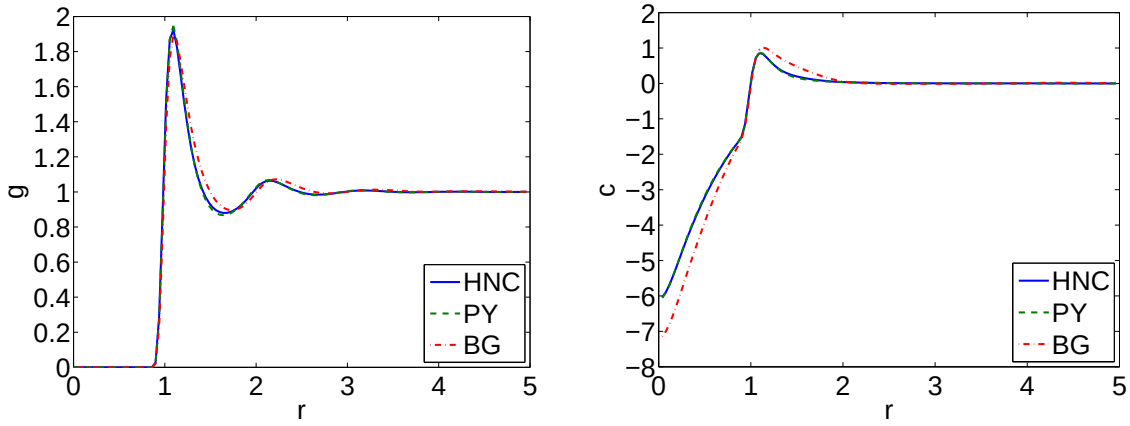

Figure 2.2. *Left: Pair distribution function for $\rho = 0.5$. Right: Corresponding direct correlation function.*

correlation functions are linked together. Since the direct correlation function does not appear in the BG model, the approximation errors can lead to this behavior. But the computed pair distribution functions can still be of the same quality as in the case of the other approximations.

In the literature, solutions to the BG, HNC and PY equations have been obtained for a variety of pair potentials over a wide range of temperatures and densities. Comparison of results for the Lennard-Jones potential showed that the PY approximation is superior at all thermodynamic states that have been studied. But even this approximation produces a noticeable error for the pair distribution func-

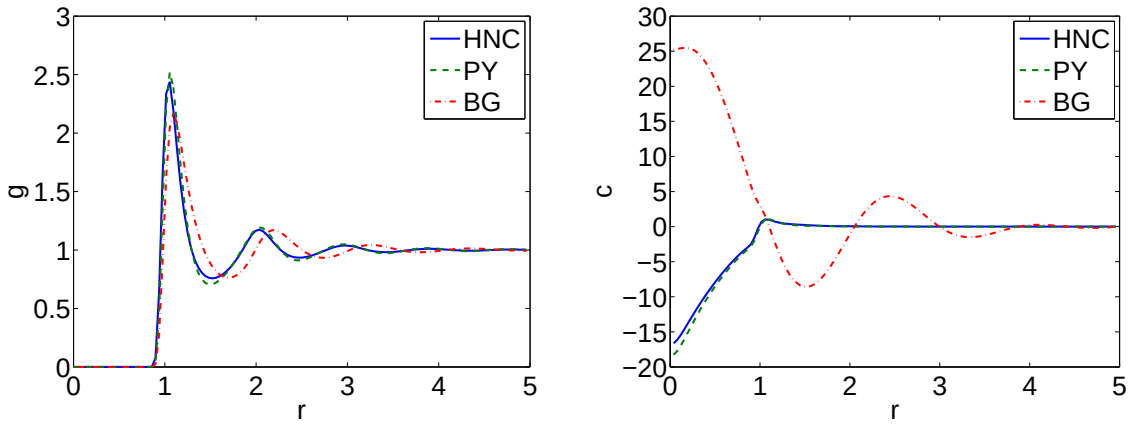

Figure 2.3. *Left: Pair distribution function for $\rho = 0.8$. Right: Corresponding direct correlation function.*

tion [47]. The main peak in $g(r)$ is too large in magnitude and appears at to small a value of r. The following pattern of oscillation is out of phase compared to the one obtained by Monte Carlo simulations. The situation gets even worse in the vicinity of critical points of the phase diagram, such as the liquid-gas transition. The characteristics of these regions are large density fluctuations. This leads to anomalous behavior of thermodynamic properties of the fluid which is difficult to extract from numerical solutions of the integral equations. Hence, all three approximation routes are not regarded to be a satisfactory theory of liquid state.

The extensions of the BG equation and the HNC and PY approximations already described in sections 2.4.2 and 2.4.4 lead to significant improvements of the computed thermodynamic properties, even though very accurate results still cannot be obtained in all cases. But these extensions all entail the drawback of being numerically less practicable. The development of new closure relations that improve the accuracy of the models and still lead to computationally tractable methods is subject of today's research.

2.4.6 Summary

The liquid state integral equation theories provide methods to compute the reduced distribution functions of pure liquids without explicitly performing the integration over phase space. They are based on the YBG-hierarchy on the one hand and on the Ornstein-Zernike equation on the other hand. Even though the equations are only different formulations of the same theory, they all require specific approximations in order to be solvable. Hence, the methods yield different results according to their approximations, whereas their respective formulations yield different numerical

algorithms. This has led to distinct fields of application for the YBG-hierarchy and the Ornstein-Zernike based methods. The latter have been applied to pure simple fluids, complex molecular fluids and also to solute-solvent systems, which we will introduce in the following chapter. Methods based on the YBG-hierarchy, however, have found less attention in the literature. Beside their application to simple fluids and fluids at interfaces they are mainly employed for the investigation of polymers. This is indeed related to molecular fluids, but, to our knowledge, these methods have never been applied to other molecular fluids or solute-solvent systems.

Chapter 3
Solute-Solvent Systems

The investigation of chemical or biological processes on the microscopic scale is a very important application of molecular simulations. To this end, molecular systems in liquid solution are to be simulated. The protein folding problem is a very famous example of such a system. Here, the three-dimensional configuration of the protein, which consists of a linear sequence of amino acids, is to be computed. Proteins naturally appear in aqueous solution. Hence, the incorporation of the solvent effects is one of the most important aspects concerning the accurate reproduction of experimental data. The solvent can have a variety of influences on the solute. These include e.g. hydrogen-bonding or dielectric shielding of the long-range Coulomb forces.

The naive, but yet important approach is to choose an all-atom representation for the whole system, i.e., the solute as well as the solvent molecules are included explicitly. This method can provide the most detailed description of the effects the solvent has on the solute. However, it is not exempt from approximations. Beside the general uncertainty, how detailed the empirical force fields used in molecular simulations can reproduce experimental data, the evaluation of the potentials often implies approximations in order to be efficiently computable. Examples are the truncation of the potential behind a certain cut-off radius or the use of an infinite number of periodic cells when computing long-rang forces with Ewald techniques, see [46] for details.

The large computational costs of sampling the solvent degrees of freedom are, however, the major problem when explicit solvent molecules are included into the simulation box. In order to incorporate the solvent effects sufficiently well, a large number of solvent molecules is necessary to form the bulk solution. Hence, a major fraction of time is spent for computing a detailed trajectory of the solvent, although the behavior of the solute is at the center of interest. Due to these problems of explicit solvent representation, computationally less expensive models for including solvent effects are of great interest. These models should describe the effects with

appropriate accuracy without introducing new degrees of freedom to the system. In contrast to the simulation with explicit solvent molecules, such models are called implicit solvent models.

It is instructive to begin the description of implicit solvent models with a statistical mechanics description of the solvent effects. Hence, we will first give the definition of the potential of mean force. Hereafter, some of the most popular implicit solvent models are presented before we focus our attention to the liquid state integral equation methods.

3.1 Potential of Mean Force

We now consider a system consisting of a single arbitrary molecule, which we call the solute M, and a bulk of solvent molecules, the solvent S. The solute consists of N_M particles whereas the solvent consists of N_S particles. The Hamiltonian of this system can be written as

$$H(\mathbf{p}^M, \mathbf{p}^S, \mathbf{x}^M, \mathbf{x}^S) = \frac{1}{2}\sum_{i=1}^{N_M} \frac{(\mathbf{p}_i^M)^2}{m_i^M} + \frac{1}{2}\sum_{i=1}^{N_S} \frac{(\mathbf{p}_i^S)^2}{m_i^S} + V(\mathbf{x}_1^M, \cdots, \mathbf{x}_{N_M}^M, \mathbf{x}_1^S, \ldots, \mathbf{x}_{N_S}^S), \tag{3.1}$$

where $\mathbf{p}^M, \mathbf{p}^S$ are the momenta, $\mathbf{x}^M, \mathbf{x}^S$ the positions and m_i^M, m_i^S the masses of the solute and the solvent particles, respectively. The potential can be further divided into a part V_M describing the intramolecular interaction of the solute, a part V_S describing the interactions within and between the solvent molecules and a part V_{MS} which consists of the interactions between the solute and the solvent atoms

$$V(\mathbf{x}^M, \mathbf{x}^S) = V_M(\mathbf{x}^M) + V_S(\mathbf{x}^S) + V_{MS}(\mathbf{x}^M, \mathbf{x}^S). \tag{3.2}$$

Hence, the Hamiltonian can also be written in separated form

$$H(\mathbf{p}^M, \mathbf{p}^S, \mathbf{x}^M, \mathbf{x}^S) = H_M(\mathbf{p}^M, \mathbf{x}^M) + H_S(\mathbf{p}^M, \mathbf{x}^S) + H_{MS}(\mathbf{p}^M, \mathbf{p}^S, \mathbf{x}^M, \mathbf{x}^S). \tag{3.3}$$

This system still includes the solvent degrees of freedom explicitly. Observables of this system can be computed by microscopic averages in the canonical ensemble as

$$\begin{aligned} \langle a \rangle &= C^{-1} \int_{\Omega_N} a(\mathbf{x}_M, \mathbf{x}_S) e^{-\beta V(\mathbf{x}_M, \mathbf{x}_S)} \, d\mathbf{x}_M d\mathbf{x}_S \\ \text{with} \quad C &= \int_{\Omega_N} e^{-\beta V(\mathbf{x}_M, \mathbf{x}_S)} \, d\mathbf{x}_M d\mathbf{x}_S, \end{aligned} \tag{3.4}$$

where we assumed that the microscopic quantity a does not depend on the momenta. The domain Ω_N denotes the spatial part of the phase space Γ_N. If we now further

assume that a does not depend on the solvent degrees of freedom, the integral can be written as

$$\langle a \rangle = C^{-1} \int_{\Omega_{N_M}} a(\mathbf{x}_M) e^{-\beta V_M(\mathbf{x}_M)} \int_{\Omega_{N_S}} e^{-\beta(V_{MS}(\mathbf{x}_M,\mathbf{x}_S)+V_S(\mathbf{x}_S))}\, d\mathbf{x}_S d\mathbf{x}_M \tag{3.5}$$

with Ω_{N_M} and Ω_{N_S} the configurational domains of the solute and the solvent, respectively. Hence, the inner integral can be computed separately. The formal integration of this inner part leads directly to the idea of the potential of mean force (PMF). It is defined by integrating the Boltzmann factor $e^{-\beta V(\mathbf{x})}$ over the solvent degrees of freedom,

$$\begin{aligned} e^{-\beta V^{PMF}(\mathbf{x}^M)} &= C_S^{-1} \int_{\Omega_{N_S}} e^{-\beta(V_M(\mathbf{x}^M)+V_S(\mathbf{x}^S)+V_{MS}(\mathbf{x}^M,\mathbf{x}^S))}\, d\mathbf{x}^S \\ &= e^{-\beta V_M(\mathbf{x}^M)} C_S^{-1} \int_{\Omega_{N_S}} e^{-\beta(V_S(\mathbf{x}^S)+V_{MS}(\mathbf{x}^M,\mathbf{x}^S))}\, d\mathbf{x}^S \qquad (3.6) \\ \text{with } C_S &= \int_{\Omega_{N_S}} e^{-\beta V_S(\mathbf{x}^S)}\, d\mathbf{x}^S. \qquad (3.7) \end{aligned}$$

The PMF can also be written in an additive way

$$\begin{aligned} V^{PMF}(\mathbf{x}^M) &= V_M(\mathbf{x}^M) - \frac{1}{\beta} \ln \left(C_S^{-1} \int_{\Omega_{N_S}} e^{-\beta(V_S(\mathbf{x}^S)+V_{MS}(\mathbf{x}^M,\mathbf{x}^S))}\, d\mathbf{x}^S \right) \qquad (3.8) \\ &= V_M(\mathbf{x}^M) + W(\mathbf{x}^M), \qquad (3.9) \end{aligned}$$

where $W(\mathbf{x}^M)$ is defined by (3.9) and contains only the energy due to the solute-solvent interaction. It is therefore also called the solvation free energy. The Hamiltonian of the reduced system now reads as

$$H^{PMF}(\mathbf{p}^M, \mathbf{x}^M) = \frac{1}{2} \sum_{i=1}^{N_M} \frac{(p_i^M)^2}{m_i^M} + V^{PMF}(\mathbf{x}^M) \tag{3.10}$$

and the microscopic average (3.5) can be written as

$$\begin{aligned} \langle a \rangle &= C_{PMF}^{-1} \int_{\Omega_{N_M}} a(\mathbf{x}_M) e^{-\beta V^{PMF}(\mathbf{x}_M)}\, d\mathbf{x}_M \qquad (3.11) \\ \text{with} \quad C_{PMF} &= \int_{\Omega_{N_M}} e^{-\beta V^{PMF}(\mathbf{x}_M)}\, d\mathbf{x}_M. \qquad (3.12) \end{aligned}$$

A comparison of equations (3.5) and (3.11) shows that the introduction of the PMF is just a formal transformation. The ensemble averages (3.4) and (3.11) lead to

the exact same results. The transformation shifts the problem of sampling the solvent degrees of freedom to the computation of the integral defining the PMF (3.8). In practice, the integral can be solved exactly only for very simple systems that are not of interest with respect to the application in biomolecular simulations. Hence, approximate methods have to be employed in order to compute the PMF. In principle, molecular dynamics or Monte Carlo type simulations can be applied. However, due to the high dimension of the phase space the convergence is very slow. For applications, where the PMF has to be computed repeatedly, one uses much simpler models instead. These models are either based on a continuum description of the solvent or use very crude approximations to the solute-solvent interactions. They include model parameters which are fitted to reproduce results from experimentally well-known systems. Applied to other solute-solvent systems, these models can yield qualitative approximations of some important solvent effects.

Models which incorporate the solvent effects by (approximations to) the PMF (3.8) are called implicit solvent models. We will give a short overview of the most popular implicit solvent models as they are used in molecular dynamics or Monte Carlo simulations of biomolecular systems.

3.2 Implicit Solvent Models

As we have seen in the previous section, the computation of the PMF involves the solution of a high dimensional integral which is prohibitive when repeatedly evaluation of the PMF is necessary, as in molecular dynamics or Monte Carlo simulations. Hence, approximations to the PMF are used, which are computationally more efficient. In most force fields, the PMF is represented as a sum of a part induced by the short-range repulsive and by the van der Waals forces W^{vdW}, and a part due to the long-range electrostatic interactions W^{pol},

$$W(\mathbf{x}^M) = W^{vdW}(\mathbf{x}^M) + W^{pol}(\mathbf{x}^M). \tag{3.13}$$

Formally, the two parts of (3.13) are given by

$$W^{vdW}(\mathbf{x}^M) = -\frac{1}{\beta}\ln\left(C_S^{-1}\int_{\Omega_{N_S}} e^{-\beta(V_S(\mathbf{x}^S)+V_{MS}^{vdW}(\mathbf{x}^M,\mathbf{x}^S))}\,d\mathbf{x}^S\right), \tag{3.14}$$

$$W^{pol}(\mathbf{x}^M) = -\frac{1}{\beta}\ln\left(\frac{\int_{\Omega_{N_S}} e^{-\beta(V_S(\mathbf{x}^S)+V_{MS}^{vdW}(\mathbf{x}^M,\mathbf{x}^S)+V_{MS}^{pol}(\mathbf{x}^M,\mathbf{x}^S))}\,d\mathbf{x}^S}{\int_{\Omega_{N_S}} e^{-\beta(V_S(\mathbf{x}^S)+V_{MS}^{vdW}(\mathbf{x}^M,\mathbf{x}^S))}\,d\mathbf{x}^S}\right) \tag{3.15}$$

with C_S defined as in (3.7). Due to this separation, different models for the approximation of the short-range and the long-range effects of the solvent can be employed. In the next sections we will present some of the most important models for both parts (3.14) and (3.15).

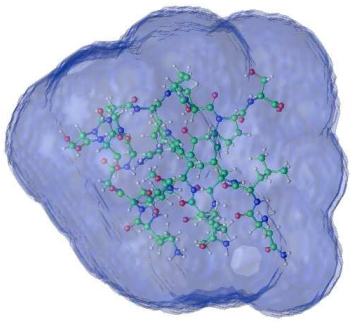

Figure 3.1. *SASA of a sample protein (Trp-cage)*

3.2.1 Solvent Accessible Surface Area

One of the most important implicit solvent models is the model based on the solvent accessible surface area (SASA). It goes back to Lee and Richards [69] who have tried to quantify the burial of hydrophobic side chains in the protein folding problem. They have observed that the hydrophobic side chains of the proteins tend to cluster together in the interior of the protein, whereas the hydrophilic side chains are better soluble in water. They concluded, that the surface tension of the solvent in direct contact with the surface of the protein is a measure for the force exerted to the solute atoms due to the solvent. This surface tension is proportional to the SASA.

The SASA encloses the volume that the solvent atoms are excluded from. Following the concept of Lee and Richards [69], it is traced out by the center of a solvent probe sphere rolling over the molecular surface of the solute. The molecular surface is simply defined by the union of the surfaces of the spheres determined by the van der Waals radius around any atom. In other words, the solvent-excluded volume is the union of the so-called expanded atoms. The expanded atom is a sphere centered at the position of the solute atom with its van der Waals radius increased by the solvent probe radius, see [101]. Figure 3.1 shows the SASA of a small protein.

Formally, the solvation free energy due to the SASA model is written as

$$W^{vdW}(\mathbf{x}^M) \approx \sum_{i=1}^{N_M} \sigma_i A_i(\mathbf{x}^M) \tag{3.16}$$

with the empirical solvation parameter σ_i of atom i and its fraction of the SASA A_i. The A_i depend on the positions of all solute atoms because they contain only the fraction of the surface accessible to the solvent. Sometimes an even simpler approach is used with

$$W^{vdW}(\mathbf{x}^M) \approx \sigma A_{tot}(\mathbf{x}^M). \tag{3.17}$$

Here, the (non-polar) solvation free energy is proportional to the total SASA and an empirical parameter σ.

During the last decades many algorithms were introduced to efficiently compute the SASA. Since the SASA model is intended to be used as an implicit solvent model within a molecular dynamics or Monte Carlo simulation, its fast computation is of major interest. The computational cost should be of the same order as the evaluation of the solute potential, i.e. $\mathcal{O}(N_M)$ or $\mathcal{O}(N_M \log(N_M))$. The existing algorithms can be divided into approximate and exact methods. The first exact analytical algorithms were introduced by Connolly [22] and Richmond [101]. Here, the accessible surfaces of intersecting spheres are computed. These algorithms were later improved with respect to computational efficiency, see [36, 124], and stability, see [31, 44]. The first approximate algorithm to compute the SASA was developed by Shrake and Rupley [107] and later improved by Legrand and Merz [71]. To this end, a large number of points is distributed on the surface of a each solute atom. The SASA is proportional to the number of points that are not located inside any other van der Waals sphere of a solute atom. Another approximate algorithm by Caflisch et al. [34] uses an approximate analytical expression with empirical fitted parameters. More recently, an algorithm has been developed that uses alpha shapes for computing the SASA, see [29, 73].

The approximate algorithms are usually computationally more efficient, but share the drawback of computing the SASA inaccurate with an error of several percents. A general problem of using the SASA in implicit solvent models for molecular dynamics is the discontinuity of the derivatives of the SASA with respect to the positions of the solute atoms. This can lead to numerical instabilities during a simulation, see [125] for some situations where these problems occur. In the context of alpha shapes, the *simulation of simplicity*-principle [30] is proposed to handle these instabilities. Basically, the input data, i.e. the positions of the solute atoms, are moved by a small displacement in order to avoid the degenerate cases without affecting the results very much.

The SASA model has become the standard tool in molecular simulation. It describes the non-polar solvent effects only on a qualitative level of detail but has the major advantage of being efficiently computable. It is included in most available codes for biomolecular simulation, e.g. CHARMM [16], AMBER [23] and TINKER [91].

A very simple approach which is related to the SASA model is the scaled particle theory [90, 96, 112]. The reversible work to produce a spherical cavity in a solvent consisting of hard spheres can be computed analytically, as long as the radius of the cavity is smaller than a certain value. Generalizations have been introduced for van der Waals liquids which involve experimental parameters. The leading terms of this expression represent the solvent-exposed surface area of the cavity and its curvature. Neglecting the curvature term leads to the idea, on which the SASA model is based.

3.2.2 The Poisson-Boltzmann Model

Popular implicit solvent models for the electrostatic part of the PMF employ a continuum description of the electrostatic interaction, where the solvent is treated as a continuum with relative permittivity ε_S (typically $\varepsilon_S = 80$ for water). To this end, the electrostatic potential $\Phi_S(\mathbf{r})$ acting on the charges of the solute can be computed by the Poisson equation

$$\nabla \cdot \varepsilon(\mathbf{r}) \nabla \Phi_S(\mathbf{r}) = -\rho(\mathbf{r}) \qquad \text{in } \Omega \tag{3.18}$$

with the charge density $\rho(\mathbf{r})$ of the solute defined by

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_M} q_i \delta(\mathbf{r} - \mathbf{x}_i^M) \tag{3.19}$$

and q_i the charge of the solute particle i. The position-dependent relative permittivity $\varepsilon(\mathbf{r})$ is defined as

$$\varepsilon(\mathbf{r}) = \begin{cases} \varepsilon_M & \text{for } \mathbf{r} \in \Omega_M \\ \varepsilon_S & \text{for } \mathbf{r} \in \Omega \setminus \Omega_M \end{cases} \tag{3.20}$$

where $\varepsilon_M = 1$ is the relative permittivity in the volume of the solute $\Omega_M \subset \Omega$ and $\Omega \subset \mathbb{R}^3$ is the total domain of the system. For simplicity, the dielectric constant in vacuum is set to $\varepsilon_0 = 1$ in our discussion. The electrostatic part of the PMF can be computed from the solution of (3.18) by the so-called reaction field $\Phi_{rf}(\mathbf{r}) = \Phi_S(\mathbf{r}) - \Phi_V(\mathbf{r})$, which is the electrostatic potential $\Phi_S(\mathbf{r})$ minus the reference potential $\Phi_V(\mathbf{r})$. The reference potential represents the solution of (3.18) with $\varepsilon(\mathbf{r})$ set to 1 everywhere. The PMF is then approximated by

$$W^{pol}(\mathbf{x}^M) \approx \sum_{i=1}^{N_M} q_i \Phi_{rf}(\mathbf{x}_i^M). \tag{3.21}$$

The Poisson equation can be extended to the Poisson-Boltzmann equation if ions in the solvent are present. The electrostatic potential is then given by

$$\nabla \cdot \varepsilon(\mathbf{r}) \nabla \Phi_S(\mathbf{r}) - \kappa^2(\mathbf{r}) \Phi_S(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{3.22}$$

with κ the Debeye length which characterizes the screening effects due to the presence of the ions. Its position-dependence is similar to that of ε, i.e., it varies sharply from 1 inside the solute to a value > 1 in the bulk solvent.

In this model, it is assumed that the relative permittivity is uniform except in the vicinity of the solute-solvent boundary. Such a form for ε can be derived from a statistical mechanics integral in the limit of small solvent molecules described as

non-polarizable hard spheres, see [11]. The results of the Poisson-Boltzmann model of continuum electrostatics depend sensitively on the atomic partial charges and the location of the dielectric boundary, i.e. the boundary $\partial\Omega_M$ of Ω_M. If constructed on the basis of the SASA, the atomic radii must be considered as free parameters of the implicit solvent model as well, see [81, 110] for parametrization schemes.

The solution of equation (3.18) or (3.22) can be obtained by standard numerical algorithms using finite difference discretization. This approach is however still to costly compared with the evaluation of the inter-solute potential V_M. Alternatively, the boundary element method, which uses finite elements distributed on the dielectric boundary, can be employed, see [128]. The computation of analytical gradients of the PMF is possible in either approach. But if repeated computation of both the PMF and its gradient is required, all numerical methods for the Poisson-Boltzmann equation are computationally too costly. Hence, methods have been developed that approximate the exact continuum electrostatic potential.

Generalized Born Model

The most popular approximation for the Poisson-Boltzmann equation is the so-called generalized Born (GB) model. The derivation of this model starts with the energy of a sphere with radius α_i and a point charge q_i in its center immersed into a medium of relative permittivity ε_S,

$$W^{born} = -\left(1 - \frac{1}{\varepsilon_S}\right)\frac{q_i^2}{2\alpha}. \tag{3.23}$$

This is the Born equation. The total electrostatic energy of a set of charged spheres at infinite distance is then given by

$$\begin{aligned} V_{tot}^{pol} &= \sum_{i=1}^{N_M}\sum_{j=i+1}^{N_M}\frac{q_i q_j}{\varepsilon_S r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon_S}\right)\sum_{i=1}^{N_M}\frac{q_i^2}{\alpha_i} \\ &= \sum_{i=1}^{N_M}\sum_{j=i+1}^{N_M}\frac{q_i q_j}{r_{ij}} - \left(1 - \frac{1}{\varepsilon_S}\right)\sum_{i=1}^{N_M}\sum_{j=i+1}^{N_M}\frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon_S}\right)\sum_{i=1}^{N_M}\frac{q_i^2}{\alpha_i} \end{aligned} \tag{3.24}$$

where $r_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$. The first term of (3.24) is the standard Coulomb energy in vacuum whereas the second and the third term of (3.24) are due to the presence of a medium of relative permittivity ε_S. In order to account for the actual shape of the solute molecule the two last terms are combined and extended to

$$W^{pol} \approx -\left(1 - \frac{1}{\varepsilon_S}\right)\sum_{i=1}^{N_M}\sum_{j=1}^{N_M}\frac{q_i q_j}{f_{GB}}, \tag{3.25}$$

with the deshielding function f_{GB} defined by

$$f_{GB} = \sqrt{(r_{ij} + \alpha_{ij})^{-D}} \quad \text{with } \alpha_{ij} = \sqrt{\alpha_i \alpha_j},\ D = \frac{r_{ij}^2}{2\alpha_{ij}^2}. \tag{3.26}$$

The α_i are treated as parameters in this derivation of Still and coworkers [111]. They are fitted to give comparable energies as other methods for small molecules. Improvements of the original deshielding function (3.26) have been proposed by other authors, see e.g. [54].

Due to the semianalytical nature of the GB model, the computational expense for the evaluation of the energy as well as of the gradient is comparatively low. The obtained accuracy is comparable with other methods to approximate the electrostatic energy [111]. However, problems arise for larger solutes such as proteins, since typical effects as the charge burial in the interior of the protein are difficult to account for. Nevertheless, the GB model is a very popular implicit solvent model in molecular dynamics simulations. Combined with the SASA model for the non-polar part of the PMF, it forms the GB/SA continuum model for solvation and is implemented in most available packages for biomolecular simulation, see also Section 3.2.1.

3.2.3 Specialized Implicit Solvent Models

If one assumes that the solvation free energy arises due to the short-range interaction between solute and solvent, the SASA model is a reasonable approximation. The Coulomb force, however, is a long-range interaction. Nevertheless, SASA models are used to approximate the full PMF including electrostatic interactions. These models are called full SASA models, see [104]. An important drawback of the full SASA models is the difficulty in taking into account the dielectric shielding of the electrostatic interactions. The shielding should vary clearly when moving a charged particle from a position fully exposed to the solvent to a position buried in the interior of the solute. One can overcome this deficiency by introducing a distance dependent relative permittivity and by neutralizing residues carrying a net charge.

The most simple but also computational most efficient methods are the so-called knowledge-based potentials. They are especially useful when an extensive search of configurations is required, as for the minimization of the free energy. These potentials are built upon experimental observations of known structures of proteins. By analyzing large data bases of these protein structures one has observed that the number of residue pairs at a certain distance follows the Boltzmann principle. Hence, potentials have been introduced that give a free energy depending on the distance of the residue pairs in the protein. In the most simple approach, the forces are repulsive for polar pairs of residues and attractive for pairs of non-polar residues, see [109]. This gives the correct behavior of folded proteins where non-polar residues

tend to form a core in the interior of the protein, whereas the polar residues tend to reside at the surface of the protein. These methods however can of course not be expected to include all solvation effects or to provide accurate results for the solvation free energy.

The so-called mixed implicit/explicit schemes try to provide a compromise between the accuracy and the computational effort. Here, a limited number of solvent molecules is inserted explicitly in the vicinity of the solute. Typically, the region containing the solute and the explicit solvent molecules is represented by a sphere. Outside this sphere, the bulk solvent is treated implicitly by an effective solvent boundary potential. Different approaches exist with respect to the definition of the sphere radius. In [17] the radius is constant, which means that the number of explicit solvent molecules must be allowed to vary in order to account for density fluctuations. In contrast to that, in [9], the number of explicit solvent molecules is constant and the sphere radius is determined by the outermost solvent molecule. This has the advantage of being more flexible with respect to large configurational changes of the solute. It has been shown in [9] that the solvation free energies do not depend sensitively on the number of explicit solvent molecules with this latter approach.

We presented some of the most popular implicit solvent methods in order to give an overview of the methods that are widely used within Monte Carlo or molecular dynamics simulations for the approximation of the solvent effects. All presented methods share the drawback that they cannot yield to accurate predictions of the solvent free energy in general. Hence, much effort has been put into the development of more accurate but still efficient approximations of the PMF. The most promising methods are based on the liquid state integral equation theories of statistical mechanics. They are able to approximate the mean solvent density around the solute, which in turn can then be used to compute the PMF.

3.3 Computing the PMF via Reduced Distribution Functions

If we consider the PMF $V^{PMF}(\mathbf{x}^M)$ (3.8) and recall the definition of the reduced distribution functions of Section 2.4.1, we can identify the integral leading to the PMF with a reduced distribution function. This way, we can simply write

$$\begin{aligned} V^{PMF}(\mathbf{x}^M) &= -\frac{1}{\beta}\ln(\rho^{N_M} g^{(N_M)}(\mathbf{x}^M)) \\ &= -\frac{1}{\beta}\ln(g^{(N_M)}(\mathbf{x}^M)) - c_\rho \end{aligned} \tag{3.27}$$

with $c_\rho = \frac{1}{\beta}\ln(\rho^{N_M})$. The constant shift c_ρ can be neglected, since it has no influence on the properties of the system. However, (3.27) is just a transformation of (3.8) and

does not simplify the problem of high-dimensional integration in any sense. However, there exist methods that approximate these reduced distribution functions directly. Unfortunately, these approximations become worse with increasing order N_M of the distribution function. It is therefore more convenient to use lower order distribution functions for the computation of the PMF. This can be done if we consider a special but yet important class of potential functions.

To see this, we first consider the computation of the forces of the PMF $-\nabla V^{PMF}$. They are exactly the forces of the full potential $V(\mathbf{x}_M, \mathbf{x}_S)$ averaged over the solvent degrees of freedom with the solute atoms in fixed position,

$$\nabla_{\mathbf{x}^M} V^{PMF}(\mathbf{x}^M) = \langle \nabla_{\mathbf{x}^M} V(\mathbf{x}^M, \mathbf{x}^S) \rangle_{(\mathbf{x}^M)} \tag{3.28}$$

with

$$\langle a(\mathbf{x}^M, \mathbf{x}^S) \rangle_{(\mathbf{x}^M)} = C_{MS}^{-1} \int_{\Omega_{N_S}} a(\mathbf{x}^M, \mathbf{x}^S) e^{-\beta V(\mathbf{x}^M, \mathbf{x}^S)} \, d\mathbf{x}^S \tag{3.29}$$

and

$$C_{MS} = \int_{\Omega_{N_S}} e^{-\beta V(\mathbf{x}^M, \mathbf{x}^S)} \, d\mathbf{x}^S.$$

If we now assume that the solute-solvent interaction potential can be written as a sum over pairwise terms,

$$V_{MS}(\mathbf{x}^M, \mathbf{x}^S) = \sum_{i=1}^{N_M} \sum_{j=1}^{N_S} v_{MS}(\mathbf{x}_i^M, \mathbf{x}_j^S), \tag{3.30}$$

we can transform (3.28) to yield

$$\begin{aligned} \nabla_{\mathbf{x}^M} V^{PMF}(\mathbf{x}^M) &= \nabla_{\mathbf{x}^M} V_M(\mathbf{x}^M) + \nabla_{\mathbf{x}^M} W(\mathbf{x}^M) \\ &= \nabla_{\mathbf{x}^M} V_M(\mathbf{x}^M) + \sum_{i=1}^{N_M} \sum_{j=1}^{N_S} \langle \nabla_{\mathbf{x}_M} v_{MS}(\mathbf{x}_i^M, \mathbf{x}_j^S) \rangle_{(\mathbf{x}_M)} \\ &= \nabla_{\mathbf{x}^M} V_M(\mathbf{x}^M) + \sum_{i=1}^{N_M} \int_{\Omega} \nabla_{\mathbf{x}_M} v_{MS}(\mathbf{x}_i^M, \mathbf{r}) \rho^{(N_M+1)}(\mathbf{r}|\mathbf{x}^M) \, d\mathbf{r} \end{aligned} \tag{3.31}$$

with the conditional probability

$$\rho^{(N_M+1)}(\mathbf{r}|\mathbf{x}^M) = \frac{\rho^{(N_M+1)}(\mathbf{r}, \mathbf{x}^M)}{\rho^{(N_M)}(\mathbf{x}^M)} \tag{3.32}$$

and $\Omega \subset \mathbb{R}^3$ the domain of the system. In this derivation, we used the definition of the reduced probability (2.34) for the integral (3.29) as well as the normalization C_{MS}, which is actually a function of the solute coordinates.

Similarly, we can compute the energy of the PMF. But unlike in the case of the forces we now have

$$V^{PMF}(\mathbf{x}^M) \neq \langle V(\mathbf{x}^M, \mathbf{x}^S) \rangle_{(\mathbf{x}^M)}.$$

However, the free energy of a solute at infinite dilution can also be defined as the reversible work which is necessary to take the solute from vacuum into the solvent by a step by step process [104]. Formally, we introduce a coupling parameter $\lambda \in [0, 1]$, which switches the solute-solvent interactions on or off, i.e., we write

$$V_{MS}(\mathbf{x}^M, \mathbf{x}^S) = V_{MS}(\mathbf{x}^M, \mathbf{x}^S; \lambda), \tag{3.33}$$

where $\lambda = 0$ corresponds to a non-interacting reference system and $\lambda = 1$ to the fully interacting system. Then, the reversible work is defined as

$$W(\mathbf{x}^M) = \int_0^1 \left\langle \frac{\partial V_{MS}(\mathbf{x}^M, \mathbf{x}^S; \lambda')}{\partial \lambda'} \right\rangle_{(\mathbf{x}^M, \lambda')} d\lambda', \tag{3.34}$$

where the average is defined as in (3.29) with the parameter λ in V_{MS} set to λ'. If we now again assume that the solute-solvent interaction potential can be written as a sum over pairwise terms,

$$V_{MS}(\mathbf{x}^M, \mathbf{x}^S; \lambda) = \sum_{i=1}^{N_M} \sum_{j=1}^{N_S} v_{MS}(\mathbf{x}_i^M, \mathbf{x}_j^S; \lambda),$$

we can transform the average in (3.34) as

$$\left\langle \frac{\partial V_{MS}}{\partial \lambda} \right\rangle_{(\mathbf{x}^M, \lambda)} = \left\langle \frac{\partial}{\partial \lambda} \sum_{i=1}^{N_M} \sum_{j=1}^{N_S} v_{MS}(\mathbf{x}_i^M, \mathbf{x}_j^S; \lambda) \right\rangle_{(\mathbf{x}^M, \lambda)} \tag{3.35}$$

$$= \sum_{i=1}^{N_M} \int_\Omega \left\langle \sum_{j=1}^{N_S} \delta(\mathbf{r} - \mathbf{x}_j^S) \right\rangle_{(\mathbf{x}^M, \lambda)} \frac{\partial v_{MS}(\mathbf{x}_i^M, \mathbf{r}; \lambda)}{\partial \lambda} \, d\mathbf{r} \tag{3.36}$$

$$= \sum_{i=1}^{N_M} \int_\Omega \langle \rho(\mathbf{r}) \rangle_{(\mathbf{x}^M, \lambda)} \frac{\partial v_{MS}(\mathbf{x}_i^M, \mathbf{r}; \lambda)}{\partial \lambda} \, d\mathbf{r}, \tag{3.37}$$

where $\langle \rho(\mathbf{r}) \rangle_{(\mathbf{x}^M, \lambda)}$ is the average solvent density at position $\mathbf{r}$ with the solute in fixed position $\mathbf{x}^M$ and the coupling parameter set to λ. By inserting this back into (3.34), we can compute the PMF by solving an integral over a single vector $\mathbf{r}$ and the parameter λ

$$W(\mathbf{x}^M) = \sum_{i=1}^{N_M} \int_0^1 \int_\Omega \langle \rho(\mathbf{r}) \rangle_{(\mathbf{x}^M, \lambda)} \frac{\partial v_{MS}(\mathbf{x}_i^M, \mathbf{r}; \lambda)}{\partial \lambda} \, d\mathbf{r} \, d\lambda. \tag{3.38}$$

The average solvent density can be identified with the conditional probability from (3.31)

$$\langle \rho(\mathbf{r}) \rangle_{(\mathbf{x}^M)} = \left\langle \sum_{j=1}^{N_S} \delta(\mathbf{r} - \mathbf{x}_j^S) \right\rangle_{(\mathbf{x}^M)} = \rho^{(N_M+1)}(\mathbf{r}|\mathbf{x}^M). \tag{3.39}$$

This relation can again easily be verified by inserting the definition of the reduced probability (2.34) into (3.39).

Now, we can compute the energy and the forces of the PMF by low-dimensional integrals if we know the reduced conditional probability

$$\rho^{(N_M+1)}(\mathbf{r}|\mathbf{x}^M) = \rho g^{(N_M+1)}(\mathbf{r}|\mathbf{x}^M). \tag{3.40}$$

We have already seen in Chapter 2 that the liquid state integral equation theories are able to compute reduced distribution functions of pure liquids. The situation in solute-solvent systems is slightly different. Here, the configuration of the solute is given and we are interested in the density of the solvent around that solute configuration. The integral equation theories can be extended to such situations. Many authors have developed methods for the computation of solvent densities based on the Ornstein-Zernike equation. These methods will be presented in the following. Further, we will derive our new method based on the YBG-hierarchy.

Chapter 4
Approximation of Solvent Densities

As we have seen in the previous chapter, the knowledge of the solvent density around the solute is sufficient to compute the potential of mean force in the case of pair potentials. The solvent density is associated with a reduced distribution function as described in Section 3.3. These reduced distribution functions are given as integrals over the probability distribution of the canonical ensemble. Such high-dimensional integrals cannot be computed in appropriate time with today's methods. They can be approximated by Monte Carlo integration or molecular dynamics simulations, which sample the whole phase space and, hence, are extremely time-consuming. No other method is known to be applicable to such integrals due to the high-dimensionality of phase space.

Nevertheless, in the past decades approximative methods have been developed to compute the pair distribution function of pure liquids. These liquid state integral equation theories, already presented in Chapter 2, are either based on the YBG-hierarchy or the Ornstein-Zernike equation. The advantages and disadvantages of the different methods concerning the computation of pair distribution functions of pure liquids are well understood, see Section 2.4.5. However, the extension to the treatment of solute-solvent systems is focused on methods based on the Ornstein-Zernike equation. Several authors, as e.g. Beglov and Roux [10–12, 28], Chandler et al. [20, 21], Kovalenko and Hirata [61, 65], Pettitt and Karplus [85, 86] or Richardi, Fries et al. [37, 99, 100] and many more have developed methods to compute the solvent density based on the Ornstein-Zernike equation and related theories. But to our knowledge, no such extension exists based on the YBG-hierarchy. This is why we investigate the value of the YBG-hierarchy concerning the computation of solvent densities.

In this chapter, we will describe the extension of the liquid state integral equation theories to the computation of solvent densities around solutes of arbitrary shape. To this end, we will consider simple monoatomic fluids as well as complex molecular solvents. We will first discuss Ornstein-Zernike based methods known from the

literature before we will derive our BGY3d equation for monoatomic solvents and the new molecular BGY3d equation (BGY3dM), both based on the YBG-hierarchy.

4.1 Definition of the Solvent Density

In order to understand, how the reduced distribution functions and the solvent density are connected, we first consider as a simple example a homogeneous monoatomic fluid for which we know the pair distribution function $g^{(2)}$. We assume that one atom is held in fixed position $\mathbf{x}_1^M$. We then know the probability to find another atom at position $\mathbf{x}$. The probability is simply given by the conditional probability $g^{(2)}(\mathbf{x}|\mathbf{x}_1^M)$, where

$$g^{(2)}(\mathbf{x}|\mathbf{x}_1^M) = \frac{g^{(2)}(\mathbf{x}, \mathbf{x}_1^M)}{g^{(1)}(\mathbf{x}_1^M)} \tag{4.1}$$

and we have $g^{(1)}(\mathbf{x}_1^M) = 1$ for a homogeneous fluid. We obtain the average density of particles around the fixed particle at $\mathbf{x}_1^M$ by multiplying with the overall density ρ of the solvent:

$$\rho^S(\mathbf{x}) = \rho g^{(2)}(\mathbf{x}|\mathbf{x}_1^M) \tag{4.2}$$

with

$$\rho^S(\mathbf{x}) := \langle \rho(\mathbf{x}) \rangle .$$

Hence, the knowledge of the pair distribution of a homogeneous fluid is equivalent to the knowledge of the average fluid density around a single fixed particle, which we hereafter identify with the solute. If the solute consists of more than one, let us say N_M, particles, the situation becomes more difficult. We now need to know the probability to find a solvent particle in position $\mathbf{x}$ with the solute atoms being fixed at positions $\mathbf{x}_1^M, \ldots, \mathbf{x}_{N_M}^M$. That is, we need to know the N_M+1-particle distribution function and it holds

$$\rho^S(\mathbf{x}) = \rho g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M) \tag{4.3}$$

with the conditional probability

$$g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M) = \frac{g^{(N_M+1)}(\mathbf{x}, \mathbf{x}^M)}{g^{(N_M)}(\mathbf{x}^M)}.$$

Hence, the methods that will be described in the next sections try to approximate the conditional (N_M+1)-particle distribution function $g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M)$. Note that all Ornstein-Zernike based methods indeed approximate the total correlation function $h = g^{(2)} - 1$ which is assumed to be disturbed by an external potential exerted by the solute as we will see in the following sections.

The above discussion considered all particles, solute as well as solvent particles, to be identical. This is of course not a reasonable assumption. The extension to more realistic systems, where the solute consists of different particle types, is

straight-forward and has only an effect on the potential functions and forces between solute and solvent particles. Further extensions to systems with molecular solvents, however, need more involved considerations and will be discussed in Sections 4.2.2 and 4.3.3.

4.2 Ornstein-Zernike based Methods

A lot of effort has been put into the application of the Ornstein-Zernike equation to solute-solvent systems by several authors. The standard equations of the liquid state integral equation theory can be applied to monoatomic solvents around a monoatomic solute, as we have seen above. The extension to more complex solutes with arbitrary shapes leads to the so-called 3d-HNC [10,53] or 3d-PY [53] methods. The molecular Ornstein-Zernike equations (MOZ) [14, 15, 38] allow to compute distribution functions of molecular solvents, as e.g. water, taking also into account the rotational degrees of freedom. As a reduction of the full angular-dependent distribution functions, the so-called radial site-site distribution functions of molecular solvents can be computed more easily. The reference interaction site model (RISM) [20] has been developed as the first theory for site-site correlation functions. A different approach is based on the density functional theory (DFT) [21, 27], which is also able to compute site-site distribution functions. However, the ultimate goal is to combine molecular solvents with solutes of arbitrary shape. Efforts in this direction have been made by combining the RISM model with the 3d-HNC method [12,28,61], by a reduction of the MOZ equations to three-dimensional space [24] and by keeping the full three-dimensional dependence of the DFT approach [12]. In the following, we will try to arrange the development of these methods and related works to give a whole picture of existing methods concerning the computation of solvent densities in solute-solvent systems.

4.2.1 Solutes in Monoatomic Solvents

The extension of the standard Ornstein-Zernike model (2.56) together with the HNC (2.57) or PY (2.59) approximations to solutes of arbitrary shape is straight-forward. Instead of reducing the equations to the spherical symmetric case, the full three-dimensional dependence is kept in equations (2.56), (2.57) and (2.59). Ikeguchi and Doi [53] computed the average distribution of a monoatomic water model by a Picard iteration of the transformed Ornstein-Zernike equation in Fourier space

$$\hat{\gamma}(\mathbf{k}) = \frac{\rho\hat{c}(\mathbf{k})^2}{1-\rho\hat{c}(\mathbf{k})} \tag{4.4}$$

with $\gamma(\mathbf{x}) = h(\mathbf{x}) - c(\mathbf{x})$ and either the HNC

$$c(\mathbf{x}) = e^{-\beta V(\mathbf{x})+\gamma(\mathbf{x})} - \gamma(\mathbf{x}) - 1 \tag{4.5}$$

or the PY closure

$$c(\mathbf{x}) = e^{-\beta V(\mathbf{x})}(1 + \gamma(\mathbf{x})) - \gamma(\mathbf{x}) - 1. \tag{4.6}$$

For details about the transformation of the Ornstein-Zernike equation, see Appendix B. Here, $V(\mathbf{x})$ is the total disturbing potential that acts on a solvent particle due to the solute,

$$V(\mathbf{x}) = \sum_{i=1}^{N_M} v(\mathbf{x}_i^M, \mathbf{x}). \tag{4.7}$$

The results are in qualitative agreement with the water density measured by a molecular dynamics simulation. However, since a monoatomic water model is used, no information about specific densities of the oxygen and hydrogen atoms can be gained.

Beglov and Roux [10] used a slightly different method to compute the average density of Lennard-Jones particles around a solute. Combination of the HNC closure and the Ornstein-Zernike equation leads to

$$h(\mathbf{x}) = e^{-\beta V(\mathbf{x}) + \rho(h*c)(\mathbf{x})} - 1, \tag{4.8}$$

where the asterix $*$ stands for the convolution

$$(h * c)(\mathbf{x}) := \int_\Omega h(\mathbf{x} - \mathbf{x}')c(\mathbf{x}')\ d\mathbf{x}', \tag{4.9}$$

and $\Omega \subseteq \mathbb{R}^3$ is the domain of the system. Multiplication with the bulk density ρ and insertion of the deviation from the bulk density $\Delta\rho(\mathbf{x}) := \rho^S(\mathbf{x}) - \rho$ yields the 3d-HNC equation

$$\Delta\rho^{k+1}(\mathbf{x}) = \lambda\rho\left(e^{-\beta V(\mathbf{x}) + (\Delta\rho^k * c)(\mathbf{x})} - 1\right) + (1 - \lambda)\Delta\rho^k(\mathbf{x}) \tag{4.10}$$

with the mixing factor $\lambda \in [0, 1]$. Here, the potential $V(\mathbf{x})$ is again defined as the total disturbing potential due to the solute as in (4.7). But unlike in the method of Ikeguchi and Doi, the direct correlation function $c(\mathbf{x})$ appears as an input to the method and is computed for the pure undisturbed fluid with the standard HNC approximation (2.57). Beglov and Roux tested the method against densities computed with molecular dynamics, the RISM method (which will be introduced in Section 4.2.2) and what they call the superposition approximation (SA)

$$\rho^{SA}(\mathbf{x}) = \rho \prod_{i=1}^{N_M} g^{(2)}(\mathbf{x}_i^M, \mathbf{x}) \tag{4.11}$$

with $g^{(2)}$ computed for the homogeneous fluid. They compared the numbers of the closest solvent particles and the height of the first maximum of the density profile.

The results of the 3d-HNC method were in good agreement with the molecular dynamics simulations when compared to the RISM and SA approximations. However, the height of the first maximum is always overestimated by 3d-HNC. Nevertheless, the method is able to reproduce fairly accurately the magnitude and the position of the density peaks around non-spherical solutes.

A first attempt to compute the spatial distribution of a more complex solvent around a non-spherical solute was made by Beglov and Roux [11] by employing the mean spherical approximation (MSA) for a water model. To this end, a water molecule is represented by a non-polarizable hard sphere with an embedded dipole at its center. The direct correlation function for pure water can be calculated analytically for the MSA model of water. A system of equations for the solvent density and the solvent polarization was then solved by an iterative scheme. The method reduces to the well-known equations of macroscopic electrostatics in the limit of infinitely small spheres. Although the 3d-MSA method yields qualitative promising results, it is quantitatively not very accurate. This is partly due to the linearized theory of polarization which has been employed.

4.2.2 Molecular Solvents

Since most fluids are not satisfyingly described by monoatomic models, the extension of the integral equations to molecular solvents is mandatory. One possible route is to describe the molecules of the liquid as rigid bodies with three translational and three rotational degrees of freedom, which we denote by $\mathbf{x}$ and $\mathbf{\Theta}$, respectively. The pair distribution function g, the total correlation function h and the direct correlation function c are functions of the translational as well as of the rotational degrees of freedom,

$$g = g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2), \qquad h = h(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2), \qquad c = c(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2). \quad (4.12)$$

Note that the functions indeed only depend on the relative positions of two molecules although the parametrization suggests a full 12-dimensional dependence. The generalization of the Ornstein-Zernike equation to molecular solvents then reads as

$$\begin{aligned} h(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2) &= c(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2) \\ &\quad + \int_\Omega \int_{\Omega_{\mathbf{\Theta}}} c(\mathbf{x}_1, \mathbf{x}_3, \mathbf{\Theta}_1, \mathbf{\Theta}_3) h(\mathbf{x}_2, \mathbf{x}_3, \mathbf{\Theta}_2, \mathbf{\Theta}_3)\, d\mathbf{x}_3 d\mathbf{\Theta}_3, \end{aligned} \quad (4.13)$$

see [45] for details. The set $\Omega \subseteq \mathbb{R}^3$ describes the domain of the system and $\Omega_{\mathbf{\Theta}}$ contains all possible orientations of a molecule. In most cases, the rotational degrees of freedom are represented by Euler-angles [37, 38, 45, 99]. For this special case we have

$$\Omega_{\mathbf{\Theta}} = \{(\chi, \theta, \phi) |\ \chi \in [0, 2\pi],\ \theta \in [0, \pi],\ \phi \in [0, 2\pi]\}\,. \quad (4.14)$$

All closure relations given in Section 2.4.4 are still applicable for the case of molecular fluids with the total correlation function h and the direct correlation function c now also depending on the rotational degrees of freedom. However, the computational effort to solve the molecular Ornstein-Zernike equation (4.13) together with an appropriate closure relation is much greater than in the case of a monoatomic fluid. All functions appearing in (4.13) and the volume of integration are six-dimensional. Hence, further approximations have to be used in order to be able to solve these equations.

One important idea, which goes back to the work of Blum and Torruella [14,15], is to write all functions as expansions of so-called rotational invariances,

$$h(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2) = \sum_i h_i(|\mathbf{x}_1 - \mathbf{x}_2|)\Phi_i(\mathbf{\Theta}_1, \mathbf{\Theta}_2, \bar{\mathbf{r}}) \tag{4.15}$$

with $\bar{\mathbf{r}} = \frac{\mathbf{x}_1 - \mathbf{x}_2}{|\mathbf{x}_1 - \mathbf{x}_2|}$. The functions h_i only depend on the distance of molecule 1 and 2, whereas the dependence on all rotational degrees of freedom has been shifted to the fixed set of Φ_i. The Φ_i are typically a linear combination of spherical harmonics. Their exact definition can be found e.g. in [45]. It is generally possible to solve the resulting system of equations for h_i and c_i by inserting these expansions into the Ornstein-Zernike equation and into the HNC closure relation. The exact derivation of these MOZ-HNC equations is very involved and will not be given here, see [38] for details. The infinite sum of (4.15) is truncated after some $i_{max} < \infty$ for practical use. Fries and Patey [38] observed that, for a dipolar hard sphere fluid, $i_{max} \approx 16$ is sufficient to reproduce the radial distribution function, the internal energy and the dielectric constant very accurately. The method has also been applied to liquid acetone and chloroform [100] and a non-aqueous electrolyte solution [37]. It is shown that the results for liquid structure and dielectric constants are in good agreement with values obtained by Monte Carlo simulations. However, the application to polar water models [99] showed that the MOZ-HNC model is not able to predict the O-H distribution functions satisfactorily. The authors conclude that the HNC approximation is not appropriate to describe the structure of H-bonded liquids as water. Nevertheless, the MOZ-HNC model is the only computational method that is able to reproduce the rotational dependency of the distribution functions.

Interaction Site Models

If the interaction of the molecules is described by a so-called site-site potential, the computation of many properties of interest only requires the knowledge of the site-site distribution functions $g_{\alpha\gamma}$. In this terminology, the sites are the atoms which constitute the molecules. Hence, $g_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma)$ is the distribution function between site α of a molecule 1 and site γ of molecule 2. The link between the site-site and the molecular distribution functions is given by integrating the molecular distribution

function over all degrees of freedom subject to the constraint that site α and site γ have a separation of $\mathbf{r}^{\alpha\gamma}$,

$$g_{\alpha\gamma}(\mathbf{r}^{\alpha\gamma}) = \int_{\Omega}\int_{\Omega}\int_{\Omega_{\mathbf{\Theta}}}\int_{\Omega_{\mathbf{\Theta}}} g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2) \times \delta(\mathbf{x}_1 - \mathbf{l}_1^{\alpha}(\mathbf{\Theta}_1) - \mathbf{x}_2 + \mathbf{l}_2^{\gamma}(\mathbf{\Theta}_2) - \mathbf{r}^{\alpha\gamma})\, d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{\Theta}_1 d\mathbf{\Theta}_2. \tag{4.16}$$

Here, $\mathbf{l}_1^{\alpha}(\mathbf{\Theta}_1)$ denotes the (orientation-dependent) vector displacement of site α in molecule 1. For molecules which are composed of s sites, there are obviously $\frac{s(s+1)}{2}$ site-site distribution functions. Some of these functions can be identical if the sites are identical, as e.g. the hydrogen sites in a H_2O molecule. But the complete set of site-site distribution functions contains less information than the molecular distribution function. Hence, $g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2)$ cannot be reconstructed from the $g_{\alpha\gamma}(\mathbf{r}^{\alpha\gamma})$. This is intuitive, since the site-site distribution function $g_{\alpha\gamma}(\mathbf{r}^{\alpha\gamma})$ depends on less degrees of freedom than $g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2)$.

The interaction site model (ISM) is a theory designed to compute the site-site correlation functions for the case, where the intermolecular pair potential is modeled by the site-site form, i.e.,

$$v(\mathbf{x}_1, \mathbf{x}_2, \mathbf{\Theta}_1, \mathbf{\Theta}_2) = \sum_{\alpha,\gamma} v_{\alpha\gamma}(|\mathbf{x}_1^{\alpha} - \mathbf{x}_2^{\gamma}|) \tag{4.17}$$

with the position $\mathbf{x}_1^{\alpha}$ of site α in molecule 1. The key ingredient of the ISM theory is an Ornstein-Zernike-like equation for the site-site correlation functions

$$h_{\alpha\gamma} = \sum_{\alpha',\gamma'} \omega_{\alpha\alpha'} * c_{\alpha'\gamma'} * \omega_{\gamma'\gamma} + \rho \sum_{\alpha',\gamma'} \omega_{\alpha\alpha'} * c_{\alpha'\gamma'} * h_{\gamma'\gamma}. \tag{4.18}$$

This can be written more conveniently in Fourier space as

$$\hat{H} = \hat{W}\hat{C}\hat{W} + \rho\hat{W}\hat{C}\hat{H} \tag{4.19}$$

with the capital letters indicating matrices with the site-site functions as entries, i.e., $(\hat{H})_{\alpha\gamma} = \hat{h}_{\alpha\gamma}$, $(\hat{W})_{\alpha\gamma} = \hat{\omega}_{\alpha\gamma}$ and $(\hat{C})_{\alpha\gamma} = \hat{c}_{\alpha\gamma}$. The hat indicates a function in Fourier space. Here, $\omega_{\alpha\gamma}$ is the intramolecular correlation function

$$\omega_{\alpha\gamma}(\mathbf{r}) = \frac{1}{4\pi(r_0^{\alpha\gamma})^2}\delta(|\mathbf{r}| - r_0^{\alpha\gamma}) \tag{4.20}$$

with $r_0^{\alpha\gamma}$ the intramolecular distance between sites α and γ. For $\alpha = \gamma$, we set $r_0^{\alpha\gamma} = 0$ for completeness. The intramolecular correlation functions describe the constraints which fully determine the structure of the rigid molecule. Equation (4.18) is called site-site Ornstein-Zernike (SSOZ) equation and was first introduced by Chandler and Andersen [20]. A detailed derivation can be found in [25] or

[45]. It is formally exact, since it can be understood as definition of the site-site direct correlation functions $c_{\alpha\gamma}$. Supplemented with a closure relation, the SSOZ equation can be solved for the site-site distribution functions $g_{\alpha\gamma} = h_{\alpha\gamma} + 1$. The closure relations can easily be transferred from the monoatomic case by replacing the correlation functions by their site-site analogs, giving one relation for every site-site pair.

Early applications of the ISM model have mainly focused on the so-called fused hard sphere model, where $v_{\alpha\gamma}(r) = \infty$ for $r \leq \sigma$ and $v_{\alpha\gamma}(r) = 0$ for $r > \sigma$, with σ the hard sphere radius, see [45] for an overview. The application of the ISM model to a hard sphere fluid is referred to as the reference interaction site model (RISM[1]). The results for the site-site distribution functions are in qualitative agreement with those from Monte Carlo simulations. The accuracy however can be low. The deviation for important features is sometimes up ot 20% [45]. When applied to other short range potential functions as the Lennard-Jones potential, the results are of similar quality. Therefore, the RISM model is called a qualitative theory [45].

In order to incorporate long range effects like the Coulomb potential, some modifications have to be introduced. This is mainly necessary because it is very difficult to determine the correct boundary conditions of a finite domain in this case. Hirata and Rossky [50] proposed a method, which they call the extended RISM (XRISM). It is based on the assumption that, for most fluids, the direct correlation function $c(r)$ behaves as $c(r) \to -\beta v(r)$ for $r \to \infty$. Hence, we replace $c(r) \to c(r) - \phi(r)$ with $\phi(r) = -\beta v(r)$, and the SSOZ equation (4.18) can be written in renormalized form, see [50] for details. By this procedure, the long-range Coulomb potential does not appear explicitly in the renormalized SSOZ equations. The XRISM model has been applied to several pure liquids, as e.g. a simplified diatomic liquid with charged sites [49, 50], liquid methanol [88] and several water models [87]. All results show a good qualitative agreement of the site-site correlation functions with those obtained by Monte Carlo simulations. The coordination numbers as well as the positions of structural features are well represented. However, as already stated above, the accuracy is not satisfactory, especially regarding the magnitudes of the peaks of the correlation functions. Moreover, the RISM model, together with a closure that obeys $c(r) \to -\beta v(r)$ for large r, is not able to correctly predict the relative permittivity of the continuum [25, 45]. However, this deficiency can easily be eliminated by replacing the known XRISM dielectric constant by a phenomenological one [51]. Finally, Perkyns and Pettitt [83] introduced the dielectrically consistent reference interaction site model (DRISM), where a modification of the bridge function in the HNC closure leads to a corrected dielectric constant of the solvent.

[1]Most authors do not distinguish between ISM and RISM. Therefore the abbreviation RISM is mostly found in the literature, although no hard sphere (reference) model is considered. In order to agree with the literature, we will also use RISM hereafter.

4.2.3 Solutes in Molecular Solvents

The extension of the RISM model to solutes surrounded by molecular solvents was first given by Hirata, Rossky and Pettitt [51]. The derivation starts with the RISM integral equation for a mixture of general composition defined by

$$\varrho H \varrho = U * C * U + U * C * \varrho H \varrho. \tag{4.21}$$

Here, the $*$ of two matrices stands for a matrix convolution product

$$(U * C)_{\alpha\gamma} = \sum_{\delta} U_{\alpha\delta} * C_{\delta\gamma}. \tag{4.22}$$

Each factor of (4.21) is a $\mathcal{S} \times \mathcal{S}$ matrix, where $\mathcal{S}$ is the total number of molecular sites in the mixture, i.e., if s_i is the number of sites in molecule i, we have

$$\mathcal{S} = \sum_{i=1}^{\mu} s_i \tag{4.23}$$

for μ different species. If we restrict ourselves to two different species, the solute M and the solvent S, the matrices can be written as

$$\begin{aligned} \varrho &= \begin{pmatrix} \rho^S & 0 \\ 0 & \rho^M \end{pmatrix}, \qquad U = \begin{pmatrix} U^S & 0 \\ 0 & U^M \end{pmatrix}, \\ H &= \begin{pmatrix} H^{SS} & H^{SM} \\ H^{MS} & H^{MM} \end{pmatrix}, \qquad C = \begin{pmatrix} C^{SS} & C^{SM} \\ C^{MS} & C^{MM} \end{pmatrix}. \end{aligned} \tag{4.24}$$

To this end, ρ^S and ρ^M are the number densities of the solvent and the solute, respectively. The submatrices are labeled to indicate the species which are included, i.e., the submatrix H^{SM} contains entries $h_{S\alpha M\gamma}$ which denote the total correlation functions between site α of the solvent and site γ of the solute. The intramolecular correlations are contained in U. Since there are no intramolecular correlations between solvent and solute, the entries U^{SM} and U^{MS} are zero. The link between the matrices W^S and W^M with the entries as defined in (4.20) is given by $W^S = \rho^S U^S$ and $W^M = \rho^M U^M$, respectively.

Using (4.24) in equation (4.21) leads to a system of equations

$$\begin{aligned} \rho^S H^{SS} \rho^S &= U^S * C^{SS} * U^S + U^S * C^{SS} * \rho^S H^{SS} \rho^S \\ &\quad + U^S * C^{SM} * \rho^M H^{MS} \rho^S, \qquad (4.25) \\ \rho^M H^{MS} \rho^S &= U^M * C^{MS} * U^S + U^M * C^{MS} * \rho^S H^{SS} \rho^S \\ &\quad + U^M * C^{MM} * \rho^M H^{MS} \rho^S, \qquad (4.26) \\ \rho^M H^{MM} \rho^M &= U^M * C^{MM} * U^M + U^M * C^{MS} * \rho^S H^{SM} \rho^M \\ &\quad + U^M * C^{MM} * \rho^M H^{MM} \rho^M, \qquad (4.27) \end{aligned}$$

where the fourth equation for $\rho^S H^{SM} \rho^M$ can be left out since $H^{SM} = (H^{MS})^T$. Now, we consider the case where the solute is in infinite dilution, i.e. $\rho^M \to 0$. Then, the third equation (4.26) can be neglected in comparison to the others and we get

$$H^{SS} = W^S * C^{SS} * W^S + W^S * C^{SS} * \rho^S H^{SS}, \tag{4.28}$$

$$H^{MS} = W^M * C^{MS} * W^S + W^M * C^{MS} * \rho^S H^{SS}, \tag{4.29}$$

$$H^{MM} = W^M * C^{MM} * W^M + W^M * C^{MS} * \rho^S H^{SM}. \tag{4.30}$$

Equation (4.28) can be solved separately, if endowed with an appropriate closure relation, since it is an equation of solvent correlation functions alone. Equation (4.29) can then be solved to give the solute-solvent site-site correlation functions and equation (4.30) determines the solute site-site correlation functions. The closure relations can again simply be transferred from the monoatomic case, since they only relate correlation functions of the same site-site pair.

If the superposition approximation is used in order to compute the potential of mean force (PMF) by the solute site-site distribution functions of the RISM/XRISM equations, equation (3.27) of Section 3.3 leads directly to a term for the solvent mediated PMF between sites α and γ of the solute,

$$V_{\alpha\gamma}^{PMF}(r_{\alpha\gamma}) = -\frac{1}{\beta} \ln(h_{\alpha\gamma}(r_{\alpha\gamma}) + 1) \tag{4.31}$$

with $r_{\alpha\gamma} = |\mathbf{x}_\alpha^M - \mathbf{x}_\gamma^M|$. This is a very crude approximation of the PMF, since (4.29) and (4.30) do not take into account the spatial distribution of the solute. This approximation deteriorates significantly with increasing number of solute sites.

Nevertheless, the solute-solvent RISM/XRISM equations have been applied to a wide variety of solute-solvent systems. Among these are ions in a simplified diatomic polar solvent [51], polar diatomic solutes in water [84], *N*-butane and 1,2-dichloroethane in water and CCl_4 [130], *N*-methylalanylacetamide in water [86], alkali halides in water [89] and alanine dipeptide in water [85]. The quantitative accuracy of the results for these systems is difficult to assess, but it is expected that the method yields to reasonable results, comparable to those obtained by RISM for pure molecular fluids. In [92], the DRISM model is used for computing the PMF of alanine tetrapeptide during a molecular dynamics simulation. Since the solute-solvent RISM equations do not take into account the spatial distribution of the solute, all site-site distribution functions can be computed once in advance and stored for fast access. This, together with the fast computation of the PMF by equation (4.31), make the procedure computationally efficient. The results have been compared to simple implicit solvent models with a constant and with a linear distance dependent dielectric constant. When compared to a simulation with explicit water molecules, the DRISM method showed to be superior to both implicit solvent models. This is not very surprising since it also incorporates short-range effects such as solvent packing. Nevertheless, the accuracy of the computed properties is very inaccurate.

Density Functional Theory

A different approach for the computation of site density distributions is based on the so-called density functional theory (DFT). The DFT for nonuniform fluids is based on the fact that the free energy F of a fluid is a function of the density $F = F(\rho(\mathbf{x}))$ and has its minimum for the mean density of the fluid $\langle\rho(\mathbf{x})\rangle$. This formulation goes back to the work of Lebowitz and Percus [68] and has been used for the study of interfacial phenomena, mean field treatments of first order phase transitions, the analysis of wetting transitions, the study of non-periodic crystals and more, see [21] and the references therein. Chandler et al. [21] developed the general formulation for the application of the DFT to nonuniform polyatomic systems. Since the exact expression for the free energy cannot be computed, approximations have to be introduced. It can be shown that, for a specific approximation in the free energy functional, the RISM equations are recovered [21].

Donley et al. [27] developed a DFT approach similar to that of Chandler et al. [21]. To this end, the site-site pair distribution functions are expressed as an average over the coordinates of two molecules in an effective potential. This effective potential can be computed by RISM-like equations, see [27] for details. But in contrast to the RISM theory, the DFT model reproduces the exact site-site distribution functions in the limit of low density. Donley et al. applied their model to a simple diatomic fluid which led to a good agreement with the results of molecular dynamics simulations. Reddy et al. [93] tested the DFT approach of Donley et al. against two different integral equation theories for the modified simple point charge (SPC) model of water. In this comparison, the DFT method was superior to the other methods. The application to the standard SPC model of water showed however qualitative differences in the site-site distribution functions due to the missing repulsive potential between the oxygen and the hydrogen sites. Sumi et al. [114] employed the DFT ansatz to derive their site-density integral equation (SDIE). Here, a set of equations for the three-dimensional site-density functions is formulated as closure for the RISM equations. Application of the SDIE to a diatomic Lennard-Jones fluid yielded good agreement with a simulation at low densities [114]. The extension of the SDIE method to polymer fluids also led to accurate results compared to molecular dynamics simulations [113]. In [115], a more efficient implementation of the SDIE has been applied to Cl_2, HCl and water. The computed site-site pair distribution functions for Cl_2 and HCl again showed good qualitative agreement with simulation results. However, the O-O distribution function for water could not be reproduced correctly, as this is the case with any other integral equation method. A similar method has been developed by Yethiraj et al. [127] in the context of polymeric liquids modeled by hard sphere chains. Here, a different free-energy functional is used and the method is tested against Monte Carlo simulations with reasonable results for the pair distribution functions.

Three-Dimensional Density Distribution of Molecular Solvents

As already noted, the solute-solvent RISM equations do not contain any information about the spatial distribution of the solute. All site-site solute-solvent correlation functions are spherically symmetric. This is an inappropriate approximation especially for larger solute molecules, since it assumes that all solute sites are equally exposed to the solvent. Hence, the contribution to the solvent-mediated free energy of partially or completely buried solute atoms is clearly overestimated. To overcome this deficiency, several authors developed methods to include the full three-dimensional shape of the solutes.

Cortis et al. [24] derived a three-dimensional integral equation by averaging the molecular Ornstein-Zernike equation over the orientation of the solvent molecule consistent with one site of the solvent remaining at a fixed distance from the solute at the origin. The method is called the molecular origin-site Ornstein-Zernike integral equation (MSOZ). The resulting site distribution functions contain full three-dimensional information regarding the structure of the solute molecule. The equation is supplemented with several three-dimensional generalizations of the HNC closure. The MSOZ method has been applied to several pure non-polar and polar fluids. To this end, the site-site distribution functions have been obtained by angle averaging of the three-dimensional site distribution functions around a fixed solvent molecule. The accuracy of the spherical symmetric site-site distribution functions has been shown to be more accurate than those obtained by the RISM method.

Kovalenko and Hirata [61] introduced a new view on the solute-solvent RISM equations which allows to treat solutes of arbitrary shapes. Equation (4.28) is used to compute the site-site correlation functions for the pure solvent. The solute is seen as a single particle of arbitrary form. Therefore, the intramolecular correlation matrix of the solute W^M is simply a scalar of value one and equation (4.29) becomes

$$H^{MS} = C^{MS} * W^S + C^{MS} * \rho^S H^{SS} \tag{4.32}$$

or

$$h^{MS}_\gamma = \sum_\alpha c^{MS}_\alpha * \omega^S_{\alpha\gamma} + \sum_\alpha c^{MS}_\alpha * \rho^S h^S_{\alpha\gamma} \tag{4.33}$$

for all sites γ of the solvent. However, the interaction between solute and solvent is still described by a site-site potential, i.e., the total solute-solvent potential of solvent site γ is

$$V^{MS}_\gamma(\mathbf{x}) = \sum_{\alpha=1}^{N_M} v^{MS}_{\alpha\gamma}(\mathbf{x}^M_\alpha, \mathbf{x}). \tag{4.34}$$

Endorsed with the three-dimensional HNC closure (4.5) with V^{MS}_γ as above, equation (4.33) can be solved to give the full spatial correlation function of each solvent site. The method was applied to compute the spatial distribution of water around

a single water molecule and the density profile of water near a crystalline layer of Lennard-Jones sites [61]. Together with the partially linearized hypernetted chain approximation (PLHNC) [62], a modified closure to account for very strong electrostatic attractions, the 3d-RISM equations were employed to compute the PMF for the N,N-dimethylaniline cation and the anthracene anion in acetonitrile as solvent [62]. A comparison of the results for sodium chloride in water with results from a molecular dynamics simulation showed a good agreement [65, 66]. An improvement was introduced by applying an empirical bridge function in the 3d-HNC closure in order to counter the overestimation of water ordering around a hydrophobic solute [63]. This yields a drastically improved agreement with simulation data for the thermodynamics of hydration for rare gases and alkanes in water. The 3d-RISM-PLHNC method has also been applied to study met-enkephalin in water with reasonable results [64]. In [67], a so-called self-consistent three-dimensional reference interaction site model (SC-3d-RISM-HNC) is proposed for the special case of ionic solutes in a polar molecular solvent. To this end, the first equation of the RISM model for the pure solvent is replaced by equation (4.33), where one solvent atom is considered as the solute in fixed position. Hence, full three-dimensional resolution of the direct correlation function c_α^S is already obtained for the pure solvent. Inserted into the Ornstein-Zernike equation together with a three-dimensional HNC closure, this yields the total correlation function of the ion around the fixed solvent molecule. This concept is particular, since the roles of the solute and the solvent seem to be exchanged. But this is really a transformation of coordinates and therefore an equivalent representation of the total correlation function. The SC-3d-RISM-HNC model has been applied to Na^+ and Cl^- ions in water with an improved agreement of the ion hydration structure and thermodynamics when compared to the conventional RISM-HNC approach.

Beglov and Roux [12] use the DFT ansatz to derive an integral equation method for the computation of three-dimensional solvent densities around solutes of arbitrary shape. To this end, they use the free energy functional of Chandler et al. [21] truncated at second order of the intramolecular and the intermolecular correlations. Minimization of this functional with respect to the site density leads to an integral equation for the mean density which can also be written as a system of two coupled equations,

$$c_\gamma^{MS}(\mathbf{x}) = e^{-\beta V_\gamma^{MS}(\mathbf{x}) + h_\gamma^{MS}(\mathbf{x}) - c_\gamma^{MS}(\mathbf{x})} - h_\gamma^{MS}(\mathbf{x}) + c_\gamma^{MS}(\mathbf{x}) - 1, \tag{4.35}$$

$$\rho h_\gamma^{MS}(\mathbf{x}) = \sum_\alpha c_\alpha^{MS} * \chi_{\alpha\gamma}^S(\mathbf{x}) \tag{4.36}$$

with the susceptibility of the pure solvent

$$\chi_{\alpha\gamma}^S = \rho\omega_{\alpha\gamma}^S + \rho^2 h_{\alpha\gamma}^S. \tag{4.37}$$

The first equation (4.35) is indeed the HNC approximation which is recovered by

the DFT approach. The second equation introduces the intra- and intermolecular correlations of the solvent through the susceptibility function which appears as an input. It is computed by the standard RISM-HNC method as described above. This method is also referred to as 3d-RISM-HNC method. In [12], the equations (4.35) and (4.36) were used to compute the density of water around a single water molecule and *N*-methylacetamide. The solute-solvent site-site radial distribution functions were computed from the mean solvent densities and showed reasonable hydrogen bonding features. However, the O-H site-site distribution functions differs from the H-O site-site distribution function, which should not be the case in an exact theory due to symmetry. This is a deficiency of the method, which has already been noticed for the RISM equations. It is due to the inadequate incorporation of the intramolecular correlations which are only considered to second order. Hence, Du and coworkers [28] introduced a so-called hydrogen bridge function in order to incorporate the short-range hydrogen-oxygen intramolecular correlation to lowest order. This bridge function includes a free parameter which can be used to fit the results to experimental values. The susceptibility of pure water computed by molecular dynamics and the RISM-HNC model were combined in order to improve the short-range structures of the hydrogen-oxygen correlation functions. Then, the 3d-RISM-HNC equations were used to compute the free energy of some *N*-alkanes, *N*-alcohols, *N*-carboxylic acids and simple amides [28]. The computation of the free energies employed the thermodynamic integration, see equation (3.34) of Section 3.3. The comparison of the results with experimental data showed a qualitatively good agreement. The lack of accuracy is due to an overestimation of the water hydrogen density in the neighborhood of negatively charged groups and the overestimation of the pressure in the HNC closure. The authors propose to further improve the empirical bridge function by an extensive adjustment in order to give accurate results for a large training set of molecular solutes.

4.2.4 Summary

The integral equation theories based on the Ornstein-Zernike equation for computing the solvent density around a solute provide a powerful framework for the development of new approaches to incorporate solvent effects in molecular simulations. So far however, the methods still involve computations that are too complex in order to be used in simulations, where a repeated evaluation of the potential of mean force is required. On the other hand, the accuracy of the computation of thermodynamic properties such as the free energy needs to be improved. This can be achieved by adjusting free parameters, as for example in the empirical bridge functions, in order to fit the method to known (experimental) data. One finally concludes that the approximations that pose a compromise between accuracy and numerical tractability still have a substantial effect on the results. This is clearly undesirable but cannot

be avoided with today's theories. Hence, further research that is also concerned with new theories and concepts that have not been considered so far is necessary. This is why we focus on the YBG-hierarchy together with the Kirkwood approximation which, to our knowledge, have not been employed in the context of solute-solvent systems in the literature.

4.3 Methods based on the YBG-Hierarchy

Concerning the computation of solvent density distributions in solute-solvent systems, the literature focuses on methods that are somehow linked to the Ornstein-Zernike equation. This is quite surprising, since equations from the YBG-hierarchy (2.48) have already been used in the context of non-uniform fluids. In this case, the first equation of the hierarchy can be written as

$$\nabla \ln(\rho(\mathbf{x})) = \mathbf{F}^{ext}(\mathbf{x}) + \int_\Omega \mathbf{F}(\mathbf{x},\mathbf{x}')g^{(2)}(\mathbf{x},\mathbf{x}')\rho(\mathbf{x}')\ d\mathbf{x}' \tag{4.38}$$

with external force $\mathbf{F}^{ext}$. This relation can be used to compute the density $\rho(\mathbf{x})$ for a fluid which is disturbed by the external force. Examples of applications are the computation of the density profile in the vicinity of walls or free liquid surfaces, liquid-liquid interfaces or liquids in narrow slits, see [47] for an overview. It is important to note that, due to the non-uniform disturbance, the pair distribution function $g^{(2)}(\mathbf{x},\mathbf{x}')$ in this formulation does not only depend on the distance $|\mathbf{x}-\mathbf{x}'|$, but also on the absolute positions $\mathbf{x}$ and $\mathbf{x}'$. Hence, equation (4.38) is exact and a closure relation is again required in order to express $g^{(2)}$ by known functions. The easiest relation one can consider is to replace $g^{(2)}$ by the pair distribution function of the undisturbed fluid at some averaged density ρ. But more sophisticated closure relations are also known, see [47]. The application of equation (4.38) to a solvent disturbed by a solute seems to be obvious, but has never been considered so far. Therefore, we will investigate an equation which is very similar to (4.38) in this context. We call it BGY3d equation. Indeed, (4.38) and the BGY3d equation are identical if the pair distribution function of the undisturbed fluid in (4.38) is employed. But to see this, we will now derive the BGY3d equation in detail.

4.3.1 Derivation of the BGY3d Equation

We consider a system with a solute consisting of N_M atoms and N_S monoatomic solvent particles surrounding the solute in a box $\Omega = [0,L]^3$. The box has to be chosen large enough such that the effects of the solute atoms on the solvent can be neglected at the boundaries of the box. For now, we assume that all solute as well as all solvent particles are identical. We want to compute the solvent density for a fixed position $\mathbf{x}^M$ of the solute, where we write short $\mathbf{x}^M$ for $\mathbf{x}_1^M,\dots,\mathbf{x}_{N_M}^M$. This

density $\rho^S(\mathbf{x})$ is given by the N_M+1-particle distribution function with N_M particle positions fixed, i.e.

$$\rho^S(\mathbf{x}) = \rho g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M) \tag{4.39}$$

with the conditional probability

$$g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M) = \frac{g^{(N_M+1)}(\mathbf{x},\mathbf{x}^M)}{g^{(N_M)}(\mathbf{x}^M)}. \tag{4.40}$$

As we have seen in section 2.4.2, we can compute the N_M+1-distribution function by the corresponding equation from the YBG-hierarchy together with an appropriate closure relation. The equation to solve for the distribution function $g^{(N_M+1)}$ is given by

$$\begin{aligned}\nabla_{\mathbf{x}} g^{(N_M+1)}(\mathbf{x},\mathbf{x}^M) = \beta \sum_{i=1}^{N_M} \mathbf{F}(\mathbf{x},\mathbf{x}_i^M) g^{(N_M+1)}(\mathbf{x},\mathbf{x}^M) \\ + \beta\rho \int_\Omega \mathbf{F}(\mathbf{x},\mathbf{x}') g^{(N_M+2)}(\mathbf{x},\mathbf{x}',\mathbf{x}^M)\, d\mathbf{x}'.\end{aligned} \tag{4.41}$$

Hence, we need an approximation of $g^{(N_M+2)}(\mathbf{x},\mathbf{x}',\mathbf{x}^M)$ by lower order distribution functions in order to be able to solve (4.41).

We are going to consider the so-called n-level Kirkwood closure relations [108]. They relate the n-particle distribution function to distribution functions of lower order:

$$g^{(n)}(\mathbf{x}_1,\ldots,\mathbf{x}_n) \approx \prod_{k=2}^{n-1} \prod_{1\le i_1<i_2<\cdots<i_k\le n} g^{(k)}(\mathbf{x}_{i_1},\mathbf{x}_{i_2},\ldots,\mathbf{x}_{i_k})^{(-1)^{n-1-k}}. \tag{4.42}$$

Examples of (4.42) are the Kirkwood approximation (2.51) for $n=3$ and the Fisher-Kopeliovich closure (2.54) for $n=4$. Approximation of $g^{(N_M+2)}(\mathbf{x}_1,\ldots,\mathbf{x}_{N_M+2})$ by this closure gives

$$\begin{aligned} g^{(N_M+2)}(\mathbf{x}_1,\ldots,\mathbf{x}_{N_M+2}) \approx g^{(N_M+1)}(\mathbf{x}_1,\ldots,\mathbf{x}_{N_M+1}) g^{(N_M+1)}(\mathbf{x}_1,\ldots,\mathbf{x}_{N_M},\mathbf{x}_{N_M+2}) \\ \times \prod_{1\le i_1<i_2<\cdots<i_{N_M-1}\le N_M} g^{(N_M+1)}(\mathbf{x}_{i_1},\ldots,\mathbf{x}_{i_{N_M-1}},\mathbf{x}_{N_M+1},\mathbf{x}_{N_M+2}) \\ \times \prod_{k=2}^{N_M} \prod_{1\le i_1<i_2<\cdots<i_k\le N_M+2} g^{(k)}(\mathbf{x}_{i_1},\mathbf{x}_{i_2},\ldots,\mathbf{x}_{i_k})^{(-1)^{N_M+2-1-k}}. \end{aligned} \tag{4.43}$$

The first two terms of the right hand side of (4.43) are identical with the N_M+1-particle distribution functions of (4.41). All other terms are recursively further approximated by (4.42). This finally yields

$$\begin{aligned} g^{(N_M+2)}(\mathbf{x},\mathbf{x}',\mathbf{x}^M) \approx \\ \frac{g^{(N_M+1)}(\mathbf{x},\mathbf{x}^M)\; g^{(N_M+1)}(\mathbf{x}',\mathbf{x}^M)\; g^{(2)}(\mathbf{x},\mathbf{x}')}{g^{(N_M)}(\mathbf{x}_M)}, \end{aligned} \tag{4.44}$$

where we identified $\mathbf{x}_{N_M+1}$ and $\mathbf{x}_{N_M+2}$ with $\mathbf{x}$ and $\mathbf{x}'$ from equation (4.41), respectively. Inserting (4.44) into (4.41) leads to an integral equation for the $N_M + 1$-particle distribution function

$$\nabla_{\mathbf{x}} g^{(N_M+1)}(\mathbf{x}, \mathbf{x}^M) = \beta \sum_{i=1}^{N_M} \mathbf{F}(\mathbf{x}, \mathbf{x}_i^M) g^{(N_M+1)}(\mathbf{x}, \mathbf{x}^M) + \beta\rho \, \frac{g^{(N_M+1)}(\mathbf{x}, \mathbf{x}^M)}{g^{(N_M)}(\mathbf{x}_M)} \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}') g^{(2)}(\mathbf{x}, \mathbf{x}') \; g^{(N_M+1)}(\mathbf{x}', \mathbf{x}^M) \; d\mathbf{x}'. \tag{4.45}$$

Since we are looking for the probability to find a solvent particle at position $\mathbf{x}$ provided that the solute particles are at positions $\mathbf{x}^M$, we insert the conditional probability (4.40) into (4.45) and obtain

$$\nabla_{\mathbf{x}} g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M) = \beta \sum_{i=1}^{N_M} \mathbf{F}(\mathbf{x}, \mathbf{x}_i^M) g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M) + \beta\rho \; g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M) \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}') g^{(2)}(\mathbf{x}, \mathbf{x}') \; g^{(N_M+1)}(\mathbf{x}'|\mathbf{x}^M) \; d\mathbf{x}'. \tag{4.46}$$

We call (4.46) the BGY3d equation. The pair distribution function $g^{(2)}$ appears as an input in equation (4.46). It can be computed for example by the Born-Green equation (2.52), which can further be regarded as special case of the BGY3d equation. If the solute consists of one solvent particle, we have $g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M) = g^{(2)}(\mathbf{x}|\mathbf{x}^M)$ and equation (4.46) can be used to compute the pair distribution function $g^{(2)}$.

As stated above, the BGY3d equation is identical to equation (4.38) for a non-uniform fluid with a special approximation to $g^{(2)}$, if one regards the influence of the solute particles on the solvent as external force disturbing the solvent. To our knowledge, equation (4.38) has never been solved with full three-dimensional resolution of the density $\rho(\mathbf{x})$. Hence, all external forces $\mathbf{F}^{ext}$ considered so far in the literature result in symmetry properties that allow to use a simplified form of (4.38). This is not reasonable for our application, since we are going to consider solutes of arbitrary shape. Hence, we have to solve the BGY3d equation (4.46) in three dimensions. However, equation (4.46) exhibits some properties which are disadvantageous for its numerical treatment. This is why we introduce some transformations that allow for an efficient numerical solution on a three-dimensional grid.

4.3.2 Transformation of the BGY3d equation

In order to simplify matters, we first introduce some notation. Since we will always consider a system with N_M solute atoms, we will abbreviate $g(\mathbf{x}) = g^{(N_M+1)}(\mathbf{x}|\mathbf{x}^M)$,

$\mathbf{F}(\mathbf{x}, \mathbf{x}^M) = \sum_{i=1}^{N_M} \mathbf{F}(\mathbf{x}, \mathbf{x}_i^M)$ and $\nabla = \nabla_{\mathbf{x}}$ in the following. The BGY3d equation then reads as

$$\begin{aligned} \nabla g(\mathbf{x}) = \beta\ \mathbf{F}(\mathbf{x}, \mathbf{x}^M) g(\mathbf{x}) \\ + \beta\rho\ g(\mathbf{x}) \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}') g^{(2)}(\mathbf{x}, \mathbf{x}') g(\mathbf{x}')\ d\mathbf{x}'. \end{aligned} \tag{4.47}$$

First, we want to investigate the case of $\rho \to 0$, which can be interpreted as a system consisting of two single solvent particles around the solute. Equation (4.47) then becomes

$$\nabla g_0(\mathbf{x}) = \beta\ \mathbf{F}(\mathbf{x}, \mathbf{x}^M) g_0(\mathbf{x}). \tag{4.48}$$

This is a system of partial differential equations. If we remember that $\mathbf{F}(\mathbf{x}, \mathbf{x}^M) = -\nabla \sum_{i=1}^{N_M} v(\mathbf{x} - \mathbf{x}_i^M)$, we can easily give the analytic solution of equation (4.48) which we denote by

$$g_0(\mathbf{x}) = e^{-\beta \sum_{i=1}^{N_M} v(\mathbf{x} - \mathbf{x}_i^M)}. \tag{4.49}$$

This holds for any pair potential v, e.g., the Lennard-Jones or the Coulomb potential. An important point to note is that these potentials have a singularity for $|\mathbf{x} - \mathbf{x}_i^M| = 0$. This is reasonable from a physical point of view, since in classical mechanics two particles cannot be at the exact same position. This is why we postulate that the probability of this situation is zero and, indeed, we have $g_0(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_i^M} = 0$ for $i = 1, 2, \ldots, N_M$. This also holds for the solution of (4.47), because, for small values of $|\mathbf{x} - \mathbf{x}_i^M|$, the singular force term strongly dominates the right hand side. For this, the integral term is assumed to be bounded. This is not immediately clear due to the singularity of the force $\mathbf{F}(\mathbf{x}, \mathbf{x}')$ for $|\mathbf{x} - \mathbf{x}'| = 0$. But we know that the pair distribution function of the pure solvent behaves as

$$g^{(2)}(\mathbf{x}, \mathbf{x}') \to e^{-\beta v(\mathbf{x} - \mathbf{x}')} \quad \text{for} \quad \mathbf{x} \to \mathbf{x}'. \tag{4.50}$$

If we note that

$$x^{-a} e^{-x^{-b}} \to 0 \quad \text{for} \quad x \to 0 \tag{4.51}$$

holds for any $a, b > 0$, we can conclude that the product of the force and the pair distribution function vanishes for $|\mathbf{x} - \mathbf{x}'| = 0$ and all popular types of potential functions. Figure 4.1 illustrates the situation for the Lennard-Jones potential. The exponential decrease of the pair distribution function dominates the polynomial increase of the force, and all components of the product $\mathbf{F} g^{(2)}$ are continuous and $< \infty$ for every $\mathbf{x}$. Hence, the integral term of (4.47) is bounded.

In summary, we can conclude that the solution of the full problem (4.47) approaches the solution of the reduced problem (4.49) at the solute particle positions

$$g(\mathbf{x}) \to g_0(\mathbf{x}) \quad \text{for} \quad \mathbf{x} \to \mathbf{x}_i^M, \quad i = 1, 2, \ldots, N_M, \tag{4.52}$$

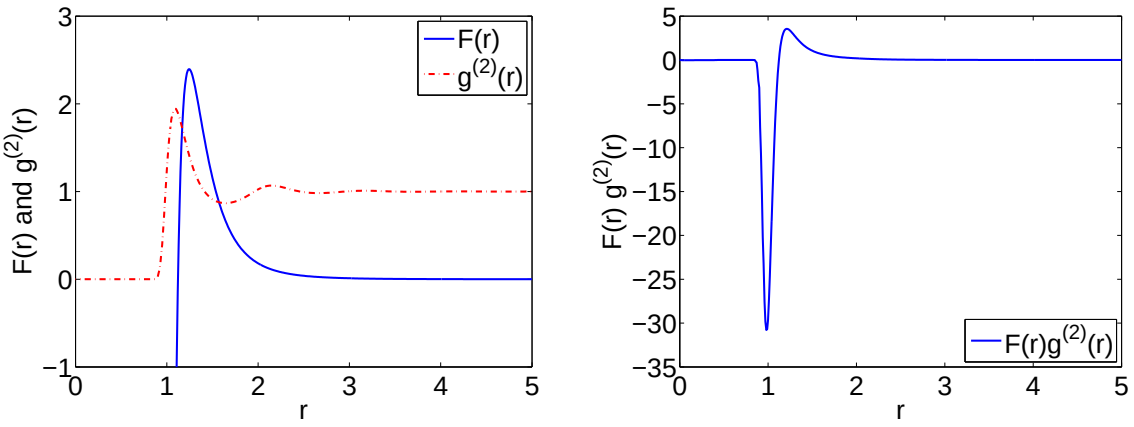

Figure 4.1. *Left: Radial component of the force* $\mathbf{F}$ *and the pair distribution function* $g^{(2)}$ *for a Lennard-Jones fluid. Right: Radial component of the product of* $\mathbf{F}g^{(2)}$.

which effectively means

$$g(\mathbf{x}) \to 0 \quad \text{for} \quad \mathbf{x} \to \mathbf{x}_i^M, \quad i = 1, 2, \ldots, N_M, \tag{4.53}$$

and the solution g is well-defined at the solute particle positions.

Product Approach

Even though the solution of (4.47) is well-defined everywhere, the singular force term will be problematic to handle numerically. The singularity causes very stiff systems that are difficult to solve. Hence, we will introduce an approach which will eliminate the singular term in (4.47). From the observation that the solution g approaches the solution g_0 at the solute particle positions, we conclude that g can be written as a perturbation of g_0. Therefore, we write the solution as

$$g = g_0 \tilde{g}. \tag{4.54}$$

By this choice, we have fixed $g = 0$ at the solute particle positions, but we do not further restrict the solution by this approach, since $g_0(\mathbf{x}) \neq 0$ for all $\mathbf{x} \neq \mathbf{x}_i^M, i = 1, 2, \ldots, N_M$. Next, we insert (4.54) into (4.47) and obtain

$$\begin{aligned} g_0(\mathbf{x})\nabla\tilde{g}(\mathbf{x}) + \tilde{g}(\mathbf{x})\nabla g_0(\mathbf{x}) &= \beta\ \mathbf{F}(\mathbf{x}, \mathbf{x}^M) g_0(\mathbf{x})\tilde{g}(\mathbf{x}) \\ &\quad + \beta\rho\ g_0(\mathbf{x})\tilde{g}(\mathbf{x}) \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}') g^{(2)}(\mathbf{x}, \mathbf{x}') g(\mathbf{x}')\ d\mathbf{x}'. \end{aligned} \tag{4.55}$$

We know that

$$\tilde{g}(\mathbf{x}) \left(\nabla g_0(\mathbf{x}) - \beta\ \mathbf{F}(\mathbf{x}, \mathbf{x}^M) g_0(\mathbf{x}) \right) = 0, \tag{4.56}$$

because we have chosen g_0 as solution of (4.48). We end up with

$$g_0(\mathbf{x}) \left(\nabla \tilde{g}(\mathbf{x}) - \beta\rho \; \tilde{g}(\mathbf{x}) \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}') g^{(2)}(\mathbf{x}, \mathbf{x}') g(\mathbf{x}') \; d\mathbf{x}' \right) = 0. \tag{4.57}$$

Since g_0 is zero only at the solute particle positions, the term in brackets has to vanish. This term is easier to handle numerically, since no singular term is involved.

However, we will further investigate an enhancement of the approach (4.54). We now assume that $\tilde{g}$ can be expressed as

$$\tilde{g}(\mathbf{x}) = e^{-u(\mathbf{x})}. \tag{4.58}$$

This restricts $\tilde{g}$ to be a strict positive function, which is again reasonable from the physical point of view, since probability distributions have to be strictly positive, see Section 2.2. Inserting (4.58) into (4.57) gives

$$e^{-u(\mathbf{x})} \left(\nabla u + \beta\rho \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}') g^{(2)}(\mathbf{x}, \mathbf{x}') g(\mathbf{x}') \; d\mathbf{x}' \right) = 0. \tag{4.59}$$

Again, the term in the brackets has to vanish, which yields

$$\nabla u = -\beta\rho \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}') g^{(2)}(\mathbf{x}, \mathbf{x}') g(\mathbf{x}') \; d\mathbf{x}'. \tag{4.60}$$

This equation is actually a system of three equations for a scalar function u. In order to combine these equations, we apply the divergence to each side and get

$$\Delta u = -\beta\rho \; \nabla \cdot \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}') g^{(2)}(\mathbf{x}, \mathbf{x}') g(\mathbf{x}') \; d\mathbf{x}' \quad \text{with} \quad g(\mathbf{x}) = g_0(\mathbf{x}) e^{-u(\mathbf{x})}. \tag{4.61}$$

Again, the $\cdot$ denotes the scalar product of two vectors in $\mathbb{R}^3$. Equation (4.61) is our final result for the computation of the density of monoatomic solvents around a solute of arbitrary shape. We will solve (4.61) numerically in Chapter 5. Then, it will become evident that the approach (4.54) together with (4.58) is the key idea which makes an efficient numerical solution of the BGY3d equation possible.

4.3.3 BGY3dM Equation for Molecular Solvents

With the original BGY3d equation (4.46), only monoatomic solvents can be considered. This is a restriction which inhibits its use for most real solvents, as e.g. water. Thus, we now investigate how the BGY3d model can be extended to molecular solvents.

The YBG-hierarchy has already been used to compute properties of isolated molecules and molecular fluids. However, all investigations in the literature only

consider very simple potential functions in the context of polymers. Whittington and Dunfield [126] were the first to derive a BGY equation for a single isolated polymer. They employed two simple closure approximations, the so-called independence approximation and the Markov approximation, see [126]. By this, they were able to reproduce well-known self-consistent field equations for a single polymer. Later, Lipson [74], Lipson and Andrews [75] and Taylor and Lipson [120] have used the BGY equation together with the Kirkwood approximation for lattice polymers. To this end, the polymer can only assume discrete configurations, since the positions of the polymer sites are restricted to a spatial lattice. Eu and Gan [32] have analyzed the Kirkwood hierarchy for polymeric fluids. In their approach, integral equations are provided for both the intermolecular and the intramolecular pair distribution functions. They applied their integral equations to an isolated polymer comprised of soft spheres [40], to a polymeric fluid, where the sites are modeled as hard spheres [41, 42], to single polymers with soft and hard spheres and Lennard-Jones potential [42], and to isolated hard-sphere and square-well chains with lengths of up to 1000 sites [43]. Taylor and Lipson [116] have derived a site-site BGY equation for hard-sphere dimers from the molecular BGY equation. In this derivation, triplet as well as quadruplet distribution functions appear. Motivated by results for isolated chains, they employ a so-called normalized site-site superposition approximation for the intramolecular triplet and quadruplet distribution functions, which they found to give better results when dealing with multi-site distribution functions involving connected particles. In [117, 118], they tested the equations for longer chains and rings of hard-spheres with good agreement when compared to other simulation data. Taylor, Lipson et al. [119,122] have extensively studied their equations for isolated square-well chains and rings. The results are in very good agreement with Monte Carlo simulations for shorter chains and higher temperatures. In [121], they consider a fluid of square-well dimers and get satisfactory results for lower densities of the fluid. Finally, Attard [4] derived a site-site BGY equation for polymeric fluids directly from the YBG-hierarchy. He introduced a new triplet superposition approximation for the intramolecular distribution functions and tested the model for fluids of hard-sphere chains with mostly accurate results.

In summary, quite a few models based on the YBG-hierarchy have been developed for polymeric fluids. These models include flexible as well as stiff bonds and mainly differ in the approximations applied to the intramolecular distribution functions. In the literature, these models have been used to compute the pair distribution functions of different polymeric fluids. To this end, the models were always reduced to lower-dimensional equations exploiting the spherical symmetry of the distribution functions.

We will now derive our molecular BGY3d equation designed to compute the site distribution functions of a molecular solvent in the vicinity of an arbitrary solute. Here, no symmetry can be taken advantage of and the full three-dimensional dis-

tribution has to be computed for realistic potential functions, such as the Lennard-Jones and the Coulomb potential.

Derivation of the Site-Site BGY3dM Equation

We begin with the Liouville equation (2.40) at equilibrium and without external forces

$$\sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial \pi}{\partial \mathbf{x}_i} + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \mathbf{F}_{ij} \cdot \frac{\partial \pi}{\partial \mathbf{p}_i} = 0. \tag{4.62}$$

We apply the same factorization of the probability distribution as in Section 2.4.2:

$$\pi(\mathbf{p}, \mathbf{x}) = \mathcal{P}(\mathbf{p})\rho(\mathbf{x}) \tag{4.63}$$

with

$$\mathcal{P}(\mathbf{p}) = \prod_{i=1}^{N} \left(\frac{\beta}{2\pi m_i} \right)^{\frac{d}{2}} e^{-\beta \frac{|\mathbf{p}_i|^2}{2m_i}} \tag{4.64}$$

and

$$\frac{\partial}{\partial \mathbf{p}_i} \mathcal{P}(\mathbf{p}) = -\frac{\beta}{m_i} \mathbf{p}_i \mathcal{P}(\mathbf{p}). \tag{4.65}$$

Inserting this factorization in (4.62) yields

$$\sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial \rho(\mathbf{x})}{\partial \mathbf{x}_i} - \beta \sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i} \cdot \left(\sum_{j=1, j\neq i}^{N} \mathbf{F}_{ij} \right) \rho(\mathbf{x}) = 0. \tag{4.66}$$

The above equation has to be valid independently of the momenta $\mathbf{p}_i$. Hence, it must hold term by term, i.e.,

$$\frac{\partial \rho(\mathbf{x})}{\partial \mathbf{x}_i} - \beta \sum_{j=1, j\neq i}^{N} \mathbf{F}_{ij} \ \rho(\mathbf{x}) = 0 \tag{4.67}$$

for every $i = 1, \ldots, N$.

In contrast to Section 2.4.2, we now consider the case where the particles are not identical. We assume s different species of particles. The index of a quantity assigns the quantity to the respective species, e.g.,

$$N = \sum_{\alpha=1}^{s} N_\alpha, \qquad \rho = \sum_{\alpha=1}^{s} \rho_\alpha \tag{4.68}$$

for the total number of particles and the number densities. As a consequence of the existence of different species, not all particles are indistinguishable any more. Thus,

we have to distinguish in the definition of the reduced particle densities what type of particles are integrated over,

$$\rho^{(n_1,\dots,n_s)}(\mathbf{x}_1^1,\dots,\mathbf{x}_{n_1}^1,\mathbf{x}_1^2,\dots,\mathbf{x}_{n_s}^s) = \frac{N_1!}{(N_1-n_1)!}\cdots\frac{N_s!}{(N_s-n_s)!} \times \int_\Omega \rho(\mathbf{x}_1^1,\dots,\mathbf{x}_{N_1}^1,\mathbf{x}_1^2,\dots,\mathbf{x}_{N_s}^s)\; d\mathbf{x}_{(N_1-n_1)}\cdots d\mathbf{x}_{(N_s-n_s)}, \tag{4.69}$$

where we use the notation $d\mathbf{x}^1_{(N_1-n_1)} = d\mathbf{x}^1_{n_1+1}\cdots d\mathbf{x}^1_{N_1}$. The index of $\rho^{(n_1,\dots,n_s)}$ now explicitly denotes the remaining number of particles for each species. The relation between the particle density and the distribution function reads as

$$\rho^{(n_1,\dots,n_s)} = \rho_1^{n_1}\cdots\rho_s^{n_s} g^{(n_1,\dots,n_s)}. \tag{4.70}$$

We mainly use the alternative notation $g^{(2)}_{\alpha\gamma}$ and $g^{(3)}_{\alpha\gamma\eta}$ for the distribution functions, where $\alpha,\gamma,\eta = 1,\dots,s$ specify the species of particle one, two or three, respectively.

We now integrate equation (4.67) over $N-2$ particle degrees of freedom. Note that there are $\frac{s(s+1)}{2}$ different possibilities to choose the $n_1,\dots,n_s$ such that $n_1+\cdots+n_s = 2$. Further multiplication by $\frac{N_1!}{(N_1-n_1)!}\cdots\frac{N_s!}{(N_s-n_s)!}$ and taking into account relation (4.70) yields

$$\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma) = \beta\mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma) + \beta\sum_{\eta=1}^{s}\rho_\eta\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_3^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma,\mathbf{x}_3^\eta)\; d\mathbf{x}_3^\eta \tag{4.71}$$

for every $\alpha,\gamma = 1,\dots,s$. Here, $\mathbf{F}_{\alpha\gamma}$ denotes the force between atoms of the species α and γ,

$$\mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma) = -\nabla_{\mathbf{x}_1^\alpha} v_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma). \tag{4.72}$$

Equation (4.71) requires a closure relation in order to be solvable. If we insert the Kirkwood approximation (2.51) for the triplet distribution function $g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma,\mathbf{x}_3^\eta)$, we get a coupled system of $\frac{s(s+1)}{2}$ equations for the pair distribution functions

$$\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma) = \beta\mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma) + \beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)\sum_{\eta=1}^{s}\rho_\eta\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_3^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_3^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma,\mathbf{x}_3^\eta)\; d\mathbf{x}_3^\eta \tag{4.73}$$

$$\text{for} \quad \alpha,\gamma = 1,2,\dots,s.$$

That is, we have derived a system of equations for the pair distribution functions of a mixture of s different particle species. However, we are interested in the distribution functions for molecular solvents. Hence, we will further advance the above equations to also handle intramolecular interactions.

We consider an arbitrary molecular fluid. The molecules consist of s not necessarily different particles, the so-called sites. Hence, we compose a mixture of s particle species where the first particle of every species belongs to molecule one, the second particle of every species belongs to molecule two, and so on. Formally, we write

$$\mathbf{X}_i = (\mathbf{x}_i^1, \mathbf{x}_i^2, \ldots, \mathbf{x}_i^s) \quad \text{for} \quad i = 1, \ldots, N_S, \tag{4.74}$$

where $\mathbf{X}_i$ contains all $3s$ coordinates of the particles that constitute molecule i. The number of particles and their density is equal for all species

$$N_S := N_1 = N_2 = \cdots = N_s \quad \text{and} \quad \rho_S := \rho_1 = \rho_2 = \cdots = \rho_s. \tag{4.75}$$

Hence, N_S is the total number of molecules and ρ_S is the molecular density. So far, we did not say anything about the structure of a molecule. That is, because it is not necessary to know anything more than the fact, that each particle has an additional property which makes it possible to distinguish between particles of different molecules. The consequence of this is reflected in the definition of the reduced n-particle density

$$\begin{aligned}\rho^{(n_1,\ldots,n_s)}(\mathbf{x}_1^1, \ldots, \mathbf{x}_{n1}^1, \mathbf{x}_1^2, \ldots, \mathbf{x}_{n_s}^s) = \frac{N_S!}{(N_S - n_S)!} \\ \times \int_\Omega \rho(\mathbf{x}_1^1, \ldots, \mathbf{x}_{N_1}^1, \mathbf{x}_1^2, \ldots, \mathbf{x}_{N_s}^s)\, d\mathbf{x}_{(N_1-n_1)} \cdots d\mathbf{x}_{(N_s-n_s)},\end{aligned} \tag{4.76}$$

where the only difference to equation (4.70) is the factor $\frac{N_S!}{(N_S-n_S)!}$. Here, N_S is the total number of molecules in the system and n_S is the number of molecules that the reduced density $\rho^{(n_1,\ldots,n_s)}$ depends on. We give a small example: If we consider water (H_2O) as solvent and wish to compute $\rho^{(1,0,2)}(\mathbf{x}_1^1, \mathbf{x}_1^3, \mathbf{x}_2^3)$, we have $n_S = 2$, since the reduced density depends on molecules one and two. But we have $n_S = 3$ for $\rho^{(1,0,2)}(\mathbf{x}_1^1, \mathbf{x}_2^3, \mathbf{x}_3^3)$, since now molecules one, two and three are involved. In other words, the reason for the factor $\frac{N_S!}{(N_S-n_S)!}$ is that the molecules are indistinguishable and there are $\frac{N_S!}{(N_S-n_S)!}$ ways to choose n_S different molecules. In consequence, the relation between the reduced particle density and the distribution function reads as

$$\rho^{(n_1,\ldots,n_s)} = \rho_S^{n_S} g^{(n_1,\ldots,n_s)}. \tag{4.77}$$

We now want to compute the two-particle distribution functions by means of the Liouville equation (4.67). Hence, we choose two arbitrary particles $\mathbf{x}_1^\alpha$ and $\mathbf{x}_2^\gamma$ of molecule 1 and 2, integrate (4.67) over the remaining $N-2$ particle degrees of freedom, multiply the whole equation by $\frac{N_S!}{(N_S-n_S)!}$ and insert relation (4.77). This

gives

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) = \beta \mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \\ + \beta \sum_{\eta=1}^{s} \rho_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_3^\eta)\, d\mathbf{x}_3^\eta \\ + \beta \sum_{\eta=1,\eta\neq\alpha}^{s} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta \\ + \beta \sum_{\eta=1,\eta\neq\gamma}^{s} \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta, \end{aligned} \tag{4.78}$$

where the superscript i denotes the intramolecular interaction. If we choose the two particles $\mathbf{x}_1^\alpha$ and $\mathbf{x}_1^\gamma$ from the same molecule, this similarly yields

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) = \beta \mathbf{F}^i_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) \\ + \beta \sum_{\eta=1}^{s} \rho_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma, \mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta \\ + \beta \sum_{\eta=1,\eta\neq\alpha,\eta\neq\gamma}^{s} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta. \end{aligned} \tag{4.79}$$

Equations (4.78) and (4.79) form a set of $\frac{s(s+1)}{2}$ different equations for all possible site-site pair distribution functions of the system. They look similar to equation (4.71) for a simple monoatomic mixture, but contain additional terms which lack the ρ_S factor in front of the integral. These are the terms that account for the intramolecular correlations. To close the system of equations, we again insert the Kirkwood approximation (2.51) and get

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) = \beta \mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \\ + \beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1}^{s} \rho_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_3^\eta)\, d\mathbf{x}_3^\eta \\ + \beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\alpha}^{s} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta \\ + \beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\gamma}^{s} \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta \end{aligned} \tag{4.80}$$

together with

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) &= \beta \mathbf{F}^i_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) \\ &+\beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) \sum_{\eta=1}^{s} \rho_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_1^\gamma, \mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta \\ &+\beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) \sum_{\eta=1, \eta\neq\alpha, \eta\neq\gamma}^{s} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_1^\gamma, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta. \end{aligned} \tag{4.81}$$

Next, we have to specify how the intramolecular interactions differ from the intermolecular ones. The intermolecular interactions are, as always, described by pair potentials such as the Lennard-Jones or the Coulomb potential. They are responsible for the properties of the solvent. The intramolecular interactions, however, account for the geometric structure of the molecules that constitute the solvent. To this end, the molecules are considered to be rigid bodies, i.e., the distance between every pair of particles within the same molecule is constant. This implies that this model is only reasonable for solvents whose molecules have a single geometric configuration at the considered conditions. Water, as one of the most important solvents, is an example for such a fluid. For it to be more realistic, the particles of a molecule should be allowed to fluctuate around their mean positions, i.e., a molecule should be able to contain energy. Nevertheless, these internal vibrations are neglected in our model as in any other liquid state integral equation theory for molecular fluids, see Section 4.2. This is reasonable, since the effect of the internal degrees of freedom on the site-site distribution functions are small.

In order to model the rigid body molecules, we first introduce a harmonic potential as intramolecular interaction:

$$v^i(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) = \kappa (r_1^{\alpha\gamma} - r_0^{\alpha\gamma})^2, \qquad \forall\ \alpha \neq \gamma \tag{4.82}$$

with $r_1^{\alpha\gamma} = |\mathbf{x}_1^\alpha - \mathbf{x}_1^\gamma|$ and $r_0^{\alpha\gamma}$ the desired intramolecular distance between particles of species α and γ. The constant κ defines the strength of the potential. The $\frac{s(s-1)}{2}$ different distances $r_0^{\alpha\gamma}$ completely specify the configuration of the molecule. The potential (4.82), however, does not lead to fixed distances within the molecule, but allows fluctuations around the desired distances. Therefore, we investigate the limit where the constant κ, which determines the strength of the force that constrains two particles to their desired distance, goes to infinity, i.e., we consider $\lim_{\kappa\to\infty} v^i(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa)$. We examine equation (4.81), which determines the intramolecular pair distribution functions, and assume that, in this limit, the solution of equation (4.81) is strongly dominated by the first term of the right hand side, and that all integral terms can be neglected, i.e.,

$$\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa) = \beta \mathbf{F}^i_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa) \qquad \text{for } \kappa \to \infty. \tag{4.83}$$

The dependence on κ is explicitly written in the arguments in order to distinguish between $g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa)$ and the final version of $g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma)$ which will not depend on this parameter. The solution of (4.83) is

$$\begin{aligned} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa) &= c(\kappa, r_0^{\alpha\gamma}) e^{-\beta v^i(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa)} \\ &= c(\kappa, r_0^{\alpha\gamma}) e^{-\beta\kappa (r_1^{\alpha\gamma} - r_0^{\alpha\gamma})^2}. \end{aligned} \tag{4.84}$$

In particular, this is a solution of (4.83) for any factor $c(\kappa, r_0^{\alpha\gamma})$. We introduce this factor such that we get a reasonable limit $\lim_{\kappa\to\infty}$ which obeys the desired normalization condition

$$\lim_{\kappa\to\infty} \int_\Omega g^{(2)}_{\alpha\gamma}(\mathbf{r}_1^{\alpha\gamma}; \kappa)\, d\mathbf{r}_1^{\alpha\gamma} = 1 \tag{4.85}$$

with $\mathbf{r}_1^{\alpha\gamma} = \mathbf{x}_1^\alpha - \mathbf{x}_1^\gamma$. Remember that the pair distribution function only depends on the distance of the sites α and γ. We choose

$$c(\kappa, r_0^{\alpha\gamma}) = \sqrt{\frac{4\beta\kappa}{\pi}} \frac{1}{4\pi (r_0^{\alpha\gamma})^2}. \tag{4.86}$$

With this choice of $c(\kappa, r_0^{\alpha\gamma})$ we can investigate the limit of $\kappa \to \infty$ in more detail. We find for the convolution with an arbitrary function f

$$\begin{aligned} \int_\Omega f(\mathbf{r}_1^{\alpha\gamma} - \mathbf{r}') g^{(2)}_{\alpha\gamma}(\mathbf{r}_1^{\alpha\gamma})\, d\mathbf{r}_1^{\alpha\gamma} &= \lim_{\kappa\to\infty} \int_\Omega f(\mathbf{r}_1^{\alpha\gamma} - \mathbf{r}') g^{(2)}_{\alpha\gamma}(\mathbf{r}_1^{\alpha\gamma}; \kappa)\, d\mathbf{r}_1^{\alpha\gamma} \\ &= \lim_{\kappa\to\infty} \int_\Omega f(\mathbf{r}_1^{\alpha\gamma} - \mathbf{r}') \sqrt{\frac{4\beta\kappa}{\pi}} \frac{e^{-\beta\kappa (r_1^{\alpha\gamma} - r_0^{\alpha\gamma})^2}}{4\pi (r_0^{\alpha\gamma})^2}\, d\mathbf{r}_1^{\alpha\gamma} \\ &= \int_\Omega f(\mathbf{r}_1^{\alpha\gamma} - \mathbf{r}') \frac{\delta(r_1^{\alpha\gamma} - r_0^{\alpha\gamma})}{4\pi (r_0^{\alpha\gamma})^2}\, d\mathbf{r}_1^{\alpha\gamma}. \end{aligned} \tag{4.87}$$

Equation (4.87) represents the definition of the delta distribution as the limit of a Dirac sequence, see e.g. [60]. The result is very intuitive, since the two particles $\mathbf{x}_1^\alpha$ and $\mathbf{x}_1^\gamma$ have to remain exactly at a distance of $r_0^{\alpha\gamma}$ if the restoring force is infinite. The factor $4\pi(r_0^{\alpha\gamma})^2$ represents the surface of the sphere with radius $r_0^{\alpha\gamma}$ and ensures the correct normalization.

Now, we have developed a formalism to describe the molecules as rigid bodies within equations (4.80) and (4.81). We even know all solutions of equations (4.81),

$$g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) = \frac{\delta(r_1^{\alpha\gamma} - r_0^{\alpha\gamma})}{4\pi (r_0^{\alpha\gamma})^2}, \qquad \forall\ \alpha \neq \gamma \tag{4.88}$$

and only have to take care of equations (4.80). But before we insert (4.88) into (4.80), we take a closer look at the second integral term of equation (4.80), since it

contains the intramolecular force $\mathbf{F}^i$. Inserting v^i from (4.82), using relation (4.88) and taking the limit $\kappa \to \infty$ leads to

$$\begin{aligned}
&\lim_{\kappa\to\infty} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta; \kappa) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta; \kappa) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta \\
&= \lim_{\kappa\to\infty} \int_\Omega \left(-\nabla_{\mathbf{x}_1^\alpha} \kappa (r_1^{\alpha\eta} - r_0^{\alpha\eta})^2\right) \sqrt{\frac{4\beta\kappa}{\pi}} \frac{e^{-\beta\kappa(r_1^{\alpha\eta} - r_0^{\alpha\eta})^2}}{4\pi (r_0^{\alpha\eta})^2} g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta \\
&= \lim_{\kappa\to\infty} \frac{1}{\beta} \int_\Omega \left(\nabla_{\mathbf{x}_1^\alpha} \sqrt{\frac{4\beta\kappa}{\pi}} \frac{e^{-\beta\kappa(r_1^{\alpha\eta} - r_0^{\alpha\eta})^2}}{4\pi (r_0^{\alpha\eta})^2} \right) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta \\
&= \lim_{\kappa\to\infty} \frac{1}{\beta} \int_\Omega \sqrt{\frac{4\beta\kappa}{\pi}} \frac{e^{-\beta\kappa(r_1^{\alpha\eta} - r_0^{\alpha\eta})^2}}{4\pi (r_0^{\alpha\eta})^2} \left(\nabla_{\mathbf{x}_1^\eta} g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\right)\ d\mathbf{x}_1^\eta \\
&= \frac{1}{\beta} \int_\Omega \frac{\delta(r_1^{\alpha\eta} - r_0^{\alpha\eta})}{4\pi (r_0^{\alpha\eta})^2} \left(\nabla_{\mathbf{x}_1^\eta} g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\right)\ d\mathbf{x}_1^\eta \\
&= \frac{1}{\beta} \nabla_{\mathbf{x}_1^\alpha} \int_\Omega \frac{\delta(r_1^{\alpha\eta} - r_0^{\alpha\eta})}{4\pi (r_0^{\alpha\eta})^2} g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta. \qquad (4.89)
\end{aligned}$$

In lines four and six of (4.89), we used the property of the convolution,

$$\frac{d}{dx}(a(x) * b(x)) = \frac{da(x)}{dx} * b(x) = a(x) * \frac{db(x)}{dx}, \qquad (4.90)$$

to shift the gradient to $g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)$ and finally outside the integral. Altogether, we can write down our final result for the site-site pair distribution functions

$$\begin{aligned}
\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) &= \beta \mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \qquad (4.91) \\
&\quad + \beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1}^{s} \rho_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_3^\eta)\ d\mathbf{x}_3^\eta \\
&\quad + g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1, \eta\neq\alpha}^{s} \nabla_{\mathbf{x}_1^\alpha} \int_\Omega \omega(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta \\
&\quad + \beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1, \eta\neq\gamma}^{s} \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) \omega(\mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)\ d\mathbf{x}_2^\eta
\end{aligned}$$

with

$$\omega(\mathbf{x}_i^\alpha, \mathbf{x}_i^\eta) = \frac{\delta(r_i^{\alpha\eta} - r_0^{\alpha\eta})}{4\pi (r_0^{\alpha\eta})^2}. \qquad (4.92)$$

We call equations (4.91) the site-site BGY3dM (SS-BGY3dM) equations.

Molecular BGY3d (BGY3dM) Equation

With (4.91) we have developed equations for the site-site pair distribution functions of molecular fluids. However, the goal is to compute the site densities of a molecular solvent around an arbitrary solute

$$\rho_\alpha^S(\mathbf{x}_1^\alpha)|_{\mathbf{x}^M} = \rho_S g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) \tag{4.93}$$

with the conditional probability $g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M)$ and $\mathbf{x}^M \in \mathbb{R}^{3N_M}$ the fixed configuration of the solute. Analog to the derivation of the site-site BGY3dM equations, we obtain the corresponding equation for $g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha, \mathbf{x}^M)$ as

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha, \mathbf{x}^M) &= \beta \mathbf{F}_\alpha(\mathbf{x}_1^\alpha, \mathbf{x}^M) g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha, \mathbf{x}^M) \\ &+\beta \sum_{\eta=1}^{s} \rho_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g_{\alpha\eta}^{(N_M+2)}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta, \mathbf{x}^M)\, d\mathbf{x}_2^\eta \\ &+\beta \sum_{\eta=1,\eta\neq\alpha}^{s} \int_\Omega \mathbf{F}_{\alpha\eta}^i(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g_{\alpha\eta}^{(N_M+2)}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta, \mathbf{x}^M)\, d\mathbf{x}_1^\eta. \end{aligned} \tag{4.94}$$

As before, $\mathbf{F}_\alpha(\mathbf{x}_1^\alpha, \mathbf{x}^M)$ is the total force exerted on solvent particle $\mathbf{x}_1^\alpha$ due to the solute. We again insert the n-level Kirkwood closure relations (4.42) and divide by $g^{(N_M)}(\mathbf{x}^M)$. This yields

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) &= \beta \mathbf{F}_\alpha(\mathbf{x}_1^\alpha, \mathbf{x}^M) g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) \\ &+\beta g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) \sum_{\eta=1}^{s} \rho_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g_{\alpha\eta}^{(2)}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g_\eta^{(N_M+1)}(\mathbf{x}_2^\eta|\mathbf{x}^M)\, d\mathbf{x}_2^\eta \\ &+g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) \sum_{\eta=1,\eta\neq\alpha}^{s} \nabla_{\mathbf{x}_1^\alpha} \int_\Omega g_\eta^{(N_M+1)}(\mathbf{x}_1^\eta|\mathbf{x}^M) \omega(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta, \end{aligned} \tag{4.95}$$

where we already considered the result of (4.89). This is the molecular BGY3d (BGY3dM) equation. The site-site pair distribution functions $g_{\alpha\eta}^{(2)}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta)$ appear as input into the equations and can be computed in advance by equations (4.91). The most important innovation of equation (4.95) is the second integral term of the right hand side. This part accounts for the intramolecular correlations of the solvent molecules.

4.3.4 Improved Approximation for the Intramolecular Interactions

In the derivation of the molecular BGY3d equation we employed the n-level Kirkwood closure relation (4.42) for all occurring n-particle distribution functions. It

was shown by Reiss [95] for monoatomic fluids that, in the thermodynamic limit, i.e. infinite number of particles in an infinite volume, the Kirkwood approximation is the optimal superposition approximation for the triplet distribution function. Attard [4] expanded this statement to the purely intermolecular triplet distribution functions of molecular liquids. An important observation of this statement is that the Kirkwood approximation

$$g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \approx \frac{g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}(\mathbf{x}_2, \mathbf{x}_3)}{g^{(1)}(\mathbf{x}_1) g^{(1)}(\mathbf{x}_2) g^{(1)}(\mathbf{x}_3)} \tag{4.96}$$

is the only superposition of pair functions which obeys the correct asymptotic conditions [4], i.e.

$$g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \to g^{(2)}(\mathbf{x}_2, \mathbf{x}_3) g^{(1)}(\mathbf{x}_1) \quad \text{for} \quad |\mathbf{x}_1| \to \infty. \tag{4.97}$$

This asymptotic limit is only exact to order N^{-1}, with N the number of particles. Equation (4.97) states that particle one becomes uncorrelated to the remaining two particles when it is far away.

For molecular fluids, the situation is different. Here it is known that the Kirkwood approximation for triplet distribution functions is not a satisfying choice if the triplet distribution function includes intramolecular distributions. To see this, we recall that the error terms of the asymptotic limit (4.97) are of the order N^{-1}, i.e., they can be neglected in the thermodynamic limit. This does however not apply for the intramolecular parts, since the number of sites per molecule is constant. Hence, the approximation has to be modified if the correct asymptote and normalization conditions are to be satisfied. These conditions for the mixed intra- and intermolecular triplet distribution function read as follows:

$$g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2, \mathbf{x}^\eta_2) \to g^{(2)}_{\gamma\eta}(\mathbf{x}^\gamma_2, \mathbf{x}^\eta_2) g^{(1)}_\alpha(\mathbf{x}^\alpha_1) \quad \text{for} \quad |\mathbf{x}^\alpha_1| \to \infty, \tag{4.98}$$

$$g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2, \mathbf{x}^\eta_2) \to 0 \quad \text{for} \quad |\mathbf{x}^\gamma_2| \to \infty \tag{4.99}$$

for the asymptotic conditions and further

$$\int_\Omega g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2, \mathbf{x}^\eta_2)\, d\mathbf{x}^\alpha_1 = \frac{N_S - 1}{\rho_S} g^{(2)}_{\gamma\eta}(\mathbf{x}^\gamma_2, \mathbf{x}^\eta_2), \tag{4.100}$$

$$\int_\Omega g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2, \mathbf{x}^\eta_2)\, d\mathbf{x}^\gamma_2 = g^{(2)}_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2) \tag{4.101}$$

for the normalization conditions. As before, N_S is the number of molecules in the fluid and ρ_S is the molecular density. The asymptotic conditions are very intuitive, since the distribution function obviously has to approach zero if the distance of two sites of the same molecule becomes large. Attard [4] has developed a formalism

to compute the optimal pair functions to superpose. They can be computed by iteration of the two equations describing the normalization condition (4.101),

$$g^{(2)}_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2) = \Gamma_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2) \int_\Omega g^{(2)}_{\gamma\eta}(\mathbf{x}^\gamma_2, \mathbf{x}^\eta_2)\Gamma_{\alpha\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2)\ d\mathbf{x}^\gamma_2,$$
$$g^{(2)}_{\alpha\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2) = \Gamma_{\alpha\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2) \int_\Omega g^{(2)}_{\gamma\eta}(\mathbf{x}^\gamma_2, \mathbf{x}^\eta_2)\Gamma_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2)\ d\mathbf{x}^\eta_2. \tag{4.102}$$

Hence, the superposition approximation in this case can be explicitly written as

$$g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2, \mathbf{x}^\eta_2) = g^{(2)}_{\gamma\eta}(\mathbf{x}^\gamma_2, \mathbf{x}^\eta_2)\Gamma_{\alpha\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2)\Gamma_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2) \tag{4.103}$$

with the two functions $\Gamma_{\alpha\gamma}$ and $\Gamma_{\alpha\eta}$ as the solutions of equations (4.102). Note, that only the intermolecular pair functions have to be computed. This approximation satisfies all conditions (4.98) – (4.101) and is therefore the optimal choice. The obvious disadvantage is that it requires the iterative solution of two coupled equations. Moreover, approximation (4.103) would complicate the numerical solution of our BGY3dM model since the product approach could not be employed, as we will see later. This is why we pursue a more pragmatic approach.

In [116], Taylor and Lipson studied the BGY equation for hard sphere dimers. To this end, they derive an equation for the site-site pair distribution functions from the molecular BGY equation. Here, the dimers are treated as rigid bodies with translational and rotational degrees of freedom. This model requires approximations for triplet and quadruplet distribution functions which contain both intra- and intermolecular parts. They employ the so-called normalized site-site superposition approximations (NSSA). For the triplet distribution function, this approximation can (in our notation) be written as

$$g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2, \mathbf{x}^\eta_2) = g^{(2)}_{\gamma\eta}(\mathbf{x}^\gamma_2, \mathbf{x}^\eta_2)\tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2)\tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2) \tag{4.104}$$

with

$$\tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2) = \frac{g^{(2)}_{\alpha\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2)}{n^\eta_{\alpha\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2)}, \qquad \tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2) = \frac{g^{(2)}_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2)}{n^\gamma_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2)} \tag{4.105}$$

and the normalization functions

$$n^\eta_{\alpha\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2) = \int_\Omega g^{(2)}_{\gamma\eta}(\mathbf{x}^\gamma_2, \mathbf{x}^\eta_2)g^{(2)}_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2)\ d\mathbf{x}^\eta_2,$$
$$n^\gamma_{\alpha\eta}(\mathbf{x}^\alpha_1, \mathbf{x}^\eta_2) = \int_\Omega g^{(2)}_{\gamma\eta}(\mathbf{x}^\gamma_2, \mathbf{x}^\eta_2)g^{(2)}_{\alpha\gamma}(\mathbf{x}^\alpha_1, \mathbf{x}^\gamma_2)d\mathbf{x}^\gamma_2. \tag{4.106}$$

These conditional probability functions account for the correlations between the site-site functions due to the constraints imposed by the presence of connected sites.

It is obviously a simplified version of equation (4.103) but has the advantage of being readily evaluated. Nevertheless, it leads to a much better approximation than the Kirkwood superposition approximation when dealing with multi-site distribution functions involving connected pairs [116].

We follow the argumentation of Taylor and Lipson and use the approximation (4.104) in our BGY3dM model. However, in order to keep the computational effort on an acceptable level, we have to further simplify the approximation. To be more precise, we insert (4.104) for those triplet distribution functions which include intramolecular correlations. In the case of the site-site BGY3dM equations, these are the triplet distribution functions $g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)$ and $g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)$, see equation (4.78). We then transform and simplify the corresponding integral term to yield a term which can be numerically evaluated with acceptable effort, see Chapter 5 for all numerical details. For the first triplet distribution function and the respective integral term this gives

$$
\begin{aligned}
&\sum_{\eta=1,\eta\neq\alpha}^{s} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\; d\mathbf{x}_1^\eta \\
&\approx g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\alpha}^{s} \frac{\nabla_{\mathbf{x}_1^\alpha} \int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) \tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\; d\mathbf{x}_1^\eta}{n^\eta_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma)} \\
&= g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\alpha}^{s} \frac{\nabla_{\mathbf{x}_1^\alpha} \int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) \tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\; d\mathbf{x}_1^\eta}{\int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\; d\mathbf{x}_1^\eta} \\
&\approx g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\alpha}^{s} \frac{\nabla_{\mathbf{x}_1^\alpha} \int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) \tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\; d\mathbf{x}_1^\eta}{\int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) \tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\; d\mathbf{x}_1^\eta} \\
&= g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\alpha}^{s} \nabla_{\mathbf{x}_1^\alpha} \ln\left(\int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) \tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\; d\mathbf{x}_1^\eta \right)
\end{aligned}
\tag{4.107}
$$

with $g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) = \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)$ in our case. In the second line, we already considered the limit of an infinite intramolecular restoring force, compare the derivation (4.89). In the fourth line, we replaced $g^{(2)}_{\gamma\eta}$ by its normalized form $\tilde{g}^{(2)}_{\gamma\eta;\alpha}$. By this, it is possible to transform the term so that it involves a single integral only, which is numerically advantageous. Numerical tests in Chapter 5 will reveal that this approximation leads to reasonable results. For the second triplet distribution function and the corresponding integral term, we employ the full NSSA approximation for now:

$$
\begin{aligned}
&\sum_{\eta=1,\eta\neq\gamma}^{s} \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)\; d\mathbf{x}_2^\eta \qquad (4.108) \\
&\approx g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\gamma}^{s} \frac{1}{n^\eta_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma)} \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) \tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) \omega_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)\; d\mathbf{x}_2^\eta.
\end{aligned}
$$

The actual implementation of this term will be discussed in Chapter 5.

Now we have to transfer the NSSA approximation to the conditional $N_M + 2$-particle distribution function as it appears in the BGY3dM equation (4.94) (third line). For this, we insert the n-level Kirkwood approximation (4.42) and simply divide by the normalization functions analog to (4.104). This yields

$$\begin{aligned}
&\sum_{\eta=1,\eta\neq\alpha}^{s} \frac{1}{g^{(N_M)}(\mathbf{x}^M)} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(N_M+2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta, \mathbf{x}^M)\ d\mathbf{x}_1^\eta \\
&\approx g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) \sum_{\eta=1,\eta\neq\alpha}^{s} \frac{\nabla_{\mathbf{x}_1^\alpha} \int_\Omega \tilde{g}^{(N_M+1)}_{\eta;\alpha}(\mathbf{x}_1^\eta|\mathbf{x}^M)\omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta}{n_\alpha^\eta(\mathbf{x}_1^\alpha)} \\
&= g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) \sum_{\eta=1,\eta\neq\alpha}^{s} \frac{\nabla_{\mathbf{x}_1^\alpha} \int_\Omega \tilde{g}^{(N_M+1)}_{\eta;\alpha}(\mathbf{x}_1^\eta|\mathbf{x}^M)\omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta}{\int_\Omega g_\eta^{(N_M+1)}(\mathbf{x}_1^\eta|\mathbf{x}^M)\omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta} \\
&\approx g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) \sum_{\eta=1,\eta\neq\alpha}^{s} \frac{\nabla_{\mathbf{x}_1^\alpha} \int_\Omega \tilde{g}^{(N_M+1)}_{\eta;\alpha}(\mathbf{x}_1^\eta|\mathbf{x}^M)\omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta}{\int_\Omega \tilde{g}_{\eta;\alpha}^{(N_M+1)}(\mathbf{x}_1^\eta|\mathbf{x}^M)\omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta} \\
&= g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) \sum_{\eta=1,\eta\neq\alpha}^{s} \nabla_{\mathbf{x}_1^\alpha} \ln\left(\int_\Omega \tilde{g}^{(N_M+1)}_{\eta;\alpha}(\mathbf{x}_1^\eta|\mathbf{x}^M)\omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta\right). \quad (4.109)
\end{aligned}$$

The derivation is equivalent to (4.107). The replacement of $g_\eta^{(N_M+1)}$ by its normalized form again simplifies the computation of this term. The normalization functions are defined by

$$\begin{aligned}
n_\alpha^\eta(\mathbf{x}_1^\alpha) &= \int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g_\eta^{(N_M+1)}(\mathbf{x}_1^\eta|\mathbf{x}^M)\ d\mathbf{x}_1^\eta, \\
n_\eta^\alpha(\mathbf{x}_1^\eta) &= \int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) d\mathbf{x}_1^\alpha. \quad (4.110)
\end{aligned}$$

We are now able to insert the terms with improved approximation for the intramolecular interactions into the SS-BGY3dM and the BGY3dM equations. The resulting equations can be transformed along the lines of Section 4.3.1 for the monoatomic BGY3d equation in order to simplify their numerical treatment.

4.3.5 Transformations of the Site-Site BGY3dM and BGY3dM Equations

The structure of the site-site BGY3dM as well as the BGY3dM equations is very similar to that of the original BGY3d equation (4.46). This is advantageous, since we can adopt the results of Section 4.3.1 for the original BGY3d equation. With the exact same arguments, the solution of the BGY3dM equations (4.95) approaches

zero at the solute particle positions for any site of the solvent molecules. Hence, we can adopt the transformations of the BGY3d equation in Section 4.3.1 in order to facilitate the numerical treatment of the equations.

We first consider the BGY3dM equations (4.95). To this end, we can insert the product as well as the exponential ansatz and write

$$g_\alpha(\mathbf{x}_1^\alpha) = g_\alpha^0(\mathbf{x}_1^\alpha) e^{-u_\alpha(\mathbf{x}_1^\alpha)} \tag{4.111}$$

with

$$g_\alpha^0(\mathbf{x}_1^\alpha) = e^{-\beta \sum_{i=1}^{N_M} v(\mathbf{x}_1^\alpha - \mathbf{x}_i^M)} \tag{4.112}$$

as the solution of

$$\nabla_{\mathbf{x}_1^\alpha} g_\alpha(\mathbf{x}_1^\alpha) = \beta \mathbf{F}_\alpha(\mathbf{x}_1^\alpha, \mathbf{x}^M) g_\alpha(\mathbf{x}_1^\alpha), \tag{4.113}$$

where we again use the short notation and leave out the $\mathbf{x}^M$ indices and the superscript $(N_M + 1)$. Following the argumentation of Section 4.3.2, we insert (4.111) into (4.95) and apply the divergence. This results in

$$\begin{aligned} \Delta_{\mathbf{x}_1^\alpha} u_\alpha(\mathbf{x}_1^\alpha) = &-\beta \sum_{\eta=1}^{s} \rho_S \nabla_{\mathbf{x}_1^\alpha} \cdot \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g_{\alpha\eta}^{(2)}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g_\eta(\mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta \\ &- \sum_{\eta=1, \eta\neq\alpha}^{s} \Delta_{\mathbf{x}_1^\alpha} \ln \left(\int_\Omega \tilde{g}_{\eta;\alpha}(\mathbf{x}_1^\eta) \omega_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta \right). \end{aligned} \tag{4.114}$$

Here, we also considered the intramolecular term (4.109) with the improved approximation. This is the transformed BGY3dM equation, our final result for the computation of site distribution functions of molecular solvents around an arbitrary solute.

Secondly, we can equivalently transform the site-site BGY3dM equations and use

$$g_{\alpha\gamma}^{(2)}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) = g_{\alpha\gamma}^0(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) e^{-u_{\alpha\gamma}^{(2)}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma)} \tag{4.115}$$

with

$$g_{\alpha\gamma}^0(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) = e^{-\beta v_{\alpha\gamma}(\mathbf{x}_1^\alpha - \mathbf{x}_2^\gamma)} \tag{4.116}$$

as the solution of

$$\nabla_{\mathbf{x}_1^\alpha} g_{\alpha\gamma}^{(2)}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) = \beta \mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) g_{\alpha\gamma}^{(2)}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma), \tag{4.117}$$

where $v_{\alpha\gamma}(\mathbf{x}_1^\alpha - \mathbf{x}_2^\gamma)$ is the potential between sites α and γ of molecules one and two, respectively. As above, we now insert (4.115) into (4.91), consider the improved approximation terms (4.107) and (4.108), apply the divergence and finally end up

with

$$\begin{aligned}\Delta_{\mathbf{x}_1^\alpha} u^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) = -\beta \sum_{\eta=1}^{s} \rho_S \nabla_{\mathbf{x}_1^\alpha} \cdot \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_3^\eta)\ d\mathbf{x}_3^\eta \\ - \sum_{\eta=1,\eta\neq\alpha}^{s} \Delta_{\mathbf{x}_1^\alpha} \ln\left(\int_\Omega \omega(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) \tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta \right) \\ -\beta \sum_{\eta=1,\eta\neq\gamma}^{s} \nabla_{\mathbf{x}_1^\alpha} \cdot \frac{\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) \tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) \omega(\mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)\ d\mathbf{x}_2^\eta}{n^\eta_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma)}. \quad (4.118)\end{aligned}$$

We employ the NSSA approximation without any modifications in the last line. The numerical evaluation of this term is challenging, but it will lead to reasonable results, as we will see in Chapter 5. Equations (4.118) are the transformed site-site BGY3dM equations, which can be used to compute the site-site pair distribution functions that are further necessary as input to the BGY3dM equations (4.114).

The advantage of the formulations (4.118) and (4.114) is, as already described, that the singular force terms do not appear outside the integrals, thereby facilitating the numerical treatment. Since we will compute the convolution integrals by means of Fourier transformations, the Laplace operator can be realized as diagonal scaling in Fourier space, which is numerically very efficient. Together, this yields an efficient method to compute the site distribution functions around arbitrary solutes. Numerical tests of the site-site BGY3dM and BGY3dM equations will be investigated in Chapter 5. Further, we will consider actual applications of the BGY3dM model in Chapter 6.

Chapter 5
Numerical Aspects

In Chapter 4 we have derived the BGY3d and the BGY3dM models for the computation of solvent densities around solutes of arbitrary shape. To this end, the BGY3d model deals with monoatomic solvents, whereas the BGY3dM model is able to consider complex molecular solvents. In this Chapter, we are going to present the numerical algorithms for solving the BGY3d as well as the BGY3dM equations. We will present their discretization and discuss the numerical error. Furthermore, we will also investigate the approximation error of the models due to the closure relations involved. For this, we will compare results obtained by the new BGY3d and BGY3dM models with results from molecular dynamics simulations and the 3d-HNC method of Beglov and Roux [10].

The solution of the BGY3d and of the BGY3dM models requires the prior computation of the pair distribution functions of the pure solvent. To this end, we employ the Born-Green equation (2.52) in the case of monoatomic solvents and the SS-BGY3dM equations (4.91) in the case of molecular solvents. The structure of these equations is almost identical to that of the BGY3d and BGY3dM models. They involve the same approximations and are discretized the same way. It follows that the Born-Green equation and the SS-BGY3dM model are suited better for the investigation of the discretization error, since they do not involve a precomputed function. This is why we will discuss the discretization error on the basis of these equations. In order to quantify the approximation error of the models, we will compute pair distribution functions and site density distributions around simple solutes and compare them to results from molecular dynamics simulations. In the case of three-dimensional density distributions, this investigation is complicated by the slow convergence of the molecular dynamics results. They will still exhibit significant fluctuations even after hundreds of millions of time steps. Hence, we will finally also compare the computational effort of our BGY3d and BGY3dM models and a molecular dynamics simulation.

5.1 Numerical Solution of the BGY3d Equation

We are now going to investigate in full detail the numerical solution of the BGY3d model, which we derived in Section 4.3.1. Therefore, we introduce a short notation and write the BGY3d equation (4.61) as

$$\Delta u = -\beta\rho \ \nabla \cdot \mathbf{K}_g \qquad \text{in } \Omega \tag{5.1}$$

with

$$\mathbf{K}_g(\mathbf{x}) = \int_\Omega \mathbf{F}(\mathbf{x},\mathbf{x}')g^{(2)}(\mathbf{x},\mathbf{x}')g(\mathbf{x}') \ d\mathbf{x}' \tag{5.2}$$

and

$$g(\mathbf{x}) = g_0(\mathbf{x}) \ e^{-u(\mathbf{x})}. \tag{5.3}$$

Equation (5.1) is a non-linear integro-differential equation. As before, $\Omega \subseteq \mathbb{R}^3$ denotes the spatial domain of the system. We have to choose the domain large enough such that finite size effects can be neglected, see also Section 2.2. That is, we assume $u(\mathbf{x}) \approx 0$ outside the domain and can define Dirichlet boundary conditions $u(\partial\Omega) = 0$. This is possible due to the short-range character of the distribution functions[1]. Furthermore, we can always enlarge the domain without affecting the solution. In particular, we can choose $\Omega = \mathbb{R}^3$.

To cope with the non-linearity, we apply a fixed point iteration, which leads to a series of linear integro-differential equations.

Algorithm 5.1 ((Damped) Fixed Point Iteration of the BGY3d Equation).

1. $u^0 = 0$; $l = 0$;
2. $l \leftarrow l + 1$; $g^{l-1} = g_0 \ e^{-u^{l-1}}$; solve

$$\Delta u^l = -\beta\rho\nabla \cdot \mathbf{K}_{g^{l-1}} \qquad \text{in } \Omega \tag{5.4}$$

   and set

$$u^l \leftarrow \nu u^l + (1-\nu)u^{l-1}. \tag{5.5}$$

3. If

$$||u^l - u^{l-1}||_{L_\infty} < \nu\chi \tag{5.6}$$

   stop; else go to 2.

Here, $0 < \nu \leq 1$ is a factor that damps the iteration in order to guarantee convergence. The constant χ defines the stopping criterion of the iteration.

[1] For many types of interaction potentials, the short-range character of pair distribution functions can even be shown rigorously, see [1].

By the application of the fixed point iteration, the problem is transformed into a series of linear PDEs (5.4). After evaluating the right hand side of (5.4), we get a Poisson problem with Dirichlet boundary conditions $u^l(\partial\Omega) = 0$. The boundary condition results from the fact that we want to choose the finite domain such that the solution g is constant outside the domain. We choose this constant to be 1, which leads us to the described boundary condition. We apply the Laplace operator in Fourier space in order to solve the Poisson equation. To this end, the Laplace operator is represented as a diagonal matrix. This has the advantage of fast evaluation but assumes periodic boundary conditions. We can however choose the domain large enough such that the solution of the Poisson problem (5.4) is identical for periodic and Dirichlet boundary conditions, assumed that the force and the pair distribution function are of short range.

We now consider the evaluation of the right hand side of (5.4). For this, we first introduce a short notation $A_i(\mathbf{x}) := F_i(\mathbf{x})g^{(2)}(\mathbf{x})$ for $i = 1, 2, 3$. If we set the domain to $\Omega = \mathbb{R}^3$, we can apply the convolution theorem which states that

$$\mathcal{F}_3(A_i * g) = \mathcal{F}_3(A_i)\mathcal{F}_3(g), \tag{5.7}$$

where the asterix $*$ denotes again the convolution. Furthermore, $\mathcal{F}_3$ is the Fourier transform in three dimensions

$$\hat{g}(\mathbf{k}) := \mathcal{F}_3(g)(\mathbf{k}) = \int_{\mathbb{R}^3} g(\mathbf{x}) e^{-2\pi\imath\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x}, \tag{5.8}$$

and the inverse Fourier transform reads as

$$\mathcal{F}_3^{-1}(\hat{g})(\mathbf{x}) = \int_{\mathbb{R}^3} \hat{g}(\mathbf{k}) e^{2\pi\imath\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{k}. \tag{5.9}$$

Hence, the convolution integral (5.2) can be computed as

$$(\mathbf{K}_g)_i = \mathcal{F}_3^{-1}\left(\mathcal{F}_3(A_i)\mathcal{F}_3(g)\right), \qquad i = 1, 2, 3 \tag{5.10}$$

The representation of the convolution integral by means of Fourier transformations enables a very efficient computation with complexity $\mathcal{O}(n^3 \log(n^3))$ if the fast Fourier transform (FFT) is employed. Here, n denotes the degrees of freedom for one dimension.

In order to cope with the divergence of $\mathbf{K}_g$, we make the following observations: Since the F_i are antisymmetric and $g^{(2)}$ is a symmetric function, A_i is antisymmetric as well. Hence, the convolution of A_i with a constant is zero and we can write

$$A_i * g = A_i * (g - 1) = A_i * h \tag{5.11}$$

with $h = g - 1$. Now, we use the fact that the derivative of the convolution can be shifted to its arguments

$$\partial_{x_i}(A_i * h) = (\partial_{x_i} A_i) * h = A_i * (\partial_{x_i} h). \tag{5.12}$$

Therefore, we obtain for the derivative of the convolution

$$\partial_{x_i} \left(\mathbf{K}_g\right)_i = \mathcal{F}_3^{-1}(\mathcal{F}_3(A_i)\mathcal{F}_3(\partial_{x_i} h)) \tag{5.13}$$

with

$$\mathcal{F}_3(\partial_{x_i} h) = 2\pi\imath k_i \mathcal{F}_3(h), \tag{5.14}$$

since $h(\mathbf{x}) \to 0$ for $|\mathbf{x}| \to \infty$.

If we now expand the solution u^l in a Fourier series

$$u^l = \int_{\mathbb{R}^3} \hat{u}^l e^{2\pi\imath\mathbf{k}\cdot\mathbf{x}} \, d\mathbf{k} = \mathcal{F}_3^{-1}(\hat{u}^l), \tag{5.15}$$

where the $e^{2\pi\imath\mathbf{k}\cdot\mathbf{x}}$ are the basis functions of the Fourier space, application of the Laplace operator simply yields

$$\begin{aligned} \Delta\mathcal{F}_3^{-1}(\hat{u}^l) &= \int_{\mathbb{R}^3} \hat{u}^l \Delta e^{2\pi\imath\mathbf{k}\cdot\mathbf{x}} \, d\mathbf{k} \\ &= -\int_{\mathbb{R}^3} \hat{u}^l (2\pi)^2 |\mathbf{k}|^2 e^{2\pi\imath\mathbf{k}\cdot\mathbf{x}} \, d\mathbf{k} \\ &= -(2\pi)^2 \mathcal{F}_3^{-1}(\hat{u}^l |\mathbf{k}|^2), \end{aligned} \tag{5.16}$$

i.e., the derivative can be shifted to the basis function. Summing up, we can write equation (5.4) as

$$(2\pi)^2 \mathcal{F}_3^{-1}(\hat{u}^l |\mathbf{k}|^2) = 2\pi\imath\beta\rho\mathcal{F}_3^{-1}\left(\sum_{i=1}^{3} k_i \mathcal{F}_3(A_i)\mathcal{F}(h^{l-1})\right). \tag{5.17}$$

Here, the arguments of the inverse Fourier transforms have to be identical up to a constant and we have

$$\hat{u}^l(\mathbf{k}) = \frac{\imath\beta\rho}{2\pi} \sum_{i=1}^{3} \frac{k_i}{|\mathbf{k}|^2} \mathcal{F}_3(A_i)\mathcal{F}_3(h^{l-1}) \tag{5.18}$$

with

$$h^{l-1} = g^{l-1} - 1 \quad \text{and} \quad g^{l-1} = g_0 \, e^{-u^{l-1}}. \tag{5.19}$$

The value corresponding to the zero wavelength is not defined by this relation. We therefore set $\hat{u}^l(0) = 0$ and thereby enforce the normalization

$$\int_\Omega u^l(\mathbf{x}) \, d\mathbf{x} = 0 \tag{5.20}$$

in every step of the fixed point iteration. The function u^l is constant except for a very localized region at the center of the domain, if we assume a short-range

potential. Hence, the normalization (5.20) approximates the Dirichlet boundary condition $u^l(\partial\Omega) = 0$ sufficiently well, if the domain Ω is large enough. The solution u^l can then simply be computed by (5.15),

$$u^l = \mathcal{F}_3^{-1}(\hat{u}^l) \tag{5.21}$$

with $\hat{u}^l$ from (5.18) and $\hat{u}^l(0) = 0$. This way, the solution of the Poisson equation (5.4) requires only one Fourier transform and one inverse Fourier transform in each step of the fixed point iteration. Note, that the Fourier transforms of the A_i, $i = 1, 2, 3$, can be computed and stored in advance.

Solution of the Born-Green Equation

The Born-Green equation (2.52) for the pair distribution function of the pure solvent can be regarded as special case of the BGY3d equation, where a single fixed solvent atom acts as solute. Then, we have $g(\mathbf{x}) = g^{(2)}(\mathbf{x}|\mathbf{x}_M)$ and equation (5.1) is the Born-Green equation with

$$\mathbf{K}_g^{BG}(\mathbf{x}) = \int_\Omega \mathbf{F}(\mathbf{x}, \mathbf{x}')g(\mathbf{x} - \mathbf{x}')g(\mathbf{x}')\ d\mathbf{x}' \tag{5.22}$$

and

$$g(\mathbf{x}) = g_0(\mathbf{x})\ e^{-u(\mathbf{x})}. \tag{5.23}$$

This way, we can follow the same argumentation as above, apply the fixed point iteration and use the Fourier transformation for the convolution of (5.22). This leads to the solution of the Poisson problem

$$u^l = \mathcal{F}_3^{-1}(\hat{u}^l) \tag{5.24}$$

with

$$\hat{u}^l(\mathbf{k}) = \frac{\imath\beta\rho}{2\pi} \sum_{i=1}^{3} \frac{k_i}{|\mathbf{k}|^2} \mathcal{F}_3(A_i^{l-1})\mathcal{F}_3(h^{l-1}), \qquad \hat{u}^l(0) = 0 \tag{5.25}$$

and

$$A_i^{l-1}(\mathbf{x}) = F_i(\mathbf{x})g^{l-1}(\mathbf{x}), \quad h^{l-1} = g^{l-1} - 1 \quad \text{and} \quad g^{l-1} = g_0\ e^{-u^{l-1}}, \tag{5.26}$$

as one step of the fixed point iteration. The only difference to equation (5.21) is that the vector $\mathbf{A}^l$ now also depends on the solution g^l. Hence, three additional Fourier transformations are required in each iteration step, but the same numerical implementation can be used for the Born-Green equation and the BGY3d equation. Both methods are identical, except for the definition of the vector $\mathbf{A}^l$.

5.1.1 Discretization

In order to discretize equations (5.21) and (5.24), we choose a computational domain $\Omega = [0, L]^3 \subset \mathbb{R}^3$. We approximate all functions on a regular grid of size $N = n^3$ with n the number of grid points in one dimension. The grid points are defined by $\mathbf{x}_h(\mathbf{i}) = \mathbf{i}h$ for $\mathbf{i} \in [0, n-1]^3 \subset \mathbb{N}^3$ with the mesh size $h = \frac{L}{n}$ and the length L of the domain in one dimension. We call Ω_h the set of grid points

$$\Omega_h = \{\mathbf{x}_h(\mathbf{i}) \mid \mathbf{i} \in [0, n-1]^3\}. \tag{5.27}$$

All functions are approximated (in the function space) on this grid and are represented as vectors in $\mathbb{R}^N$. The discrete (approximated) version of a function $f : \Omega \to \mathbb{R}$ is denoted by $f_h \in \mathbb{R}^N$. It holds for a known function f at the grid points that

$$(f_h)_{\mathbf{i}} = f(\mathbf{x}_h(\mathbf{i})), \qquad \forall\ \mathbf{i} \in [0, n-1]^3. \tag{5.28}$$

Functions in Fourier space are approximated on a grid with mesh size $h_k = \frac{1}{L}$, i.e., the discrete points are given by $\mathbf{k}(\mathbf{i}) = \mathbf{i}h_k$ for $\mathbf{i} \in [0, n-1] \subset \mathbb{N}^3$. Here, we set

$$\Omega_{h_k} = \{\mathbf{k}_h(\mathbf{i}) \mid \mathbf{i} \in [0, n-1]^3\}. \tag{5.29}$$

Hence, we have chosen the same number of grid points $N = n^3$ for the discretization in real and Fourier space. We employ the fast Fourier transform (FFT) algorithm as it is implemented in the FFTW [39] to compute the discrete Fourier transforms.

The numerical procedure is as follows: All given functions, the force $\mathbf{F}$, the pair distribution function $g^{(2)}$ and the initial guess u, are approximated on the grid. Then, we apply the discrete version of Algorithm 5.1, i.e., we solve (5.4) in every step of the fixed point iteration by (5.21) and (5.24) in the case of the BGY3d equation and the Born-Green equation, respectively. We stop the iteration if $\|u_h^l - u_h^{l-1}\|_{L_\infty^h} < \nu\chi$, where $\|.\|_{L_\infty^h}$ is the discrete version of the L_∞-norm,

$$\|u_h^l\|_{L_\infty^h} := \max_{\mathbf{i}} |(u_h^l)_{\mathbf{i}}|. \tag{5.30}$$

There are two sources of errors if we solve equation (5.1) by the numerical procedure described above. First, the discrete Fourier transform on the finite domain Ω enforces periodicity of the functions on Ω, although our continuous functions are not periodic with respect to Ω. Secondly, we have the discretization error itself. The interplay of these errors and their specific influence on the convergence will be investigated in more detail in the next section.

5.1.2 Convergence

We investigate the convergence of our discretization of the BGY3d model (5.1). As a start, we illustrate the two sources of errors, finite domain and discretization, for

	$\sigma_1 = \sigma_2 = 1.0$		$\sigma_1 = \sigma_2 = 0.1$		$\sigma_1 = 1.0, \sigma_2 = 0.1$	
dof	$e_{L_2^h}$	$e_{L_\infty^h}$	$e_{L_2^h}$	$e_{L_\infty^h}$	$e_{L_2^h}$	$e_{L_\infty^h}$
32^3	2.236_{-8}	4.329_{-5}	3.071_{-3}	1.006_{+2}	2.702_{-5}	6.589_{-2}
64^3	7.392_{-9}	4.330_{-5}	1.150_{-5}	2.448_{+0}	1.769_{-8}	1.253_{-4}
128^3	2.564_{-9}	4.330_{-5}	2.023_{-11}	1.278_{-5}	1.943_{-12}	9.567_{-8}
256^3	9.022_{-10}	4.330_{-5}	9.144_{-21}	7.105_{-15}	6.413_{-13}	9.937_{-8}
512^3	3.186_{-10}	4.330_{-5}	2.895_{-21}	7.105_{-15}	2.219_{-13}	1.003_{-7}

Table 5.1. *Errors of the numerical convolution for different values of σ_1, σ_2 and numbers of degrees of freedom (dof).*

the convolution of two Gaussians. We choose

$$f_i(\mathbf{x}) = \frac{1}{(2\pi\sigma_i^2)^{\frac{3}{2}}} e^{-\frac{\mathbf{x}^2}{2\sigma_i^2}}, \qquad i = 1, 2, \tag{5.31}$$

and compute the convolution

$$f_1 * f_2 = \int_{\mathbb{R}^3} f_1(\mathbf{x} - \mathbf{x}') f_2(\mathbf{x}') \, d\mathbf{x}'. \tag{5.32}$$

The analytical solution is given by

$$f^*(\mathbf{x}) = \frac{1}{(2\pi(\sigma_1^2 + \sigma_2^2))^{\frac{3}{2}}} \, e^{-\frac{\mathbf{x}^2}{2(\sigma_1^2+\sigma_2^2)}}. \tag{5.33}$$

The numerical convolution is computed on the finite domain $\Omega = [-5, 5]^3$ by means of the discrete Fourier transform. We measure the discrete L_2- and L_∞-error of the numerical solution for different values of σ_1 and σ_2 and different mesh sizes h of the domain discretization:

$$\begin{aligned} e_{L_2^h} &= ||f^* - f_h^*||_{L_2^h} = \frac{1}{N} \left(\sum_{\mathbf{i}} |f^*(\mathbf{x}_h(\mathbf{i})) - (f_h^*)_{\mathbf{i}}|^2 \right)^{\frac{1}{2}}, \\ e_{L_\infty^h} &= ||f^* - f_h^*||_{L_\infty^h} = \max_{\mathbf{i}} |f^*(\mathbf{x}_h(\mathbf{i})) - (f_h^*)_{\mathbf{i}}|, \end{aligned} \tag{5.34}$$

with f_h^* the discrete solution of the numerical convolution.

The results for up to $N = 512^3$ grid points are shown in Table 5.1 and Figure 5.1. First, one can clearly observe the influence of the width of the Gaussian functions on the error. The convolution of the two Gaussians with $\sigma_1 = \sigma_2 = 1.0$ shows a nearly constant error for any mesh size h, because the functions are very broad and smooth. Here, the error due to the finite domain size dominates, since the functions have a

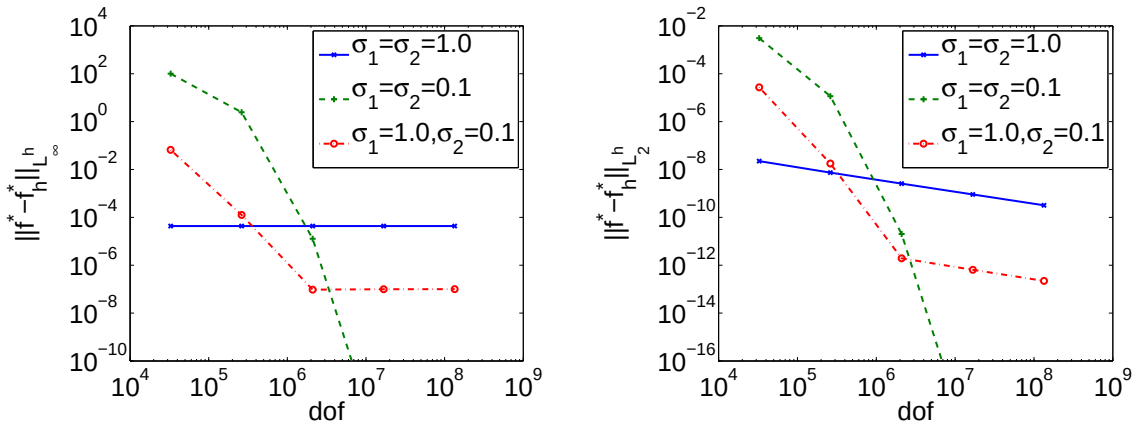

Figure 5.1. *L_∞-error (left) and L_2-error (right) of the numerical convolution for different values of σ_1, σ_2 and numbers of degrees of freedom (dof).*

non-negligible value at the boundaries. For $\sigma_1 = \sigma_2 = 0.1$, however, the influence of the finite domain is negligible. Now, the functions are more sharply defined such that they are not well-approximated with few degrees of freedom. Here, the exponential convergence of the discrete Fourier transform can be observed. The case $\sigma_1 = 1.0$ and $\sigma_2 = 0.1$ illustrates how the two sources of error interplay. The error due to the finite domain is constant, but is exceeded by the discretization error up to $N = 128^3$ grid points.

The observations of this simple test teach us that we have to appropriately choose the size of the domain and the resolution of the grid in order to get small errors at acceptable costs. We can easily choose the domain size large enough such that all functions nearly vanish at the boundaries in the case of short-range potentials such as the Lennard-Jones potential.

We will now investigate how the numerical accuracy improves when we increase the mesh size of the discretization. For this, we consider the solution of the Born-Green equation by algorithm (5.1) with the Lennard-Jones potential, i.e., we choose the force to be

$$\mathbf{F}^{LJ} = -\nabla_{\mathbf{x}} v^{LJ}(\mathbf{x}), \qquad v^{LJ}(r) = 4\epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right). \tag{5.35}$$

We employ the Born-Green equation for this convergence test, since it does not comprise a function which is only approximately known, as this is the case for the BGY3d equation. But all observations also apply to the BGY3d equation, because the discretizations are identical.

We do not know the analytical solution for the considered problem. Hence, we compare the results for different resolutions of the grid. That is, we compute the pair

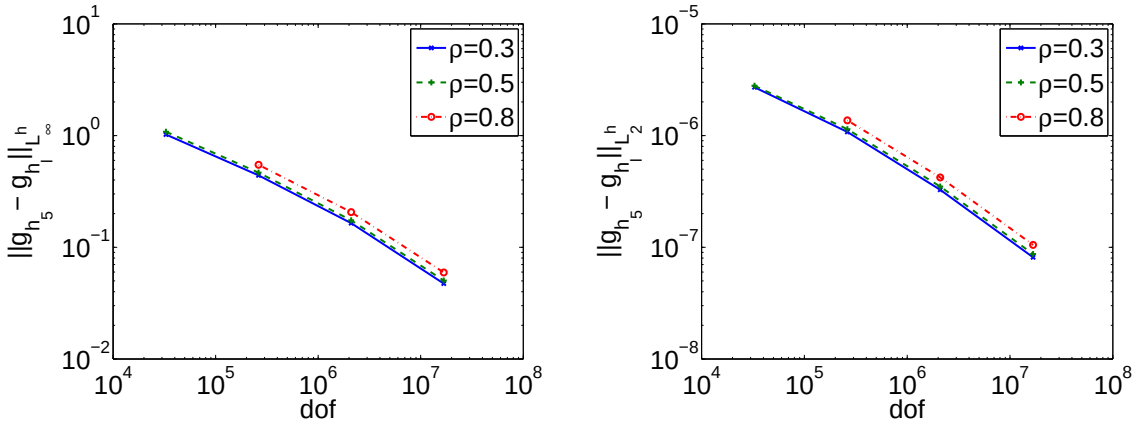

Figure 5.2. *L_∞-error (left) and L_2-error (right) for different values of ρ.*

distribution function for $N_1 = 32^3, N_2 = 64^3, N_3 = 128^3, N_4 = 256^3$ and $N_5 = 512^3$. We then measure the error between the solution on the finest grid (level 5) and all other solutions. To this end, we linearly interpolate all functions to the finest grid and compute the discrete L^2- and L^∞- errors between the interpolated solution and the solution on the finest grid g_{h_5},

$$e_{L_2^h} = \|g_{h_5} - \tilde{g}_{h_l}\|_{L_2^h} = \frac{1}{N_5}\left(\sum_{\mathbf{i}} |(g_{h_5})_{\mathbf{i}} - (\tilde{g}_h)_{\mathbf{i}}|^2\right)^{\frac{1}{2}}, \tag{5.36}$$

$$e_{L_\infty^h} = \|g_{h_5} - \tilde{g}_{h_l}\|_{L_\infty^h} = \max_{\mathbf{i}} |(g_{h_5})_{\mathbf{i}} - (\tilde{g}_h)_{\mathbf{i}}|, \tag{5.37}$$

where $\tilde{g}_h$ indicates the solution with mesh size h interpolated to the finest level. The choice of the simulation parameters can be found in Table 5.2. The values for the density ρ of the fluid and the inverse temperature are chosen to represent the liquid state of the system. The damping factor in the fixed point iteration has been chosen to be $\nu = 0.9$, $\nu = 0.5$ and $\nu = 0.3$ for $\rho = 0.3$, $\rho = 0.5$ and $\rho = 0.8$, respectively.

$\Omega = [-5, 5]^3$	$\rho = 0.3, 0.5, 0.8$	$\beta = 0.6061$
$\epsilon = 1.0$	$\sigma = 1.0$	$\chi = 10^{-6}$

Table 5.2. *Parameters of the model problem.*

Table 5.3 and Figure 5.2 show the results for different values of the density ρ. Obviously, the error significantly decreases up to $N = 256^3$. However, the magnitude of the L_∞-error shows that, even for the finest resolution, the solution is still not approximated with good accuracy. This is caused by the sharp flank

	$\rho = 0.3$		$\rho = 0.5$		$\rho = 0.8$	
dof	$e_{L_2^h}$	$e_{L_\infty^h}$	$e_{L_2^h}$	$e_{L_\infty^h}$	$e_{L_2^h}$	$e_{L_\infty^h}$
32^3	2.706_{-6}	1.022_{+0}	2.793_{-6}	1.081_{+0}	-	-
64^3	1.078_{-6}	4.403_{-1}	1.140_{-6}	4.659_{-1}	1.369_{-6}	5.472_{-1}
128^3	3.271_{-7}	1.644_{-1}	3.495_{-7}	1.743_{-1}	4.215_{-7}	2.064_{-1}
256^3	8.125_{-8}	4.729_{-2}	8.698_{-8}	5.034_{-2}	1.050_{-7}	5.947_{-2}

Table 5.3. *Errors of the BGY3d method for different values of ρ and different numbers of degrees of freedom (dof). Note that for $\rho = 0.8$ and $N = 32^3$ the fixed point iteration of the BGY3d method does not converge.*

of the pair-distribution functions from zero to the first peak, compare also Figures 2.1 – 2.3. This region is dominated by the factor $g_0 = e^{-\beta v^{LJ}(\mathbf{x})}$, which is infinitely often differentiable but very sharp and therefore difficult to approximate by our discretization. Hence, the reduction of the error is dominated by the approximation of the constant function g_0. This is why we will further investigate the solution with respect to its product structure in more detail.

The pair distribution function is computed in each step of the fixed point iteration (5.1) by our product approach

$$g^l(\mathbf{x}) = g_0(\mathbf{x}) e^{-u^l(\mathbf{x})} \quad \text{with} \quad g_0(\mathbf{x}) = e^{-\beta v^{LJ}(\mathbf{x})}, \tag{5.38}$$

where u^l is the solution of the Poisson equation (5.4) of step l. As stated above, the function g_0 is very sharp and difficult to approximate. Hence, we now repeat the simulations above with a slightly different computation of g_h: Instead of g_h, we now interpolate u_h to the finest grid before we compute g_h by (5.38) on the finest grid. This way, we can compare the solution of the fixed point iteration for different resolutions without the influence of the function g_0.

Table 5.4 and Figure 5.3 show the results for different values of the density ρ for the modified interpolation of the solution. The absolute values of the L_∞- and the L_2-errors are now considerably improved. The rate of the error reduction decreases from 0.066 between level 1 and 2 to 0.26 between level 3 and 4 in the case of the L_∞-error and of density $\rho = 0.3$. A similar behavior can be observed for the other densities and the L_2-error. Due to the small number of different grid sizes, an asymptotic rate cannot be determined yet. One may wonder why the roughness of g_0 does not have a stronger influence on the solution through the integral term (5.22) (where it is again incorporated through g from (5.38)). Since the sharp region of g_0 is very localized, the result of the integral term is very smooth and only weakly affected by the poor approximation of g_0, which explains the significantly smaller errors measured with this procedure.

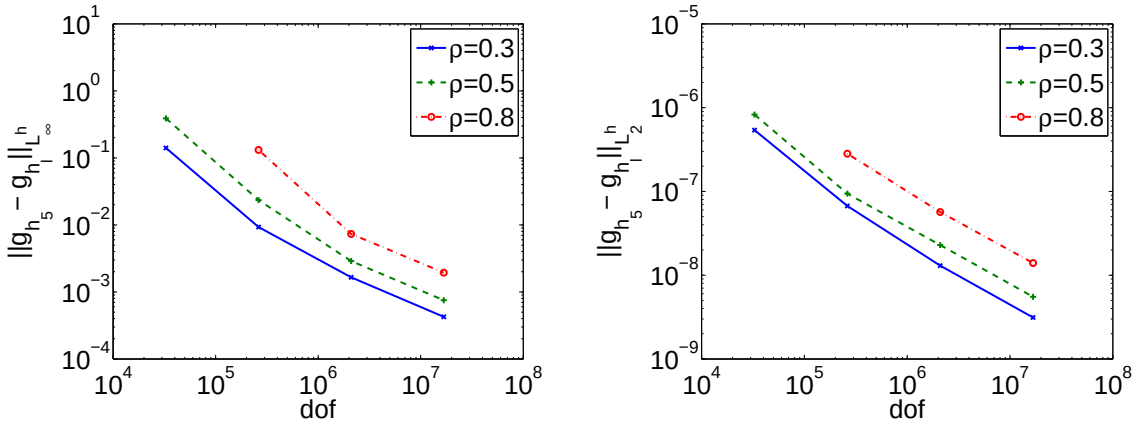

Figure 5.3. *L_∞-error (left) and L_2-error (right) for different values of ρ and different numbers of degrees of freedom (dof). The function g_0 is multiplied on the finest grid.*

	$\rho = 0.3$		$\rho = 0.5$		$\rho = 0.8$	
dof	$e_{L_2^h}$	$e_{L_\infty^h}$	$e_{L_2^h}$	$e_{L_\infty^h}$	$e_{L_2^h}$	$e_{L_\infty^h}$
32^3	5.396_{-7}	1.406_{-1}	8.250_{-7}	3.891_{-1}	-	-
64^3	6.710_{-8}	9.298_{-3}	9.418_{-8}	2.350_{-2}	2.819_{-7}	1.313_{-1}
128^3	1.301_{-8}	1.651_{-3}	2.287_{-8}	2.892_{-3}	5.674_{-8}	7.330_{-3}
256^3	3.127_{-9}	4.256_{-4}	5.516_{-9}	7.523_{-4}	1.396_{-8}	1.940_{-3}

Table 5.4. *Errors of the BGY3d method for different values of ρ and different numbers of degrees of freedom (dof). The function g_0 is multiplied on the finest grid. Note that for $\rho = 0.8$ and $N = 32^3$ the fixed point iteration of the BGY3d method does not converge.*

In summary, the solution of the BGY3d equation is satisfactorily approximated for all grid resolutions with $N \geq 256^3$, i.e., the relative L_∞-error is roughly 5% when compared to a solution with $N = 512^3$ grid points. The major part of this error is due to the difficult approximation of the constant function g_0 and is localized at the sharp peak of this function. The smooth regions of the distribution functions are well approximated even for small number of degrees of freedom, as is indicated by the low L_2-errors. Nevertheless, we have to keep in mind that the actual error between our approximated solution and the (unknown) exact solution may be greater, but this difference should be negligible, since the probability distributions are not expected to exhibit very fine structures. In the following, we will use $N = 256^3$ grid points when possible. Sometimes this has to be reduced to $N = 128^3$ for special reasons

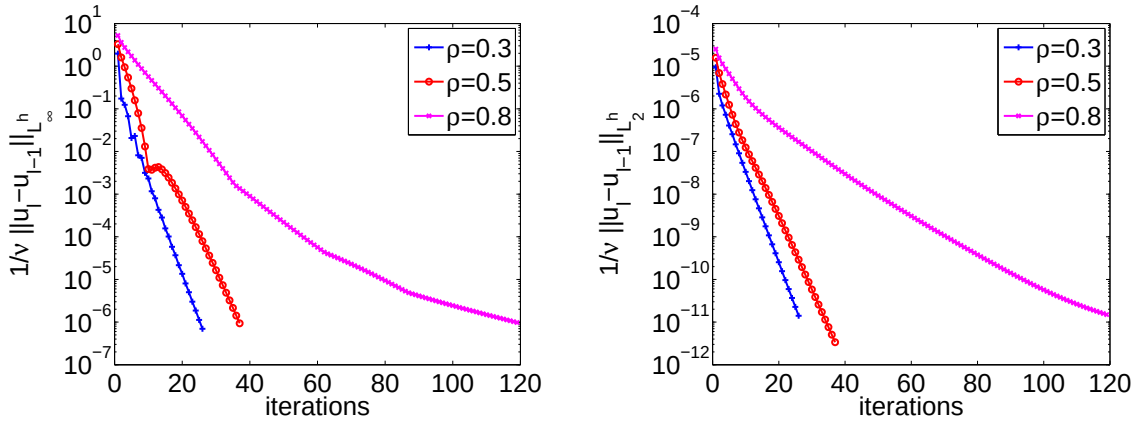

Figure 5.4. *Convergence of the fixed point iteration for different values of ρ in the L_∞-norm (left) and the L_2-norm (right).*

which will be explained in the corresponding section.

Finally, we are going to discuss the convergence of the non-linear fixed point iteration. Figure 5.4 shows the convergence history for $N = 256^3$ and different densities, i.e., the logarithmic plots of

$$\frac{1}{\nu}\|u_h^l - u_h^{l-1}\|_{L_\infty^h} \quad \text{and} \quad \frac{1}{\nu}\|u_h^l - u_h^{l-1}\|_{L_2^h}$$

with u_h^l being the discrete solution of the linearized problem (5.4) of step l of the fixed point iteration. A decrease of the norms indicates the contracting property of the fixed point iteration, which is a condition for convergence. Indeed, the L_2-norm shows a monotonic decrease during iteration, whereas regions of increase can be observed in the L_∞-norm. This is because the discrete L_∞-norm is more sensitive to the discrete approximation of the solution. Strong local changes may cause a short increase of the norm, since it is sensitive with respect to the exact location of the grid points. This happens especially at the beginning of the iteration process. The discrete L_2-norm is more robust due to the averaging process involved and shows a monotonic decrease. Hence, the fixed point iteration converges for our choice of the initial u_h^0 and the damping factor ν. Asymptotically, linear convergence can be observed for all densities. The rate of convergence however becomes worse with increasing density due to the stronger influence of the the non-linear term.

5.2 Test of the BGY3d Model

In this section we are going to investigate the model errors of the BGY3d equations in more detail. We have seen in Section 5.1.2 that the numerical errors can be

controlled if we appropriately choose the size of the domain and the resolution of the grid. We will show that the numerical errors are negligible when we compare the results of the BGY3d equation with results gained from molecular dynamics simulations. The deviations between the results will be dominated by the model error of the BGY3d model. This is due to the approximations involved in the model. In order to classify the BGY3d equation within existing methods, all results will also be compared to the 3d-HNC model of Beglov and Roux [10].

5.2.1 Computing the Solvent Density with Molecular Dynamics

In order to analyze the model error of the BGY3d equation, we are going to compare the BGY3d results with those of a molecular dynamics simulation. To be more precise, we compute the mean solvent density of a simple Lennard-Jones fluid around a solute composed of one, two, three or four particles with both the BGY3d method and by a molecular dynamics simulation. In general, the mean solvent density can be computed by

$$\begin{aligned}\langle \rho(\mathbf{x}) \rangle_{(\mathbf{x}^M)} &= \left\langle \sum_{i=1}^{N_S} \delta(\mathbf{x} - \mathbf{x}_i^S) \right\rangle_{(\mathbf{x}^M)} \\ &= C_{MS}^{-1} \int_{\Omega_{N_S}} \sum_{i=1}^{N_S} \delta(\mathbf{x} - \mathbf{x}_i^S) e^{-V(\mathbf{x}^M, \mathbf{x}^S)} \, d\mathbf{x}^S, \end{aligned} \tag{5.39}$$

with

$$C_{MS} = \int_{\Omega_{N_S}} e^{-V(\mathbf{x}^M, \mathbf{x}^S)} \, d\mathbf{x}^S, \tag{5.40}$$

where we have only used the definition of $\langle . \rangle_{(\mathbf{x}^M)}$ (3.29) from Section 3.3. The integral in (5.39) describes the spatial part of an integral over the phase space of the solvent's degrees of freedom. It can be computed by a molecular dynamics simulation by means of the ergodic hypothesis, i.e., the ensemble average $\langle . \rangle_{(\mathbf{x}^M)}$ can be replaced by a time average of the molecular dynamics trajectory $\mathbf{x}^S(t)$ for large times t_{end}. The hypothesis states that

$$C_{MS}^{-1} \int_{\Omega_{N_S}} \sum_{i=1}^{N_S} \delta(\mathbf{x} - \mathbf{x}_i^S) e^{-V(\mathbf{x}^M, \mathbf{x}^S)} \, d\mathbf{x}^S = \lim_{t_{end} \to \infty} \frac{1}{t_{end}} \int_0^{t_{end}} \sum_{i=1}^{N_S} \delta(\mathbf{x} - \mathbf{x}_i^S(t)) \, dt \tag{5.41}$$

holds, where $\mathbf{x}_i^S(t)$ is the position of solvent particle i at time t. The approximated trajectory of a computer simulation is given only at discrete time steps t_k, and the integral has to be replaced by a sum

$$\frac{1}{t_{end}} \int_0^{t_{end}} \sum_{i=1}^{N_S} \delta(\mathbf{x} - \mathbf{x}_i^S(t)) \, dt \approx \frac{1}{N_t} \sum_{k=1}^{N_t} \sum_{i=1}^{N_S} \delta(\mathbf{x} - \mathbf{x}_i^S(t_k)), \tag{5.42}$$

where N_t is the total number of time steps.

The actual implementation of (5.42) is as follows: A periodic box $\Omega = [0, L]^3$ is filled with solvent particles at the desired overall density. For convenient post-processing, the solute particles are placed in the center of the box. Then, a molecular dynamics simulation in the canonical ensemble is run with the solute particles fixed at their positions, i.e., all forces between solvent and solute particles are computed, but the solute particles do not move. At every time step t_k, the instantaneous solvent density ρ_{t_k} is computed. The densities are approximated on a regular grid as described in Section 5.1.1. The transformation of the particle distribution to the grid is done by linear interpolation,

$$(\rho_{t_k})_{\mathbf{i}} = \sum_{j=1}^{N_S} \int_{\Lambda_{h,\mathbf{i}}} \phi_{h,\mathbf{i}}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}_j^S(t_k)) \, d\mathbf{x}, \tag{5.43}$$

where the index $\mathbf{i}$ defines a grid point and $\phi_{h,\mathbf{i}}$ is the hat function in three dimensions centered at $\mathbf{x}_h(\mathbf{i})$ with support $\Lambda_{h,\mathbf{i}}$. After N_t time steps, the approximation of the mean solvent density can be computed by the sum of the particle densities at all time steps,

$$\langle \rho(\mathbf{x}_h(\mathbf{i})) \rangle_{(\mathbf{x}^M)} \approx \frac{1}{N_t} (\rho_{[0,t_{end}]})_{\mathbf{i}} \quad \text{with} \quad \rho_{[0,t_k]} = \sum_{j=1}^{k} \rho_{t_j}. \tag{5.44}$$

The situation becomes easier when we want to compute the pair distribution function $g^{(2)}$ of the pure solvent. It is given by an ensemble average over any pair of solvent particles

$$g^{(2)}(r) = \frac{1}{\rho} \langle \delta(|\mathbf{x}_i^S - \mathbf{x}_j^S| - r) \rangle \quad \text{for any} \quad i, j = 1, \ldots, N_S, \quad i \neq j. \tag{5.45}$$

This is only a one-dimensional function, and any pair of solvent particles can be used to compute it. These two aspects drastically improve the rate of convergence. Details on how the pair distribution function can be computed by molecular dynamics simulations can be found in the literature, see e.g. [46].

Remember that the equality in (5.41) is only valid in the limit of infinite time. Hence, the number of time steps required for a sufficient convergence of the right hand side of (5.42) may be very large. And indeed, we observe distinct fluctuations of the solvent densities computed from a molecular dynamics simulation. These errors can be identified especially in the settings where the full three-dimensional density is computed. Nevertheless, the computation of the solvent density by molecular dynamics allows us to quantify the model error of the approximate models, since most important features of the functions are satisfactorily reproduced, as we will see in the following.

5.2.2 Comparison of BGY3d with Molecular Dynamics

We will now to present numerical results computed with the BGY3d model, the 3d-HNC method of Beglov and Roux [10] and molecular dynamics simulations, respectively. The setting for our simulations is as follows: We choose a box Ω and place the solute atoms in the middle of the box surrounded by the solvent with number density ρ. In these tests, all particles (solute and solvent) are identical and interact via the Lennard-Jones potential

$$v^{LJ}(r) = 4\epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right). \tag{5.46}$$

We perform several simulations with different solvent densities and numbers of solute atoms. The solute atoms are placed so they have distance $\sigma^{\frac{1}{6}}$ to each other and their center of mass is at the origin. Table 5.5 lists the parameters that were used for the different tests. We employ the same discretization for the BGY3d and 3d-HNC model and the density computed by the molecular dynamics simulation. We choose $\chi = 10^{-6}$ as the stopping criterion for all fixed point iterations.

$\Omega = [-5, 5]^3$	$N = 128^3, 256^3$	$T = 1.65$
$\epsilon = 1.0$	$\sigma = 1.0$	$m = 1.0$

Table 5.5. *Parameters for domain, discretization, temperature, potential and mass of the particles.*

The 3d-HNC method is solved as described in [10]. This is basically an iteration of (4.10),

$$\Delta\rho^{k+1}(\mathbf{x}) = \nu\rho \left(e^{-\beta V(\mathbf{x}) + \Delta\rho^k(\mathbf{x}) * c(\mathbf{x})} - 1 \right) + (1-\nu)\Delta\rho^k(\mathbf{x}). \tag{5.47}$$

The direct correlation function c of the pure solvent has to be computed beforehand. To this end, the standard spherical symmetric Ornstein-Zernike equation (B.5) with HNC closure as described in Appendix B is used. Unlike in [10], we use $\Delta\rho^0 = \rho(g_0 - 1)$ with $g_0(\mathbf{x}) = e^{-v^{LJ}(\mathbf{x})}$ from (4.49) as initial guess for the fixed point iteration. The mixing factor ν is not changed during the iteration.

We have noticed that iteration of (4.10) can be numerically unstable if the condition

$$\int_\Omega \Delta\rho(\mathbf{x})\ d\mathbf{x} = 0 \tag{5.48}$$

is not fulfilled with appropriate accuracy. This can happen because of the finite size of the domain or the discretization error of the convolution integral. This error accumulates in the 3d-HNC method, since the direct correlation function $c(\mathbf{x})$ is

symmetric and hence, the convolution of $c(\mathbf{x})$ with a constant does not vanish. However, this can be easily avoided by enforcing condition (5.48) at every iteration step. We implement this during the computation of the convolution by setting the zero wavelength component of the Fourier transform of $\Delta\rho^l$ to zero, i.e., we set $\widehat{\Delta\rho^l}(0) = 0$ for every l of the fixed point iteration.

We use the molecular dynamics package TREMOLO [123] for the molecular dynamics simulations. We employ periodic boundary conditions, and the potential is truncated and smoothed to zero at a distance r_{cut}. The smoothing function is defined as

$$S(r) = \begin{cases} 1 & \text{for} \quad r \leq r_l, \\ 1 - (r - r_l)^2(3r_{\text{cut}} - r_l - 2r)/(r_{\text{cut}} - r_l)^3 & \text{for} \quad r_l < r < r_{\text{cut}}, \\ 0 & \text{for} \quad r \geq r_{\text{cut}}, \end{cases} \tag{5.49}$$

with $r_l = 2.3\sigma$ and $r_{cut} = 2.5\sigma$. Hence, we actually use the potential

$$v_{MD}^{LJ}(r) = 4\epsilon S(r)\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right). \tag{5.50}$$

This implies that we do not use the exact same potentials for the different methods. If one however considers the deviations between the approximative models and the molecular dynamics results, this error can be neglected. In the case of one solute particle, the solvent density corresponds to the pair distribution function of the pure solvent

$$\langle \rho(\mathbf{x}) \rangle_{(\mathbf{x}_1^M)} = \rho g^{(2)}(r) \quad \text{with} \quad r = |\mathbf{x}|. \tag{5.51}$$

The pair distribution functions can be more efficiently computed by molecular dynamics as described above, resulting in the radial component of the three-dimensional spherical symmetric distribution. For a comparison with the BGY3d and 3d-HNC results, we interpolate the radial component of the density distribution onto the three-dimensional grid.

In order to improve the convergence of the molecular dynamics simulation concerning the systems where full three-dimensional resolution of the density is necessary, we reduce the grid resolution to $N = 128^3$ and repeat every simulation with different initial values for the momenta of the solvent particles. Afterwards, the resulting densities of the two simulations are averaged.

We use time steps of size $\Delta t = 0.005$ for systems with solvent density $\rho = 0.3$ and time steps of size $\Delta t = 0.001$ for all other systems. The constant temperature ensemble is realized by a Nosé-Hoover-Thermostat, see [46] for details.

In order to compare the results of the different methods, we will plot the deviations between the results of molecular dynamics and the respective other method (BGY3d or 3d-HNC),

$$g_h^{diff} = |g_h - g_h^{MD}|. \tag{5.52}$$

Here, g_h stands for the discrete result of the BGY3d (g_h^{BGY}) or the 3d-HNC (g_h^{HNC}) method. Moreover, we will compute certain quantities which allow conclusions about the method's accuracy, namely the discrete L_2-norm of the deviation

$$e_{L_2^h} = \|g_h - g_h^{MD}\|_{L_2^h} = \frac{1}{N}\left(\sum_{\mathbf{i}} |(g_h)_{\mathbf{i}} - (g_h^{MD})_{\mathbf{i}}|^2\right)^{\frac{1}{2}}, \tag{5.53}$$

the discrete L_∞-norm of the deviation

$$e_{L_\infty^h} = \|g_h - g_h^{MD}\|_{L_\infty^h} = \max_{\mathbf{i}} |(g_h)_{\mathbf{i}} - (g_h^{MD})_{\mathbf{i}}| \tag{5.54}$$

and the difference between the maxima of each density

$$e_{max} = |\max_{\mathbf{i}}(g_h)_{\mathbf{i}} - \max_{\mathbf{i}}(g_h^{MD})_{\mathbf{i}}|. \tag{5.55}$$

Note that the molecular dynamics densities g_h^{MD} also are not exempt from errors, see the discussion above. Hence, the resulting error quantities should be interpreted carefully.

We first have to transform the results of the different methods to solvent distributions with identical normalization

$$\frac{1}{|\Omega|}\int_\Omega g(\mathbf{x})\ d\mathbf{x} = 1, \tag{5.56}$$

where $|\Omega|$ is the volume of the domain Ω. The BGY3d method directly computes $g(\mathbf{x})$, the 3d-HNC method computes the deviation of the density $\Delta\rho(\mathbf{x})$, and the molecular dynamics simulation results in the solvent density[2] $\rho(\mathbf{x})$. The relation between $g(\mathbf{x})$, $\rho(\mathbf{x})$ and $\Delta\rho(\mathbf{x})$ is

$$\rho(\mathbf{x}) = \bar{\rho} + \Delta\rho(\mathbf{x}) = \bar{\rho}g(\mathbf{x}) \quad \text{with} \quad \bar{\rho} = \frac{1}{|\Omega|}\int_\Omega \rho(\mathbf{x})\ d\mathbf{x}. \tag{5.57}$$

It can be used to compute $g(\mathbf{x})$ from either $\rho(\mathbf{x})$ in the case of molecular dynamics or from $\Delta\rho(\mathbf{x})$ in the case of the 3d-HNC method.

Results

We now present the results of the BGY3d, the 3d-HNC method and the molecular dynamics simulations. Table 5.6 shows the computed error quantities for all results. Figures 5.5 – 5.7 show plots of the radial component of the solutions and deviations from the molecular dynamics results for the systems with spherical symmetric solute.

[2]To this end, we employ the short notation and leave out the $\langle . \rangle_{(\mathbf{x}^M)}$ indicating the ensemble average with fixed solute for all functions.

		MD	BGY3d				3d-HNC			
ρ	N_M	$\max g$	$\max g$	$e_{L_2^h}$	$e_{L_\infty^h}$	e_{max}	$\max g$	$e_{L_2^h}$	$e_{L_\infty^h}$	e_{max}
0.3	1	1.77	1.83	1.95_{-06}	0.16	0.07	1.81	1.61_{-06}	0.18	0.04
0.5	1	1.85	1.89	1.94_{-06}	0.11	0.04	1.93	2.48_{-06}	0.26	0.08
0.8	1	2.28	2.17	8.63_{-06}	0.28	0.12	2.45	5.73_{-06}	0.54	0.17
0.5	2	3.20	3.48	1.76_{-05}	0.44	0.28	3.54	1.48_{-05}	0.65	0.33
0.5	3	5.01	6.07	2.13_{-05}	1.16	1.07	6.15	1.61_{-05}	1.37	1.15
0.5	4	5.25	6.54	2.53_{-05}	1.36	1.29	6.44	1.75_{-05}	1.39	1.19

Table 5.6. *Comparison of BGY3d and 3d-HNC with molecular dynamics results for systems with different solvent densities ρ and numbers of solute particles N_M.*

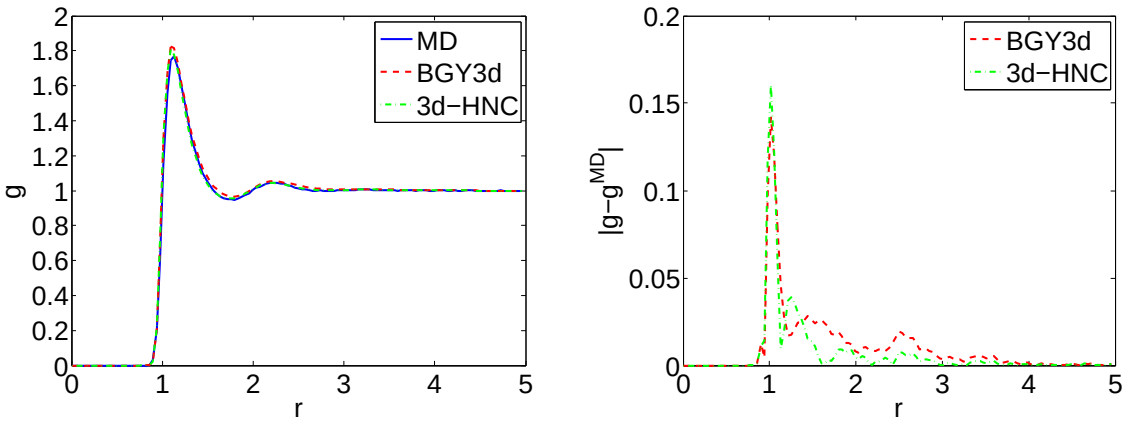

Figure 5.5. *Left: Pair distribution function for $\rho = 0.3$. Right: Deviation of pair distribution function from molecular dynamics.*

Figures 5.8 through 5.10 show cut planes of the densities and the deviations for the systems with full three-dimensional densities. As already pointed out above, the densities computed by molecular dynamics for more than one solute atom exhibit distinct fluctuations which indicate that the solution is not yet finally converged. However, the magnitude of the fluctuations is small enough when compared to the approximation errors of the two continuous models considered.

Some general observations can be made. The deviation of both models, the BGY3d model and the 3d-HNC model, grows with increasing overall solvent density and increasing number of solute particles. However, the smallest deviation for our BGY3d model appears for one solute atom and the intermediate density $\rho = 0.5$. Compared to the molecular dynamics result, the oscillation pattern behind the first

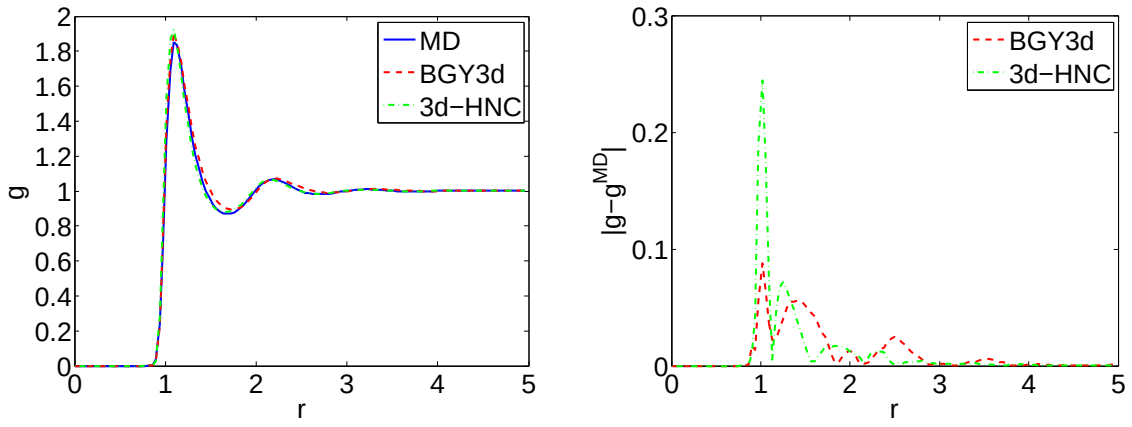

Figure 5.6. *Left: Pair distribution function for $\rho = 0.5$. Right: Deviation of pair distribution function from molecular dynamics.*

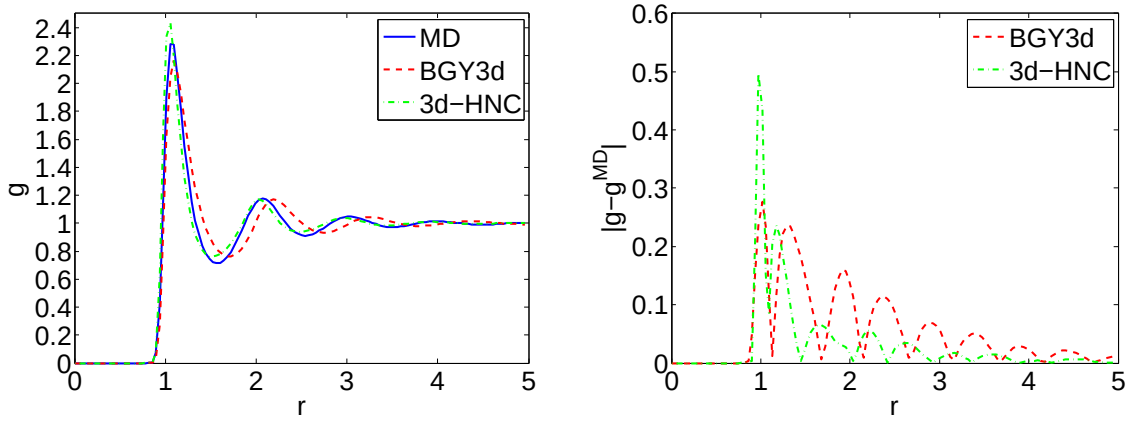

Figure 5.7. *Left: Pair distribution function for $\rho = 0.8$. Right: Deviation of pair distribution function from molecular dynamics.*

peak of both approximate models gets more and more out of phase with increasing density. This effect is stronger for the BGY3d model. It is known to be a result of the two-particle superposition approximation, see e.g. [3]. The approximation can only be improved by considering three-particle interactions in the integral terms. The L_∞-error and the error at the maximum e_{max} are smaller for BGY3d, except for the system with four solute atoms. Hence, the major difference of the approximations obtained by the BGY3d and by the 3d-HNC model can be stated as follows: The BGY3d model approximates better the position and the height of the main peak, whereas the 3d-HNC model is superior in approximating the oscillation which follows

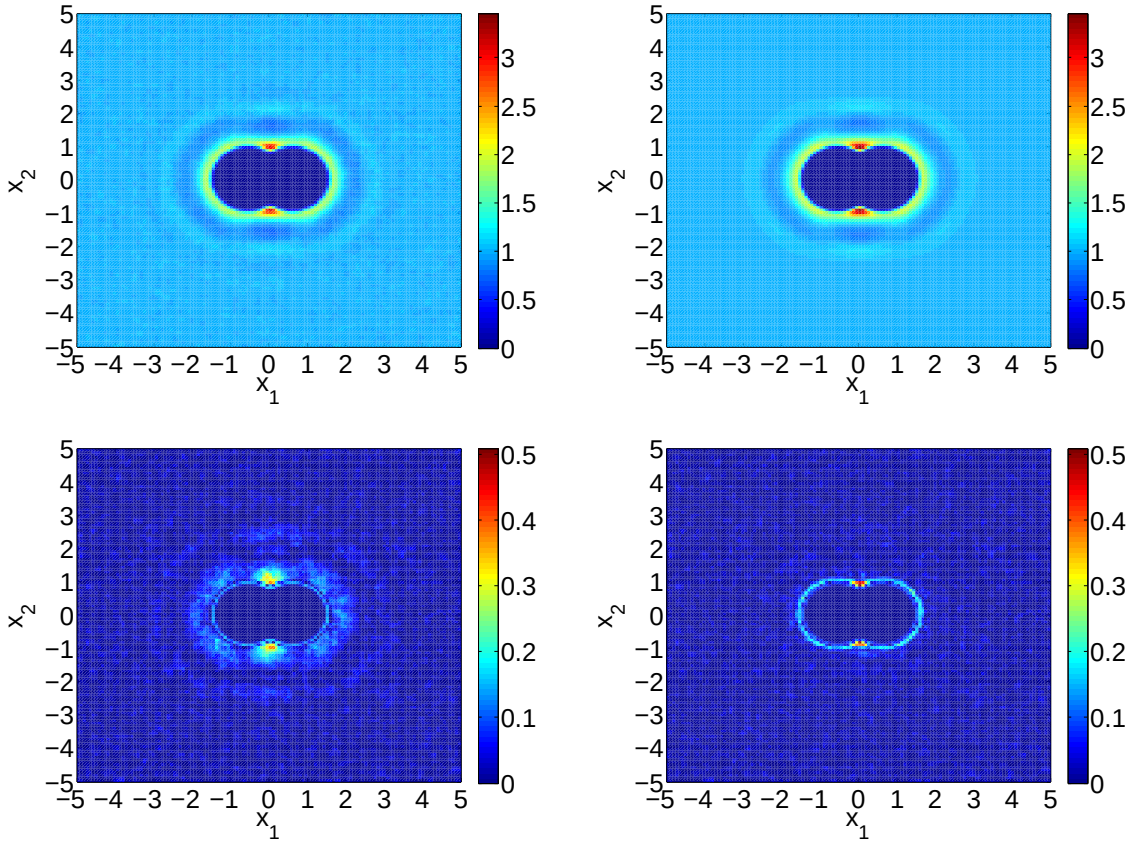

Figure 5.8. *Top: Solvent distribution for two solute particles computed with molecular dynamics (left) and the BGY3d method (right) at the $x_3 = 0$ plane. Bottom: Deviation of the BGY3d model (left) and deviation of the 3d-HNC method (right) at the $x_3 = 0$ plane.*

this first peak. This can be observed very well in the deviation plots of Figures 5.8, 5.9 and 5.10. The plots corresponding to BGY3d show a clear pattern behind the first peak but no dark red color. The 3d-HNC plots do not exhibit this pattern as clearly but feature the dark red color at the main peak.

Table 5.7 shows the computation times for all three methods. The great discrepancy of the computation times between systems with one and systems with more than one solute atoms stems from the different numerical treatments of the systems. In the case of molecular dynamics, we exploited the symmetry and improved the convergence by computing the pair distribution function for all pairs of particles of the solvent. This drastically reduces the computing time compared to the full three-dimensional density. In case of the BGY3d and 3d-HNC models, the solution

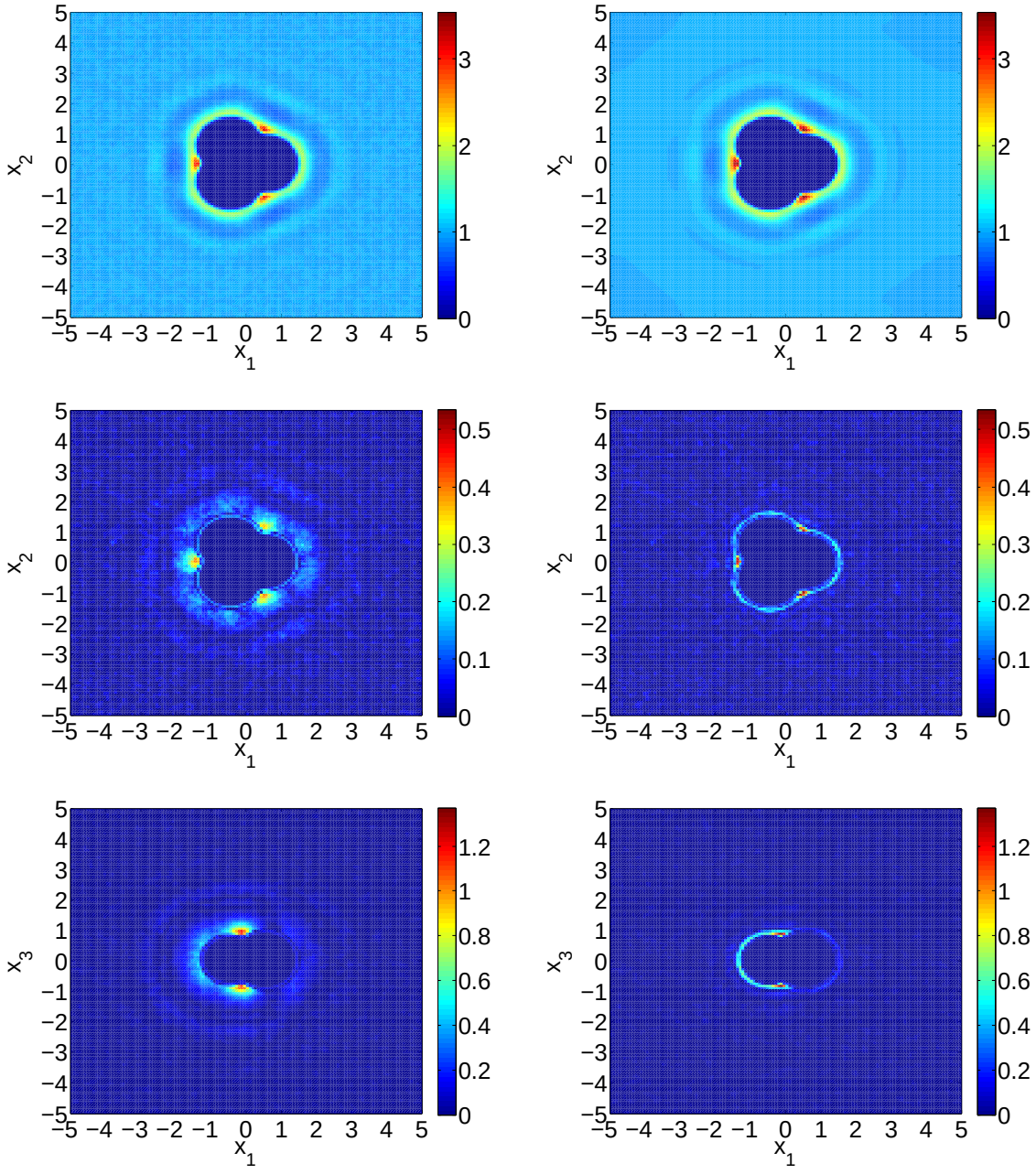

Figure 5.9. *Top: Solvent distribution for three solute particles computed with molecular dynamics (left) and the BGY3d method (right) at the $x_3 = 0$ plane. Middle: Deviation of the BGY3d model (left) and deviation of the 3d-HNC method (right) at the $x_3 = 0$ plane. Bottom: Deviation of the BGY3d model (left) and deviation of the 3d-HNC method (right) at the $x_2 = 0$ plane.*

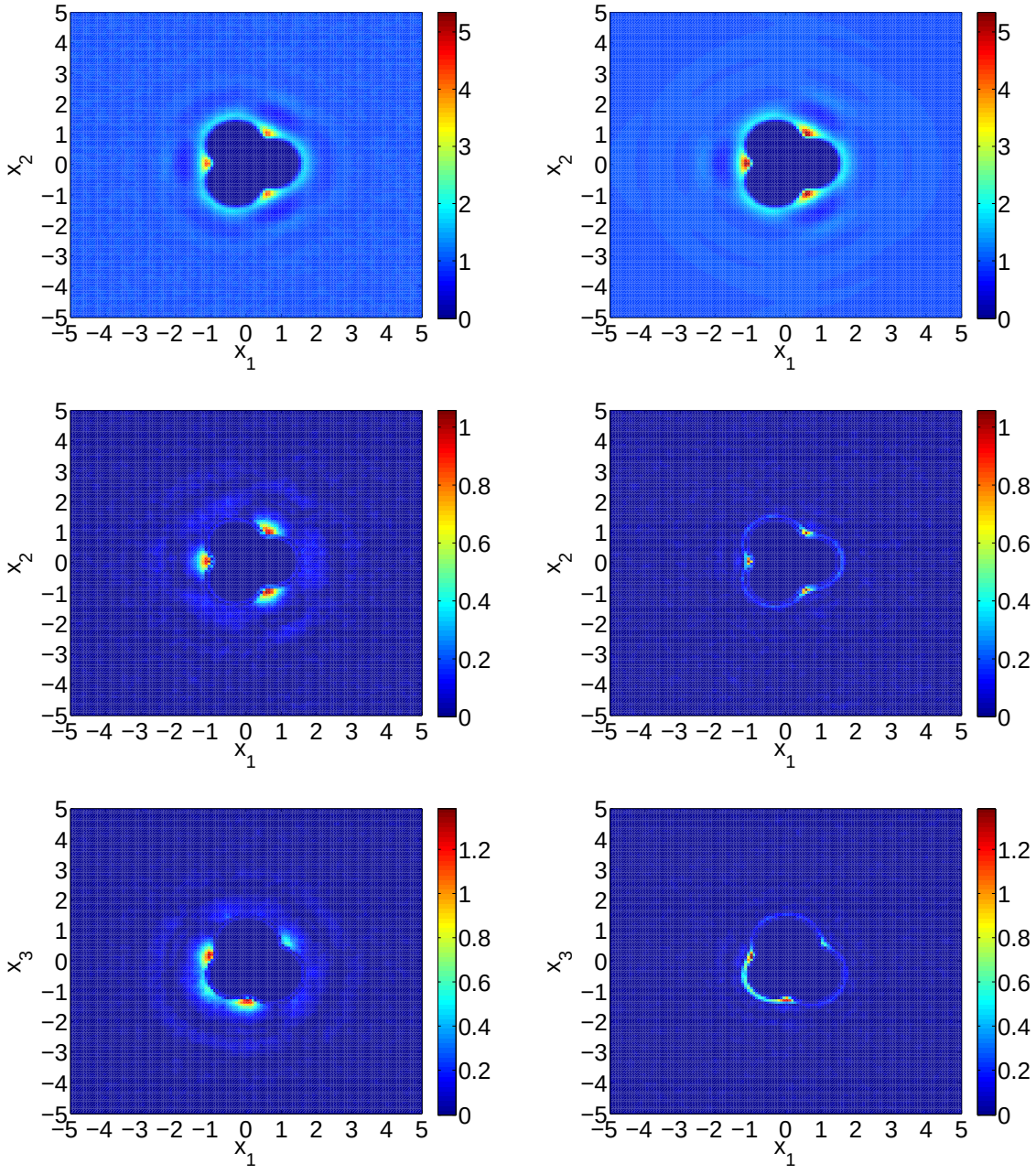

Figure 5.10. *Top: Solvent distribution for four solute particles computed with molecular dynamics (left) and the BGY3d method (right) at the $x_3 = 0$ plane. Middle: Deviation of the BGY3d model (left) and deviation of the 3d-HNC method (right) at the $x_3 = 0$ plane. Bottom: Deviation of the BGY3d model (left) and deviation of the 3d-HNC method (right) at the $x_2 = 0$ plane.*

		MD		BGY3d			3d-HNC		
ρ	N_M	steps	time	ν	steps	time	ν	steps	time
0.3	1	4.0_{+6}	$13200s$	0.9	26	$1951s$	0.9	58	$1397s$
0.5	1	4.0_{+6}	$26820s$	0.5	37	$1762s$	0.5	56	$1270s$
0.8	1	4.0_{+6}	$51780s$	0.3	119	$5624s$	0.1	168	$3998s$
0.5	2	$2 \times 1.4_{+8}$	$2 \times 269h$	0.5	32	$98s$	0.5	86	$225s$
0.5	3	$2 \times 1.4_{+8}$	$2 \times 270h$	0.5	38	$117s$	0.5	129	$331s$
0.5	4	$2 \times 1.4_{+8}$	$2 \times 272h$	0.5	40	$127s$	0.5	117	$304s$

Table 5.7. *Comparison of molecular dynamics, BGY3d and 3d-HNC with respect to computation time for systems with different solvent densities ρ and numbers of solute particles N_M. All simulations were performed on a single Intel(R) Xeon(TM) CPU 3.20GHz with six Gigabyte of RAM. Note that the solution for systems with one solute atom were computed with $N = 256^3$ grid points, whereas $N = 128^3$ grid points were employed for the other systems, see also the corresponding remarks in the text.*

for the systems with one solute atom were computed with $N = 256^3$ grid points, whereas $N = 128^3$ grid points were used for the other systems as this was the maximum number of grid points that was feasible with respect to the molecular dynamics simulation.

The computation time of the approximate models is significantly smaller for all systems considered. The difference is particularly high for two, three and four solute atoms. More than 22 days were necessary in order to reach a reasonable convergence of the density computed by the molecular dynamics simulations. This is four orders of magnitude longer than the numerical solution of our BGY3d model. A comparison of the computation times for the BGY3d and the 3d-HNC model shows that the BGY3d model is faster for the non-symmetric cases, whereas it is slower for systems with only one solute. This is caused by the different preconditions of the two models. The 3d-HNC model requires the direct correlation function of the pure solvent as input. This function can be precomputed very efficiently by a reduced solving of an one-dimensional equation. The time spent to solve this reduced equation is negligible compared to the computing time of the 3d-HNC method. Concerning the BGY3d model, the solution of systems with only one solute atom corresponds to the solution of the Born-Green equation, as already explained in Section 5.1. However, the solution of the Born-Green equation is numerically more expensive, since it requires three additional FFTs for every step of the fixed point iteration. A simplification of the three-dimensional convolution integrals in the Born-Green equation is also possible to some extent, if one exploits the symmetry. This way, the computing time could be reduced. But since we are most interested in the

computation of the full three-dimensional density distributions of the solvent, we do not further consider the simplification of the Born-Green equation. Details can be found e.g. in [58].

In practice, the relevant systems are those that feature more complex solutes, since the computation of the pair distribution function or the computation of the direct correlation function can be done beforehand once for any density. To this end, the BGY3d and 3d-HNC models are significantly less time consuming than the molecular dynamics simulation. They compute the distribution functions four orders of magnitudes faster. Their approximation error however is still non-negligible and will produce noticeable errors for any quantity which is derived from the computed solvent densities. Nevertheless, the development of such continuous models for the computation of the solvent density is an important step in the direction of more accurate implicit solvent models for molecular simulations. By this, it would be possible to approximate the solvent effects more accurately than by any classical implicit solvent model.

5.3 Numerical Solution of the BGY3dM Equations

In this section, we are going to discuss the algorithm to solve the molecular BGY3d (BGY3dM) equations numerically. Algorithm 5.1 presented in Section 5.1 for the BGY3d equation for monoatomic fluids has to be adjusted in order to cope with the additional terms incorporating the intramolecular constraints of the solvent molecules. Furthermore, we now have to compute all different site density distribution functions simultaneously, which means that we have to solve a system of coupled non-linear equations. Last but not least, we also want to consider long-range forces with the BGY3dM equations. Namely, we will consider polar molecules and therefore have to include the Coulomb potential in our model. This potential will require a special treatment, since it has a non-vanishing value at the boundaries of the computational domain, unlike the short-range potentials.

In order to simplify our discussion of the numerical solution of the BGY3dM model, we will restrict the presentation to solvents whose molecules consist only of two different particle species, which we call species A and species B. Hence, we have to compute three site-site pair distribution functions and two site distribution functions around a solute. We will consider two- and three-site models of fluids. The configuration of the rigid molecules is given by the distance r_0^{AB} in case of a two-site model or by the distance r_0^{AB} and the angle θ_{BAB} in case of a three-site model. For rigid three-site molecules, this obviously prescribes also the distance r_0^{BB} which we need for the BGY3dM equations. Further, every site of the molecule carries a charge q_{A} or q_{B} and is associated with a set of Lennard-Jones parameters ϵ_{A}, ϵ_{B} and σ_{A}, σ_{B}. The Lennard-Jones parameters for any pair of intermolecular sites is computed

according to the Lorentz-Berthelot mixing rules:

$$\epsilon_{\alpha\gamma} = \sqrt{\epsilon_\alpha \epsilon_\gamma}, \qquad \sigma_{\alpha\gamma} = \frac{\sigma_\alpha + \sigma_\gamma}{2}, \qquad \alpha, \gamma = \text{A, B.} \tag{5.58}$$

These parameters completely determine the respective model.

In the next section, we will write down explicitly the BGY3dM and SS-BGY3dM equations for a two-site solvent model, before we discuss the discretization of the equations. Then, we will explain how to include the long-range Coulomb potential into the model. Finally, we will test the SS-BGY3dM and BGY3dM model against results from molecular dynamics simulations.

5.3.1 BGY3dM Equations for a Two-Site Model

We now consider a fluid which consists of molecules with two different particle species. Hence, we have to compute two different site density distributions, g_A and g_B. In addition, the BGY3dM equations also include the three different site-site distribution functions of the pure fluid: $g^{(2)}_\text{AA}$, $g^{(2)}_\text{AB}$ and $g^{(2)}_\text{BB}$, which we have to compute beforehand. The number densities of the two species are equal, $\rho_\text{A} = \rho_\text{B} = \rho_S$, where ρ_S is the molecular number density of the solvent. The BGY3dM equations (4.114) for the two-site model solvent read as follows:

$$\begin{aligned}
\Delta_{\mathbf{x}_1} u_\text{A}(\mathbf{x}_1) = &-\beta\rho_\text{A}\nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_\text{AA}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}_\text{AA}(\mathbf{x}_1, \mathbf{x}_2) g_\text{A}(\mathbf{x}_2)\ d\mathbf{x}_2 \\
&-\beta\rho_\text{B}\nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_\text{AB}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}_\text{AB}(\mathbf{x}_1, \mathbf{x}_2) g_\text{B}(\mathbf{x}_2)\ d\mathbf{x}_2 \\
&-\Delta_{\mathbf{x}_1} \ln\left(\int_\Omega \omega_\text{AB}(\mathbf{x}_1, \mathbf{x}_2)\ \tilde{g}_{\text{B;A}}(\mathbf{x}_2)\ d\mathbf{x}_2\right), \qquad (5.59) \\
\Delta_{\mathbf{x}_1} u_\text{B}(\mathbf{x}_1) = &-\beta\rho_\text{A}\nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_\text{AB}(\mathbf{x}_2, \mathbf{x}_1) g^{(2)}_\text{AB}(\mathbf{x}_2, \mathbf{x}_1) g_\text{A}(\mathbf{x}_2)\ d\mathbf{x}_2 \\
&-\beta\rho_\text{B}\nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_\text{BB}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}_\text{BB}(\mathbf{x}_1, \mathbf{x}_2) g_\text{B}(\mathbf{x}_2)\ d\mathbf{x}_2 \\
&-\Delta_{\mathbf{x}_1} \ln\left(\int_\Omega \omega_\text{AB}(\mathbf{x}_2, \mathbf{x}_1) \tilde{g}_{\text{A;B}}(\mathbf{x}_2)\ d\mathbf{x}_2\right), \qquad (5.60)
\end{aligned}$$

with

$$g_\alpha(\mathbf{x}_1) = g^0_\alpha(\mathbf{x}_1) e^{-u_\alpha(\mathbf{x}_1)}, \quad g^0_\alpha(\mathbf{x}_1) = e^{-\beta V^M_\alpha(\mathbf{x}_1, \mathbf{x}^M)}, \quad \alpha = \text{A, B}, \tag{5.61}$$

$$\tilde{g}_{\alpha;\gamma}(\mathbf{x}_1) = \frac{g_\alpha(\mathbf{x}_1)}{\int_\Omega \omega_\text{AB}(\mathbf{x}_1, \mathbf{x}_2) g_\gamma(\mathbf{x}_2)\ d\mathbf{x}_2}, \quad \alpha, \gamma = \text{A, B}, \tag{5.62}$$

and

$$\omega_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_2) = \frac{\delta(r_{12} - r_0^{\mathrm{AB}})}{4\pi (r_0^{\mathrm{AB}})^2}, \qquad r_{12} = |\mathbf{x}_1 - \mathbf{x}_2|. \tag{5.63}$$

The function $V_\alpha^M(\mathbf{x}_1, \mathbf{x}^M)$ denotes the potential between the solute and the solvent particle $\alpha = \mathrm{A}, \mathrm{B}$. Here, we use a rather simplified notation where we number serially the position vectors $\mathbf{x}_i$ and skip the indices that indicate their particle type and molecule number, as this is redundant information. Similarly, we have to compute the three different site-site distribution functions according to the SS-BGY3dM equations (4.91):

$$\begin{aligned}
\Delta_{\mathbf{x}_1} u^{(2)}_{\mathrm{AA}}(\mathbf{x}_1, \mathbf{x}_2) = &- \beta\rho_{\mathrm{A}} \nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_{\mathrm{AA}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{AA}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{AA}}(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3 \\
&- \beta\rho_{\mathrm{B}} \nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{AB}}(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3 \\
&- \beta \nabla_{\mathbf{x}_1} \cdot \frac{\int_\Omega \mathbf{F}_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_3) \tilde{g}^{(2)}_{\mathrm{AB;A}}(\mathbf{x}_1, \mathbf{x}_3) \omega_{\mathrm{AB}}(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3}{n^{\mathrm{B}}_{\mathrm{AA}}(\mathbf{x}_1, \mathbf{x}_2)} \\
&- \Delta_{\mathbf{x}_1} \ln \left(\int_\Omega \omega_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_3) \tilde{g}^{(2)}_{\mathrm{AB;A}}(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3 \right),
\end{aligned} \tag{5.64}$$

$$\begin{aligned}
\Delta_{\mathbf{x}_1} u^{(2)}_{\mathrm{BB}}(\mathbf{x}_1, \mathbf{x}_2) = &- \beta\rho_{\mathrm{A}} \nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_{\mathrm{AB}}(\mathbf{x}_3, \mathbf{x}_1) g^{(2)}_{\mathrm{AB}}(\mathbf{x}_3, \mathbf{x}_1) g^{(2)}_{\mathrm{AB}}(\mathbf{x}_3, \mathbf{x}_2)\ d\mathbf{x}_3 \\
&- \beta\rho_{\mathrm{B}} \nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_{\mathrm{BB}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{BB}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{BB}}(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3 \\
&- \beta \nabla_{\mathbf{x}_1} \cdot \frac{\int_\Omega \mathbf{F}_{\mathrm{AB}}(\mathbf{x}_3, \mathbf{x}_1) \tilde{g}^{(2)}_{\mathrm{AB;B}}(\mathbf{x}_3, \mathbf{x}_1) \omega_{\mathrm{AB}}(\mathbf{x}_3, \mathbf{x}_2)\ d\mathbf{x}_3}{n^{\mathrm{A}}_{\mathrm{BB}}(\mathbf{x}_1, \mathbf{x}_2)} \\
&- \Delta_{\mathbf{x}_1} \ln \left(\int_\Omega \omega_{\mathrm{AB}}(\mathbf{x}_3, \mathbf{x}_1) \tilde{g}^{(2)}_{\mathrm{AB;B}}(\mathbf{x}_3, \mathbf{x}_2)\ d\mathbf{x}_3 \right),
\end{aligned} \tag{5.65}$$

$$\begin{aligned}
\Delta_{\mathbf{x}_1} u^{(2)}_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_2) = &- \beta\rho_{\mathrm{A}} \nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_{\mathrm{AA}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{AA}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{AB}}(\mathbf{x}_3, \mathbf{x}_2)\ d\mathbf{x}_3 \\
&- \beta\rho_{\mathrm{B}} \nabla_{\mathbf{x}_1} \cdot \int_\Omega \mathbf{F}_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}_{\mathrm{BB}}(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3 \\
&- \beta \nabla_{\mathbf{x}_1} \cdot \frac{\int_\Omega \mathbf{F}_{\mathrm{AA}}(\mathbf{x}_1, \mathbf{x}_3) \tilde{g}^{(2)}_{\mathrm{AA;B}}(\mathbf{x}_1, \mathbf{x}_3) \omega_{\mathrm{AB}}(\mathbf{x}_3, \mathbf{x}_2)\ d\mathbf{x}_3}{n^{\mathrm{A}}_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_2)} \\
&- \Delta_{\mathbf{x}_1} \ln \left(\int_\Omega \omega_{\mathrm{AB}}(\mathbf{x}_1, \mathbf{x}_3) \tilde{g}^{(2)}_{\mathrm{BB;A}}(\mathbf{x}_2, \mathbf{x}_3)\ d\mathbf{x}_3 \right),
\end{aligned} \tag{5.66}$$

with

$$
\begin{aligned}
g_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2) &= g^0_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2)e^{-u_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2)}, \quad g^0_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2) = e^{-\beta v_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2)},\\
\tilde{g}_{\alpha\gamma;\eta}(\mathbf{x}_1,\mathbf{x}_2) &= \frac{g_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2)}{n^\eta_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2)},\\
n^\eta_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2) &= \int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1,\mathbf{x}_3)g_{\gamma\eta}(\mathbf{x}_2,\mathbf{x}_3)\; d\mathbf{x}_2, \qquad\qquad \alpha,\gamma,\eta = \mathrm{A},\mathrm{B}
\end{aligned}
$$

and $\omega_{\mathrm{AB}}(\mathbf{x}_1,\mathbf{x}_2)$ as in (5.63). The solutions of equations (5.64), (5.65) and (5.66) can then be used as input for the BGY3dM equations (5.59) and (5.60).

5.3.2 Algorithmic Details

The numerical treatment of the non-linearity of the BGY3dM model is analogous to that of the BGY3d model for monoatomic solvents, see Section 5.1. The only difference is that we now have to deal with a system of non-linear equations. We again introduce a short notation, where we write the right hand sides of equations (5.59) and (5.60) simply as $K_{\mathrm{A}}(\mathbf{x}; g_{\mathrm{A}}, g_{\mathrm{B}})$ and $K_{\mathrm{B}}(\mathbf{x}; g_{\mathrm{A}}, g_{\mathrm{B}})$, respectively. To this end, the dependency on the density distributions of each species is explicitly indicated. Hence, equations (5.59) and (5.60) read in short notation as

$$\Delta_{\mathbf{x}} u_{\mathrm{A}}(\mathbf{x}) = K_{\mathrm{A}}(\mathbf{x}; g_{\mathrm{A}}, g_{\mathrm{B}}) \qquad \text{in } \Omega, \tag{5.67}$$

$$\Delta_{\mathbf{x}} u_{\mathrm{B}}(\mathbf{x}) = K_{\mathrm{B}}(\mathbf{x}; g_{\mathrm{A}}, g_{\mathrm{B}}) \qquad \text{in } \Omega. \tag{5.68}$$

As in the case of the BGY3d equation, this system of non-linear equations is transformed into a series of coupled linear equations by applying the fixed point iteration, see Algorithm 5.2.

ALGORITHM 5.2 ((Damped) Fixed Point Iteration of the BGY3dM Equations).

1. $u^0_{\mathsf{A}} = 0$; $u^0_{\mathsf{B}} = 0$; $g^0_{\mathsf{A}} = e^{-\beta V^M_{\mathsf{A}}}$; $g^0_{\mathsf{B}} = e^{-\beta V^M_{\mathsf{B}}}$; $l = 0$;
2. $l \leftarrow l + 1$; Solve

$$\Delta_{\mathbf{x}} u^l_{\mathsf{A}}(\mathbf{x}) = K_{\mathsf{A}}(\mathbf{x}; g^{l-1}_{\mathsf{A}}, g^{l-1}_{\mathsf{B}}) \qquad \text{in } \Omega \tag{5.69}$$

and set

$$
\begin{aligned}
u^l_{\mathsf{A}} &\leftarrow \nu u^l_{\mathsf{A}} + (1-\nu)u^{l-1}_{\mathsf{A}},\\
g^l_{\mathsf{A}} &= g^0_{\mathsf{A}}\; e^{-u^l_{\mathsf{A}}}.
\end{aligned} \tag{5.70}
$$

3. Solve

$$\Delta_{\mathbf{x}} u^l_{\mathsf{B}}(\mathbf{x}) = K_{\mathsf{B}}(\mathbf{x}; g^l_{\mathsf{A}}, g^{l-1}_{\mathsf{B}}) \qquad \text{in } \Omega \tag{5.71}$$

and set

$$u_{\mathsf{B}}^{l} \leftarrow \nu u_{\mathsf{B}}^{l} + (1-\nu) u_{\mathsf{B}}^{l-1}, \qquad (5.72)$$
$$g_{\mathsf{B}}^{l} = g_{\mathsf{B}}^{0}\, e^{-u_{\mathsf{B}}^{l}}.$$

4. If

$$||u_{\mathsf{A}}^{l} - u_{\mathsf{A}}^{l-1}||_{L_\infty} < \nu\chi \qquad (5.73)$$

and

$$||u_{\mathsf{B}}^{l} - u_{\mathsf{B}}^{l-1}||_{L_\infty} < \nu\chi, \qquad (5.74)$$

stop; else go to 2.

Again, ν is a damping parameter which ensures convergence. The iteration is stopped if the L_∞-norm of the change between successive iterations is smaller than the fixed threshold $\nu\chi$. The discussion about the correct definition of the boundary conditions is postponed to Section 5.3.3. For now, we again employ periodic boundary conditions.

As in the monoatomic case, the linearized equations (5.70) and (5.72) can be solved easily in Fourier space. Due to the superposition principle, the contributions of each integral term on the right hand sides of (5.59) and (5.60) can be computed separately. The first two terms corresponding to the intermolecular interactions are computed exactly as in the monoatomic case, see Section 5.1. The remaining terms, corresponding to the intramolecular interactions, require a convolution of a density distribution function with a delta distribution. This is also computed in Fourier space. To this end, the Fourier transformation of the delta distribution is done analytically. It holds that

$$\mathcal{F}_3\left(\delta(r_{12} - r_0^{\alpha\gamma})\right)(\mathbf{k}) = \frac{2}{|\mathbf{k}|}\sin(2\pi|\mathbf{k}|r_0^{\alpha\gamma})r_0^{\alpha\gamma}, \qquad (5.75)$$

where we exploited the spherical symmetry of the delta distribution, see also Appendix A. Hence, we have

$$\mathcal{F}_3(\omega_{\alpha\gamma})(\mathbf{k}) = \frac{\sin(2\pi|\mathbf{k}|r_0^{\alpha\gamma})}{2\pi|\mathbf{k}|r_0^{\alpha\gamma}} \qquad (5.76)$$

and we can compute the whole intramolecular term as

$$\ln\left(\int_\Omega \omega_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2)\tilde{g}_{\alpha;\gamma}(\mathbf{x}_2)\; d\mathbf{x}_2\right) = \ln\left[\mathcal{F}_3^{-1}\left(\frac{\sin(2\pi|\mathbf{k}|r_0^{\alpha\gamma})}{2\pi|\mathbf{k}|r_0^{\alpha\gamma}}\mathcal{F}_3(\tilde{g}_{\alpha;\gamma})\right)\right]. \qquad (5.77)$$

The Laplacian present in front of this term in the BGY3dM equations is compensated by the application of the inverse of the Laplacian in Fourier space.

Additionally, we have to compute the normalized distribution function $\tilde{g}_{\alpha;\gamma}$. To this end, we can compute the normalization function similar to (5.77) by

$$\begin{aligned} n_\alpha^\gamma(\mathbf{x}_1) &= \int_\Omega \omega_{\alpha\gamma}(\mathbf{x}_1, \mathbf{x}_2) g_\gamma(\mathbf{x}_2)\, d\mathbf{x}_2 \\ &= \mathcal{F}_3^{-1}\left(\frac{\sin(2\pi|\mathbf{k}|r_0^{\alpha\gamma})}{2\pi|\mathbf{k}|r_0^{\alpha\gamma}} \mathcal{F}_3(g_\gamma) \right)(\mathbf{x}_1). \end{aligned} \tag{5.78}$$

Then, g_α has to be divided by n_α^γ. Since the distribution functions are strictly non-negative, this also holds for the normalization function. However, n_α^γ can be very small for some $\mathbf{x}_1 \in \Omega$. This may lead to numerical instabilities and makes a regularization of this term necessary. Therefore, we introduce a regularization parameter ϵ_g and compute the division by

$$\frac{g_\alpha(\mathbf{x}_1)}{n_\alpha^\gamma(\mathbf{x}_1)} \approx \frac{g_\alpha(\mathbf{x}_1)}{\max(n_\alpha^\gamma(\mathbf{x}_1), \epsilon_g)}, \qquad \forall\ \mathbf{x}_1 \in \Omega. \tag{5.79}$$

In the numerical tests, the choice $\epsilon_g = 10^{-2}$ showed to be the optimal one, since it produces a numerically stable method and leads to negligible differences compared to smaller values. Hence, this value has been used for any results computed with the BGY3dM equations.

Computation of the Site-Site Distribution Functions

The three site-site distribution functions of the diatomic molecular fluid are required as input of the BGY3dM equations (5.59) and (5.60). They are computed as the solutions of the SS-BGY3dM equations (5.64), (5.65) and (5.66). These equations are treated analogously to the BGY3dM equations. The only difference lies in the additional integral term which also models an intramolecular part of the interaction. In short notation, the SS-BGY3dM equations read as

$$\Delta_{\mathbf{x}} u_{\mathrm{AA}}(\mathbf{x}) = K_{\mathrm{AA}}(\mathbf{x}; g_{\mathrm{AA}}, g_{\mathrm{AB}}) \qquad \text{in } \Omega, \tag{5.80}$$

$$\Delta_{\mathbf{x}} u_{\mathrm{BB}}(\mathbf{x}) = K_{\mathrm{BB}}(\mathbf{x}; g_{\mathrm{BB}}, g_{\mathrm{AB}}) \qquad \text{in } \Omega, \tag{5.81}$$

$$\Delta_{\mathbf{x}} u_{\mathrm{AB}}(\mathbf{x}) = K_{\mathrm{AB}}(\mathbf{x}; g_{\mathrm{AA}}, g_{\mathrm{BB}}, g_{\mathrm{AB}}) \qquad \text{in } \Omega. \tag{5.82}$$

Again, a fixed point iteration with a damping parameter ν and a parameter χ for the stopping criterion is employed in order to solve the equations, see Algorithm 5.3. Periodic boundary conditions are assumed for the linearized equations until their correct definition in Section 5.3.3.

ALGORITHM 5.3 ((Damped) Fixed Point Iteration of the SS-BGY3dM Equations).

1. $u^0_{\mathsf{AA}} = 0$; $u^0_{\mathsf{BB}} = 0$; $u^0_{\mathsf{AB}} = 0$;
   $g^0_{\mathsf{AA}} = e^{-\beta V_{\mathsf{AA}}}$; $g^0_{\mathsf{BB}} = e^{-\beta V_{\mathsf{BB}}}$; $g^0_{\mathsf{AB}} = e^{-\beta V_{\mathsf{AB}}}$;
   $l = 0$;

2. $l \leftarrow l + 1$; Solve

$$\Delta_{\mathbf{x}_1} u^l_{\mathsf{AA}}(\mathbf{x}) = K_{\mathsf{AA}}(\mathbf{x}; g^{l-1}_{\mathsf{AA}}, g^{l-1}_{\mathsf{AB}}) \qquad \text{in } \Omega \tag{5.83}$$

   and set

$$\begin{aligned} u^l_{\mathsf{AA}} &\leftarrow \nu u^l_{\mathsf{AA}} + (1-\nu) u^{l-1}_{\mathsf{AA}}, \\ g^l_{\mathsf{AA}} &= g^0_{\mathsf{AA}}\, e^{-u^l_{\mathsf{AA}}}. \end{aligned} \tag{5.84}$$

3. Solve

$$\Delta_{\mathbf{x}} u^l_{\mathsf{BB}}(\mathbf{x}) = K_{\mathsf{BB}}(\mathbf{x}; g^{l-1}_{\mathsf{BB}}, g^{l-1}_{\mathsf{AB}}) \qquad \text{in } \Omega \tag{5.85}$$

   and set

$$\begin{aligned} u^l_{\mathsf{BB}} &\leftarrow \nu u^l_{\mathsf{BB}} + (1-\nu) u^{l-1}_{\mathsf{BB}}, \\ g^l_{\mathsf{BB}} &= g^0_{\mathsf{BB}}\, e^{-u^l_{\mathsf{BB}}}. \end{aligned} \tag{5.86}$$

4. Solve

$$\Delta_{\mathbf{x}} u^l_{\mathsf{AB}}(\mathbf{x}) = K_{\mathsf{AB}}(\mathbf{x}; g^l_{\mathsf{AA}}, g^l_{\mathsf{BB}}, g^{l-1}_{\mathsf{AB}}) \qquad \text{in } \Omega \tag{5.87}$$

   and set

$$\begin{aligned} u^l_{\mathsf{AB}} &\leftarrow \nu u^l_{\mathsf{AB}} + (1-\nu) u^{l-1}_{\mathsf{AB}}, \\ g^l_{\mathsf{AB}} &= g^0_{\mathsf{AB}}\, e^{-u^l_{\mathsf{AB}}}. \end{aligned} \tag{5.88}$$

5. If

$$||u^l_{\mathsf{AA}} - u^{l-1}_{\mathsf{AA}}||_{L_\infty} < \nu\chi \tag{5.89}$$

   and

$$||u^l_{\mathsf{BB}} - u^{l-1}_{\mathsf{BB}}||_{L_\infty} < \nu\chi \tag{5.90}$$

   and

$$||u^l_{\mathsf{AB}} - u^{l-1}_{\mathsf{AB}}||_{L_\infty} < \nu\chi \tag{5.91}$$

   stop; else go to 2.

As for the BGY3dM equations, the different integral terms of the right hand sides can be computed separately. The intermolecular integral terms are treated analogously to the Born-Green equation in the monoatomic case, see Section 5.1. The terms of type

$$\ln\left(\int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1,\mathbf{x}_3)\tilde{g}_{\gamma\eta;\alpha}(\mathbf{x}_2,\mathbf{x}_3)\ d\mathbf{x}_3\right)$$

have already been discussed above. The computation of the intramolecular terms including the intermolecular force requires some additional considerations. Terms of type

$$\Delta_{\mathbf{x}_1}^{-1}\nabla_{\mathbf{x}_1}\cdot\frac{\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1,\mathbf{x}_3)\tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}_1,\mathbf{x}_3)\omega_{\gamma\eta}(\mathbf{x}_2,\mathbf{x}_3)\ d\mathbf{x}_3}{n^{\eta}_{\alpha\gamma}(\mathbf{x}_1,\mathbf{x}_2)} \tag{5.92}$$

have to be computed explicitly. Since we do not want to further simplify (5.92), it has to be computed in a step by step procedure. This is numerically quite expensive but leads to the best results. First, $\tilde{g}^{(2)}_{\alpha\eta;\gamma}$ has to be computed as described above, see equation (5.79). Then, the denominator and the divisor of (5.92) can be computed. In case of the denominator, this is done by

$$\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1,\mathbf{x}_3)\tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}_1,\mathbf{x}_3)\omega_{\gamma\eta}(\mathbf{x}_2,\mathbf{x}_3)\ d\mathbf{x}_3 = \mathcal{F}_3^{-1}\left(\mathcal{F}_3(\mathbf{F}_{\alpha\eta}g^{(2)}_{\alpha\eta;\gamma})\frac{\sin(2\pi|\mathbf{k}|r_0^{\gamma\eta})}{2\pi|\mathbf{k}|r_0^{\gamma\eta}}\right). \tag{5.93}$$

The subsequent division again requires a regularization for the same reasons as already described above. However, since a differential operator has to be applied to the result of the division, this case is even more difficult. The operator is applied by a multiplication of $\frac{k_i}{2\pi|\mathbf{k}|^2}$ in direction $i = 1, 2, 3$ in Fourier space. Hence, a Fourier transform of the result of the division is required prior to the multiplication of the factor. The Fourier components decay faster if the function is smooth in real space. However, the division by the normalization function can lead to very rough results if the normalization functions contains small values. This may result in large errors, since the Fourier components of a rough function do not decay fast, and this again leads to large errors at the boundaries. This is why we choose a new regularization constant $\epsilon_\omega = 10^{-1}$ with a relatively large value and replace the exact division by

$$\frac{\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{r}-\mathbf{r}')\tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{r}-\mathbf{r}')\omega_{\gamma\eta}(\mathbf{r}')\ d\mathbf{r}'}{\max(n^{\eta}_{\alpha\gamma}(\mathbf{r}),\epsilon_\omega)}, \qquad \forall\ \mathbf{r}\in\Omega. \tag{5.94}$$

Numerical tests revealed that this choice produces a stable method with an error that is small compared to the approximation error, as we will see in Section 5.4.

As the last step of computing the term (5.92), the factor $\frac{k_i}{2\pi|\mathbf{k}|^2}$ is multiplied in Fourier space for each direction $i = 1, 2, 3$ and the result is summed up and transformed back to real space. This three-step procedure to compute the intramolecular

term is necessary in order to apply the full NSSA approximation (4.104). It is numerically more expensive but provides good results when compared to pair distribution functions computed with molecular dynamics, see Section 5.4.

5.3.3 Treatment of the Coulomb Potential

So far, we restricted our discussion to the Lennard-Jones potential, which is a short-range potential. The short-range potential leads to short-range distribution functions and hence, all integral terms are also of short range. This has been an important prerequisite to restrict the computation of the BGY3d equations to a finite domain, see Section 5.1.2. Now, we are going to discuss how to include a potential function that is not of short range, i.e., the potential decays more slowly than $\frac{1}{r^3}$, with r being the distance in three-dimensional space. Namely, we want to consider the Coulomb potential which decays as $\frac{1}{r}$. This potential is very important when dealing with biomolecular applications, since the partial charges of realistic solvent molecules cannot be neglected. Important examples are water (H_2O) or all alcohols (CH_3OH, C_2H_5OH, ...). However, the distribution functions remain short-ranged even though a long-range interaction is employed. In the case of Coulomb systems, this can be shown rigorously [1]. The screening effect of charges provides a more descriptive explanation provides: In a fluid, negatively charged particles tend to gather next to a positively charged particle and vice versa. Hence, no clusters of particles of the same type are built and all charges are roughly equally distributed. Hence, there is no net force on the particles caused by particles at large distances. This leads to distribution functions that are of short range, also in Coulomb systems.

Nevertheless, the incorporation of the Coulomb potential requires some changes in the numerical computation of the respective integral terms. This affects all terms that contain the intermolecular force. Since these terms differ only in one of the distribution functions included, they can all be treated in the same way. Hence, we discuss the problem exemplarily on the basis of the following (notationally simplified) integral

$$\int_\Omega \mathbf{F}(\mathbf{r}' - \mathbf{r}) g_\alpha(\mathbf{r}' - \mathbf{r}) g_\gamma(\mathbf{r}') \, d\mathbf{r}' \tag{5.95}$$

with the total force

$$\mathbf{F} = \mathbf{F}^{LJ} + \mathbf{F}^{C}$$

consisting of a Lennard-Jones and a Coulomb part. The two distribution functions g_α and g_γ are arbitrary and can in particular also be identified with intramolecular distribution functions.

The main difficulty in computing (5.95) is caused by the application of the discrete Fourier transform in order to compute the convolution integral (5.95). The discrete Fourier transform assumes the functions to be periodic with respect to the computational domain Ω. Functions that decay to zero at the boundaries of the

domain satisfy this condition, as is the case for the product of the Lennard-Jones force and a pair distribution function. However, the product $\mathbf{F}g_\alpha$ does not vanish at the boundary if a long-range force is employed. Moreover, the components of the force vector are antisymmetric, so that the periodically continued function exhibits a jump at the boundary. Hence, the use of the discrete Fourier transform is prohibitive in this situation. Yet, we can transform the integral in order to circumvent this problem. The following idea is borrowed from the particle-particle–particle-mesh method (P^3M) [52] and the particle-mesh-Ewald method (PME) [26] that have been developed for the efficient computation of long-range forces in molecular dynamics or Monte Carlo simulations. According to these concepts, the Coulomb force can be divided into a part which exhibits a singularity at zero distance but is short-range, and a part which is smooth and of long range. This is achieved by shielding the point charge by a charge distribution. This charge distribution is chosen to be Gaussian,

$$\varrho(\mathbf{r}) = \left(\frac{G}{\sqrt{\pi}}\right)^3 e^{-G^2|\mathbf{r}|^2} \tag{5.96}$$

with a parameter G that determines the width of the function. The special form of the charge distribution is chosen, such that it features fast decaying Fourier components. We will take advantage of this fact later. For now, it is enough to know that the Coulomb potential (and its force) between two particles is divided by means of this shielding function into the following parts

$$\begin{aligned} v^C(\mathbf{r}) &= v^{Cs}(\mathbf{r}) + v^{Cl}(\mathbf{r}) = q_b\Phi^s(\mathbf{r}) + q_b\Phi^l(\mathbf{r}), \\ \mathbf{F}^C(\mathbf{r}) &= \mathbf{F}^{Cs}(\mathbf{r}) + \mathbf{F}^{Cl}(\mathbf{r}) = -\nabla v^{Cs}(\mathbf{r}) - \nabla v^{Cl}(\mathbf{r}) \end{aligned} \tag{5.97}$$

with Φ^s and Φ^l the solutions of the Poisson equations

$$\begin{aligned} -\Delta\Phi^s &= \frac{1}{\epsilon_0} q_a(\delta_3 - \varrho) \qquad \text{in } \mathbb{R}^3, \\ -\Delta\Phi^l &= \frac{1}{\epsilon_0} q_a\varrho \qquad\qquad \text{in } \mathbb{R}^3 \end{aligned} \tag{5.98}$$

where q_a and q_b denote the charges of the two particles and ϵ_0 is the dielectric constant. For the special choice of the function ϱ, the solution can be given analytically. We have

$$v^{Cs}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} q_a q_b \frac{\operatorname{erfc}(G|\mathbf{r}|)}{|\mathbf{r}|}, \tag{5.99}$$

$$v^{Cl}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} q_a q_b \frac{\operatorname{erf}(G|\mathbf{r}|)}{|\mathbf{r}|} \tag{5.100}$$

with erf the error function and erfc $= 1 -$ erf the complementary error function. Figure 5.11 shows a plot of both functions $\frac{\operatorname{erf}(|\mathbf{r}|)}{|\mathbf{r}|}$ and $\frac{\operatorname{erfc}(|\mathbf{r}|)}{|\mathbf{r}|}$ compared to $\frac{1}{|\mathbf{r}|}$ and

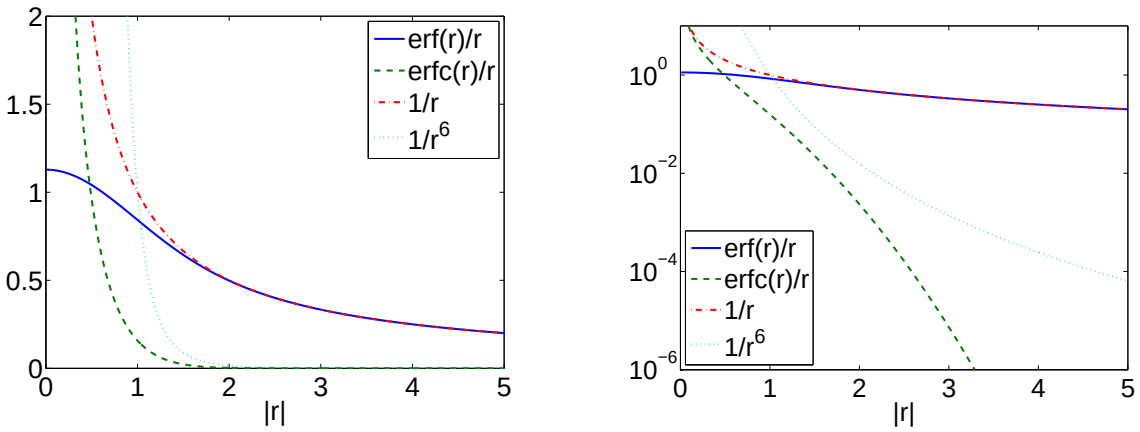

Figure 5.11. *The functions* $\frac{erf(|\mathbf{r}|)}{|\mathbf{r}|}$, $\frac{erfc(|\mathbf{r}|)}{|\mathbf{r}|}$, $\frac{1}{|\mathbf{r}|}$, *and* $\frac{1}{|\mathbf{r}|^6}$ *in a linear plot (left) and a semi-logarithmic plot (right).*

$\frac{1}{|\mathbf{r}|^6}$. The complementary error function decays rapidly, even faster than $\frac{1}{|\mathbf{r}|^6}$. On the other hand, the error function decays as slowly as $\frac{1}{|\mathbf{r}|}$ but is smooth even at $|\mathbf{r}| = 0$.

We now want to use these properties and transform the integral (5.95) in order to make it efficiently computable. The total force $\mathbf{F}$ consists of a Lennard-Jones part $\mathbf{F}^{LJ}$ and a Coulomb part $\mathbf{F}^{C}$ which can be splitted as discussed above. Hence, we transform the first part of the convolution integral (5.95) according to

$$\begin{aligned}\mathbf{F}g_\alpha &= (\mathbf{F}^{LJ} + \mathbf{F}^{Cs} + \mathbf{F}^{Cl})\ g_\alpha \\ &= (\mathbf{F}^{LJ} + \mathbf{F}^{Cs})\ g_\alpha + \mathbf{F}^{Cl}\ (g_\alpha - 1 + 1) \\ &= \left[(\mathbf{F}^{LJ} + \mathbf{F}^{Cs} + \mathbf{F}^{Cl})\ g_\alpha - \mathbf{F}^{Cl}\right] + \mathbf{F}^{Cl}. \end{aligned} \tag{5.101}$$

For the entire integral, this leads to

$$\begin{aligned}&\int_\Omega \mathbf{F}(\mathbf{r}' - \mathbf{r}) g_\alpha(\mathbf{r}' - \mathbf{r}) g_\gamma(\mathbf{r}')\ d\mathbf{r}' \\ &\qquad = \int_\Omega \left(\mathbf{F}(\mathbf{r}' - \mathbf{r}) g_\alpha(\mathbf{r}' - \mathbf{r}) - \mathbf{F}^{Cl}(\mathbf{r}' - \mathbf{r})\right) g_\gamma(\mathbf{r}')\ d\mathbf{r}' \\ &\qquad + \int_\Omega \mathbf{F}^{Cl}(\mathbf{r}' - \mathbf{r}) g_\gamma(\mathbf{r}')\ d\mathbf{r}'. \end{aligned} \tag{5.102}$$

The first term can be treated as before, since the part in outer brackets is of short range if g_α is also short-ranged. The second integral can be written as

$$\int_\Omega \mathbf{F}^{Cl}(\mathbf{r}' - \mathbf{r}) g_\gamma(\mathbf{r}')\ d\mathbf{r}' = -\nabla_{\mathbf{r}} \int_\Omega v^{Cl}(\mathbf{r}' - \mathbf{r}) g_\gamma(\mathbf{r}')\ d\mathbf{r}', \tag{5.103}$$

where we used again the fact that the derivative of a convolution can be shifted to its arguments. Now, we can take advantage of the rapid decay of the Fourier components of v^{Cl}. The Fourier transformation can even be given analytically as

$$\mathcal{F}_3(v^{Cl})(\mathbf{k}) = \frac{q_a q_b}{\epsilon_0} e^{-\frac{\pi^2}{G^2}|\mathbf{k}|^2}. \tag{5.104}$$

Hence, this integral can be computed easily by multiplying the Fourier components of v^{Cl} and g_γ and a subsequent inverse Fourier transformation of the result. The operator $\nabla_{\mathbf{r}}$ in front of the integral is neutralized by the inverse operator on the left hand side of the SS-BGY3dM or BGY3dM equations and therefore does not need to be computed explicitly.

There are also intramolecular terms which involve the long-range Coulomb force comprised in the SS-BGY3dM equations. Since the NSSA approximation is used in this case, the differential operators do not cancel out any more and have to be computed explicitly. The long-range part of these terms is then computed as

$$\Delta_{\mathbf{r}}^{-1}\nabla_{\mathbf{r}} \cdot \frac{\nabla_{\mathbf{r}} \int_\Omega v^{Cl}(\mathbf{r}-\mathbf{r}')\omega(\mathbf{r}')\, d\mathbf{r}'}{\max(n(\mathbf{r}), \epsilon_\omega)}, \qquad \forall \mathbf{r} \in \Omega, \tag{5.105}$$

where the analytical Fourier components of v^{Cl} and ω are used to compute the convolution.

Cancellation of the Long-Range Parts

Even though we know the exact Fourier components of the long-range part of the Coulomb potential v^{Cl}, we add an error source by computing the discrete inverse Fourier transformation of a long-range function. Here, it is assumed that the product $\tilde{v}^{Cl}\tilde{g}_\gamma$ of the Fourier components of v^{Cl} and g_γ is periodic with respect to the computational domain. Numerically, this is not a problem if the Fourier components decay to zero at the boundaries. Nevertheless, the assumption of periodicity produces an error at the boundaries for any convolution of a distribution function with the long-range part of the Coulomb potential. Recall that the distribution functions are of short range even in that case, see [1]. Hence, the sum of the long-range parts of the different terms in the BGY3dM or SS-BGY3dM equations has to vanish. It follows that we have to pay attention when choosing the computational order especially for our decomposition of the solution

$$g(\mathbf{x}) = g^0(\mathbf{x})e^{-u(\mathbf{x})},$$

since the function u is of long range as well as

$$g^0(\mathbf{x}) = e^{-\beta(v^{LJ}(\mathbf{x})+v^{Cs}(\mathbf{x})+v^{Cl}(\mathbf{x}))}.$$

Multiplication of both terms, as suggested by the decomposition, would lead to large errors, and the short-range nature of the solution could not be guaranteed. Instead, a different decomposition is advantageous when long-range potentials are involved:

$$g(\mathbf{x}) = \tilde{g}^0(\mathbf{x}) e^{-\beta v^{Cl}(\mathbf{x}) - u(\mathbf{x})} \quad \text{with} \quad \tilde{g}^0(\mathbf{x}) = e^{-\beta(v^{LJ}(\mathbf{x}) + v^{Cs}(\mathbf{x}))}. \tag{5.106}$$

Now, the two long-range functions v^{Cl} and u directly sum up and the result is only of short-range. By this, we do not change the SS-BGY3dM or BGY3dM equations, as they are given in (4.118) and (4.114), respectively. We only change the order of computation of the distribution function g such that it is numerically more stable. However, we still have to be careful. It cannot be guaranteed that the initial guess u^0 of the fixed point iteration of the SS-BGY3dM or BGY3dM equations fulfills the short-range condition, i.e., that the long-range parts of u^0 and v^{Cl} cancel out. The result would be a distribution function that is not of short range, which in turn would produce large errors in the next iteration step. Therefore, we have to enforce the correct boundary conditions in order to guarantee that the fixed point iteration converges and leads to a short-range solution.

Boundary Conditions

For the short-range potentials, we could choose the finite domain such that periodic and Dirichlet boundary conditions lead to identical results for the linearized problems in the fixed point iteration. Things get more complicated if the long-range Coulomb potential is employed. As discussed above, the long-range parts of the solution of the Poisson problems appearing in Algorithms 5.2 and 5.3 have to cancel out with the long-range parts of v^{Cl} in order to give a function that vanishes at the boundaries. Since this short-range condition is not guaranteed for the iterates u^l, we have to enforce the Dirichlet boundary conditions in order to stabilize the iteration.

If we omit all indices for particle type and iteration count, the Poisson problem arising in each step of our Algorithms 5.2 and 5.3 can be written as

$$\Delta\tilde{u} = K(g) + \beta\Delta v^{Cl} \qquad \text{in } \Omega, \qquad \tilde{u}(\partial\Omega) = 0, \tag{5.107}$$

where we introduced the long-range potential v^{Cl} into the differential equation in order to get zero boundary conditions. Hence, we have chosen $\tilde{u}$ such that

$$g(\mathbf{x}) = \tilde{g}^0(\mathbf{x}) e^{-\tilde{u}(\mathbf{x})} \tag{5.108}$$

with $\tilde{g}^0$ from (5.106). The solution $\tilde{u}$ can also be represented by a difference of two functions $\tilde{u} = \bar{u} - u^*$ which are also solutions of two Laplace equations

$$\begin{aligned} \Delta\bar{u} &= K(g) + \beta\Delta v^{Cl} && \text{in } \Omega, && \bar{u}(\partial\Omega) = f, \\ \Delta u^* &= 0 && \text{in } \Omega, && u^*(\partial\Omega) = f, \end{aligned}$$

with some function f at the boundary. The solution of the first line $\bar{u}$ can be identified with $u + \beta v^{Cl}$, where u is the solution computed by diagonal scaling in Fourier space. To this end, we employ periodic boundary conditions which correspond to some Dirichlet boundary conditions, represented by the function f in this case. If we now subtract u^*, the solution of

$$\Delta u^* = 0 \qquad \text{in } \Omega, \qquad u^*(\partial\Omega) = \bar{u}(\partial\Omega), \tag{5.109}$$

we get the solution of (5.107) with the correct boundary conditions,

$$\tilde{u}(\partial\Omega) = \bar{u}(\partial\Omega) - u^*(\partial\Omega) = 0. \tag{5.110}$$

Finally, we can compute the distribution function by (5.108).

The procedure described above seems very circuitous, but it has turned out to stabilize the non-linear iteration very well. Combined with a small damping factor ν, it leads to short-range distribution functions for the SS-BGY3dM as well as for the BGY3dM equations, as we will see later. This separated correction of the boundary conditions even makes it possible to keep the fast solution of the Laplacian in Fourier space, even though the involved long-range functions are not exactly periodic. This causes an error which is localized near the boundaries, assuming that the distribution functions are of short range. This error is removed by subtracting the solution of (5.109).

Equation (5.109) is solved by a standard finite difference discretization with a seven point stencil. This way, we can use the same grid as for the Poisson equation (5.107). The resulting systems of equations are solved by the iterative GMRES method with Block Jacobi preconditioning, as it is implemented in PETSc [5–7].

Instead of solving a second Laplace problem in order to correct the boundary conditions, we also could have chosen to solve the original Poisson equation (5.107) by a finite difference scheme. Then, a correction of the boundary conditions would not be necessary. However, the separation into two problems has a distinct advantage. The computation of the right-hand side of (5.107) is done by means of Fourier transformations. Hence, the additional diagonal scaling to solve for the Laplace in Fourier space can be done with linear complexity. The solution of the Laplace problem (5.109) with zero right-hand side, however, becomes easier as the fixed point iteration proceeds. This is because the change of the distribution functions at the boundary becomes very small at later iterations. This again makes the solution of the Laplacian with zero right-hand side very efficient. The solution of the last iteration is a good initial guess for the subsequent iteration. The computational effort for the correction of the boundary condition will turn out to be nearly negligible, which will be discussed in more detail in the following section.

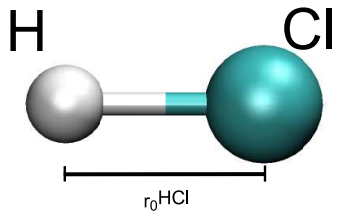

Figure 5.12. *A schematic of a HCl molecule.*

5.3.4 Discretization and Convergence

The actual discretization of the linearized equations in the fixed point iterations (5.2) and (5.3) is exactly as in the monoatomic case, see Section 5.1.1. In order to test the convergence of the discretization, we will now consider a model fluid and compute the SS-BGY3dM site-site pair distribution functions of this fluid for different grid sizes.

The model is lent from [49], where it is called a HCl-like model (hydrogen chloride) because of the properties of the particle species. We will follow this notation and use the subscripts H and Cl. Nevertheless, it should be underlined that it is not meant to be a realistic model for HCl but serves as a simple test fluid. The parameters for the diatomic HCl-like model are given in Table 5.8. Figure 5.12 shows a schematic HCl molecule.

Beside the Lennard-Jones potential, the particles now also interact via the long-range Coulomb potential. The total intermolecular potential between particles of species α and γ can be written as

$$\begin{aligned} v^I_{\alpha\gamma}(r) &= v^{LJ}_{\alpha\gamma}(r) + v^C_{\alpha\gamma}(r) \\ &= 4\epsilon_{\alpha\gamma}\left(\left(\frac{\sigma_{\alpha\gamma}}{r}\right)^{12} - \left(\frac{\sigma_{\alpha\gamma}}{r}\right)^{6}\right) + \epsilon_C \frac{q_\alpha q_\gamma}{r} \end{aligned} \tag{5.111}$$

with $r = |\mathbf{x}^\alpha - \mathbf{x}^\gamma|$ and $\alpha, \gamma = \mathrm{H}, \mathrm{Cl}$. With our choice of unit system we have

$$\epsilon_C \approx 331.84 \, \frac{\text{kcal Å}}{\text{mol e}^2}.$$

The parameters for the Lennard-Jones potential are computed according to the Lorentz-Berthelot mixing rules:

$$\epsilon_{\alpha\gamma} = \sqrt{\epsilon_\alpha \epsilon_\gamma}, \qquad \sigma_{\alpha\gamma} = 0.5(\sigma_\alpha + \sigma_\gamma), \qquad \alpha, \gamma = \mathrm{H}, \mathrm{Cl}. \tag{5.112}$$

We solve the SS-BGY3dM equations for the HCl-like model with number density $\rho = 0.018/\text{Å}^3$ and temperature $T = 420\text{K}$ ($\beta = 1.1989$). We choose $\chi = 10^{-2}$ for

$m_{\mathrm{Cl}} = 35.453$ u	$m_{\mathrm{H}} = 1.008$ u	$r_0^{\mathrm{HCl}} = 1.257$ Å
$q_{\mathrm{Cl}} = -0.2$ e	$q_{\mathrm{H}} = 0.2$ e	
$\epsilon_{\mathrm{Cl}} = 0.5143$ kcal/mol	$\epsilon_{\mathrm{H}} = 0.0397$ kcal/mol	
$\sigma_{\mathrm{Cl}} = 3.353$ Å	$\sigma_{\mathrm{H}} = 2.735$ Å	

Table 5.8. *Parameter values for the HCl-like model fluid.*

the stopping criterion of the fixed point iteration and a damping factor of $\nu = 0.01$. The parameter G, which determines the width of the charge distribution during the evaluation of the Coulomb potential, is set to $G = 0.8/\text{Å}$. Further, we use a domain $\Omega = [-10\text{Å}, 10\text{Å}]^3$ and grids with $N_1 = 32^3$, $N_2 = 64^3$, $N_3 = 128^3$ and $N_4 = 256^3$ grid points. Then, the discrete solutions[3] for $g^{(2)}_{\mathrm{HH}}$, $g^{(2)}_{\mathrm{HCl}}$ and $g^{(2)}_{\mathrm{ClCl}}$ are computed for all grid resolutions. These solutions are linearly interpolated to the finest grid and the discrete L_2- and L_∞-errors between the interpolated solutions and the solution on the finest grid are evaluated, compare Section 5.1.2 for the exact definition of the error norms. The results are presented in Table 5.9 and Figure 5.13.

	H-H		H-Cl		Cl-Cl	
dof	$e_{L_2^h}$	$e_{L_\infty^h}$	$e_{L_2^h}$	$e_{L_\infty^h}$	$e_{L_2^h}$	$e_{L_\infty^h}$
32^3	2.310_{-3}	1.397_{-1}	4.627_{-3}	3.975_{-1}	7.825_{-3}	4.594_{-1}
64^3	8.602_{-4}	4.424_{-2}	1.640_{-3}	1.513_{-1}	2.731_{-3}	1.790_{-1}
128^3	3.011_{-4}	1.269_{-2}	4.517_{-4}	4.773_{-2}	8.006_{-4}	5.224_{-2}

Table 5.9. *L_2- and L_∞-errors for the HCl-like model and different grid sizes.*

The results exhibit a similar behavior as in the monoatomic case. The L_∞-error of the H-H distribution function is reduced by a factor of 0.38 between resolution level 1 and level 2 and a factor of 0.32 between level 2 and level 3. Hence, we see a nearly linear convergence. Yet, we cannot observe any asymptotic behavior, since we have too little data. A similar behavior can be observed for the L_2-error and the other distribution functions. This may indicate that the error is again dominated by the approximation of the sharp function g_0 of our approach $g = g_0 e^{-u}$, as it has been in the case for the monoatomic solvent. It further stands out that the errors of the site-site pair distribution functions considerably differ in their absolute values. One can conclude that the H-H and H-Cl distribution functions are smoother than the Cl-Cl distribution function for this model fluid and are therefore better approximated. This will be confirmed by the results of Section 5.4, where we will present the actual

[3]We leave out the subscript h, which indicates a discrete function.

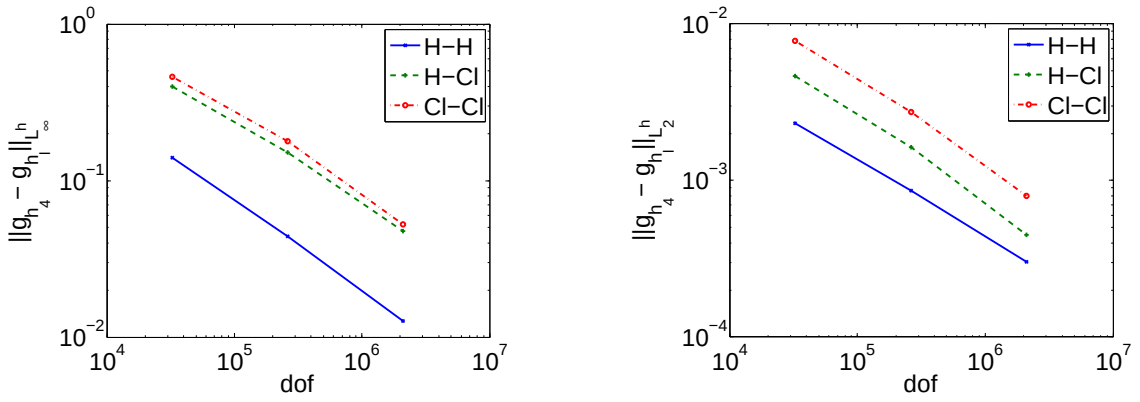

Figure 5.13. *L_∞-error (left) and L_2-error (right) for the HCl-like model and different grid sizes.*

distribution functions computed for this model. We will again use 256^3 grid points for the comparison of the SS-BGY3dM results with molecular dynamics simulations in the following. The comparison between BGY3dM and molecular dynamics has to be performed with 128^3 grid points due to the bad convergence of the molecular dynamics results in case of full three-dimensional density distributions.

Figure 5.14 shows the convergence history for the HCl-like model with $N = 128^3$, $\nu = 0.02$ and $\chi = 10^{-2}$. Recall that the equations for the H-H, H-Cl and Cl-Cl pair distribution function are solved simultaneously. Hence, the method required 822 steps in total to converge. On the one hand, this is caused by the small damping factor, which is required because of the long-range Coulomb force. On the other hand, Figure 5.14 clearly shows that the error starts to decrease monotonically for all functions only after iteration 400, even in the case of the L_2-norm. This is because the equations for the three distinct pair distribution functions are coupled and influence each other in a highly non-linear way through their respective solutions. Compared to the monoatomic case of the BGY3d equation, this causes the SS-BGY3dM and BGY3dM equations to be much more challenging to solve. Once all functions approach the neighborhood of their final solution, a monotonic decrease of all norms can be observed. The oscillations in the discrete L_∞-norm are the result of the bad approximation of the real L_∞-norm, as already explained for the monoatomic case of the BGY3d equation, see Section 5.1.2. Asymptotically, the convergence is linear and has a rate of about 0.99 for all distribution functions.

Computational Costs

In this section, we are going to discuss briefly the computational costs necessary to solve the SS-BGY3dM or the BGY3dM equations. We have already seen, that the

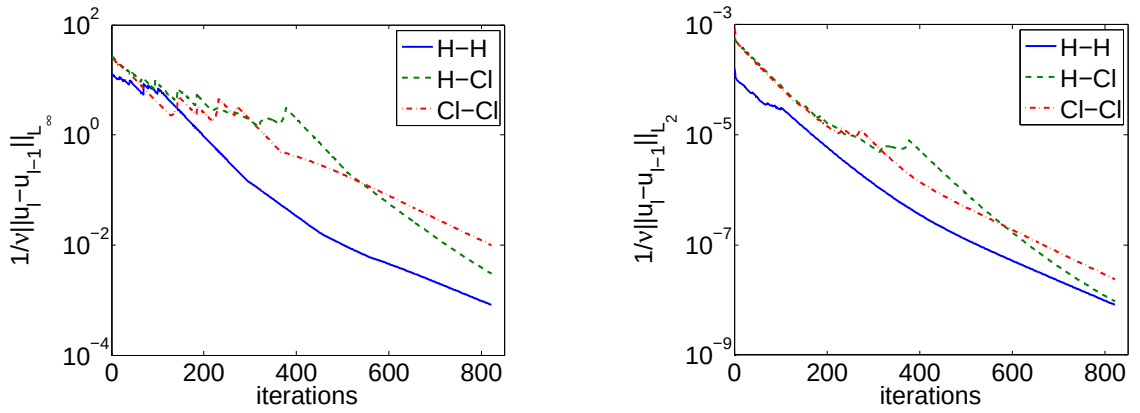

Figure 5.14. *Convergence of the fixed point iteration for the HCl-like model in the L_∞-norm (left) and the L_2-norm (right).*

number of fixed point iterations is considerably larger than in the case of monoatomic solvents and is on the order of several hundred iterations. Each iteration consists of the computation of the right hand side of the Poisson equation, the solution of the Laplacian in Fourier space and the solution of a second Laplace problem to correct the boundary conditions. For a two-site solvent like HCl, a total of 96 and 28 Fourier transforms are computed at each iteration for the SS-BGY3dM and BGY3dM equations, respectively. The amount of Fourier transforms is larger for the SS-BGY3dM equations, since there are three pair distribution functions in comparison to two single site distribution functions of the BGY3dM equations. Additionally, there is an intra-molecular term present only in the SS-BGY3dM equations. This term in particular is computationally very expensive due to the NSSA approximation involved, see also Section 5.3.2.

The computational effort for the numerical solution of the second Laplace problem (5.109), which is required to correct the boundary conditions, is comparatively small. For simplicity, we choose to solve it by the iterative GMRES method as it is implemented in PETSc [5–7]. We use the solution vector of the preceding fixed point iteration as initial guess for the GMRES iteration. Figure 5.15 shows the number of GMRES steps performed until iteration 500. As above, we solved the SS-BGY3dM equations for the HCl-like model with $N = 128^3$ grid points. Figure 5.15 exhibits an interesting characteristic. The number of GMRES steps rapidly decreases until iteration 300. From there, between zero and two GMRES steps have to be performed at each iteration in order to take the Euclidean norm of the residual below 10^{-4}, which has been chosen as stopping criterion. That is because the Laplace problems do not differ significantly at subsequent iterations. This reduces the computational costs of this boundary correction tremendously. In our special example, the per-

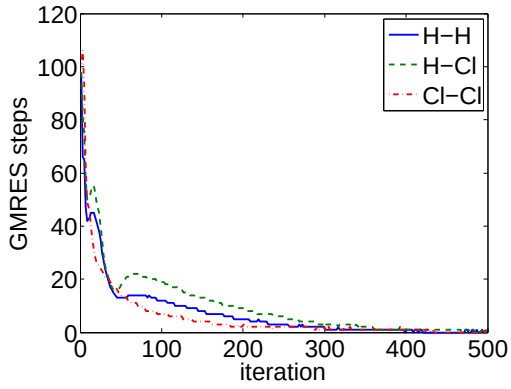

Figure 5.15. *Number of GMRES steps at each iteration of the fixed point iteration.*

centage of the computing time for the boundary correction was about six percent of the total computation time. 83% of this portion were spent during the first 300 fixed point iterations. Moreover, the choice of the GMRES solver is not optimal. Hence, the computing time necessary for the boundary correction, especially during the first iterations, could be reduced even more by employing e.g. a multigrid solver. However, this would have only a small effect on the entire computational costs and will therefore not be considered further.

It follows that the total computational costs are dominated by the number of discrete Fourier transforms to be computed at each iteration. The numerical solution of the example considered above required about 114 minutes on 32 processors of the linux cluster Himalaya [48], that is about 14 minutes per 100 iterations. The same problem computed with $N = 256^3$ grid points with the same amount of processors required approximately 140 minutes per 100 iterations, which is ten times slower. This is an acceptable factor if we consider the complexity of $\mathcal{O}(n^3 \log(n^3))$ for the three-dimensional discrete Fourier transform with n the number of grid points in one direction, and the increased costs for the parallel communication. The computing time per 100 iterations for the solution of the BGY3dM equations are about 3.2 times faster than for the SS-BGY3dM equations with the same grid resolution, as can be expected due to the reduced number of discrete Fourier transforms.

5.4 Test of the BGY3dM Model

We now investigate the model errors of the SS-BGY3dM and BGY3dM equations. To this end, we compare the results computed by our SS-BGY3dM and BGY3dM models with results from molecular dynamics simulations. In order to exemplify the dependence of the quality of the approximation on the considered solvent, we

will consider a two-site model of a fluid and compute results for two different parameter sets of this model. Moreover, we will discuss the symmetry violation due to the approximations involved in the methods. Finally, results for the site density distributions around a simple solute will be presented and compared to molecular dynamics results.

5.4.1 Comparison of SS-BGY3dM with Molecular Dynamics

We consider the two HCl-like models of [49]. The first has already been described in Section 5.3.4 with the parameter set from Table 5.8 and the structure shown in Figure 5.12. The second is a slight modification of the first. One Lennard-Jones parameter of the hydrogen atom is decreased to $\sigma_{\mathrm{H}} = 0.4\text{Å}$ which is about seven times smaller than before. We call the different models simply Model 1 and Model 2, respectively. We again choose $\chi = 10^{-2}$ for the stopping criterion of the SS-BGY3dM method and $\nu = 0.01$ for the damping factor of the fixed point iteration. The width of the charge distribution for the evaluation of the Coulomb potential is set to $G = 0.8/\text{Å}$. We further set $\Omega = [-10\text{Å}, 10\text{Å}]^3$, $\rho = 0.018/\text{Å}^3$ and $\beta = 1.1989$ for the SS-BGY3dM equation as well as for the molecular dynamics simulation. The intermolecular potential is given by (5.111). In the case of the molecular dynamics simulation, we do not assume the HCl molecules to be rigid bodies, but model the intramolecular bond by a harmonic potential:

$$v^b(r) = k_b(r - r_0^{HCl})^2, \qquad r = |\mathbf{x}^{\mathrm{H}} - \mathbf{x}^{\mathrm{Cl}}|. \tag{5.113}$$

The choice of $k_b = 500\ \frac{\text{kcal}}{\text{mol Å}^2}$ allows fluctuations of the bond length of less than 0.003Å. Hence, the difference of the resulting distribution functions between the rigid molecule and the molecule equipped with this bond-potential will not be noticeable. We again employ the molecular dynamics package TREMOLO [123]. To this end, the Coulomb interaction is computed by the smooth-particle-mesh-Ewald method (SPME), see e.g. [46]. For all further details about the molecular dynamics simulation and about the computation of the distribution functions from a molecular dynamics trajectory refer to Section 5.2.1.

In order to compare the results of SS-BGY3dM and molecular dynamics, we again compute the error values $e_{L_2^h}$, $e_{L_\infty^h}$ and e_{max}, which we already used for the monoatomic BGY3d equation in Section 5.2.2. Their respective values for the H-H, H-Cl and Cl-Cl pair distribution functions are given in Table 5.10 for Model 1 and Model 2. Figures 5.16, 5.17, 5.18 and 5.19, 5.20, 5.21 show the radial component of the solutions of both methods and the deviation between them.

The characteristic properties of a method that employs the Kirkwood approximation can be identified in each of the figures. The exact position and the height of the first peak do not match those of the molecular dynamics results except for the H-H distribution of Model 1. The frequency of the subsequent oscillation is too

	MD	SS-BGY3dM			
	max g	max g	$e_{L_2^h}$	$e_{L_\infty^h}$	e_{max}
H-H (Model 1)	1.32	1.32	3.549_{-6}	6.223_{-2}	0.00
H-Cl (Model 1)	1.29	1.37	4.385_{-6}	8.698_{-2}	0.08
Cl-Cl (Model 1)	1.99	1.93	1.491_{-5}	3.194_{-1}	0.06
H-H (Model 2)	1.14	1.13	6.893_{-6}	2.692_{-1}	0.00
H-Cl (Model 2)	1.17	1.18	7.516_{-6}	3.000_{-1}	0.01
Cl-Cl (Model 2)	2.07	2.48	2.896_{-5}	5.756_{-1}	0.41

Table 5.10. *Comparison of SS-BGY3dM with molecular dynamics results for the HCl-like models.*

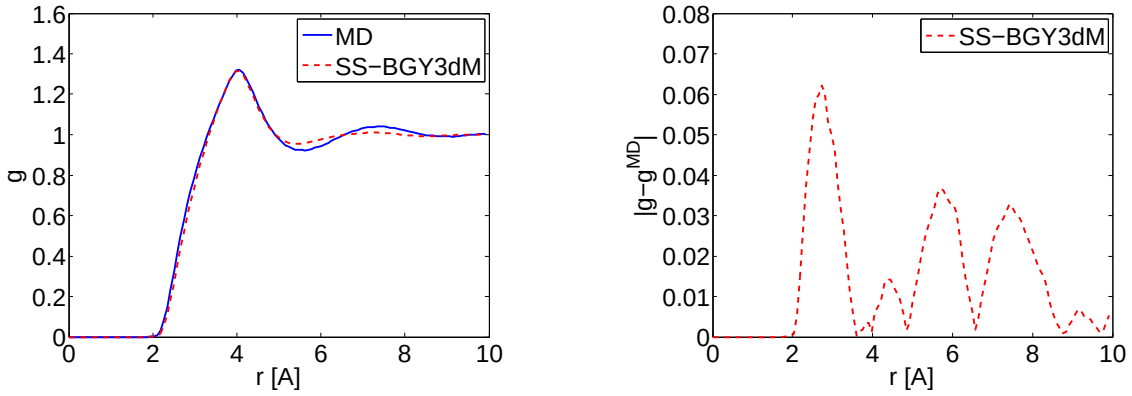

Figure 5.16. *Radial Component of the H-H distribution function (Model 1): Comparison (left) and deviation (right) between SS-BGY3dM and molecular dynamics results.*

low. These errors are known to be a consequence of the two-particle superposition approximation and can only be handled by considering a three-particle interaction, as already denoted in Section 5.2.2.

A comparison of the errors for the different site-site distribution functions of Model 1 reveals that their magnitude differs significantly. The L_2- and L_∞-errors of the Cl-Cl distribution function are about 3.5 times larger than the H-Cl errors. Hence, the quality of the approximation depends on the different potential parameters of the respective particle species. In this special example, the Cl-atoms have a much stronger Lennard-Jones interaction, which obviously influences the quality of the solution in a negative sense. The comparison to the results of Model 2 uncovers another difficulty. All errors are increased for this model, which is due to the

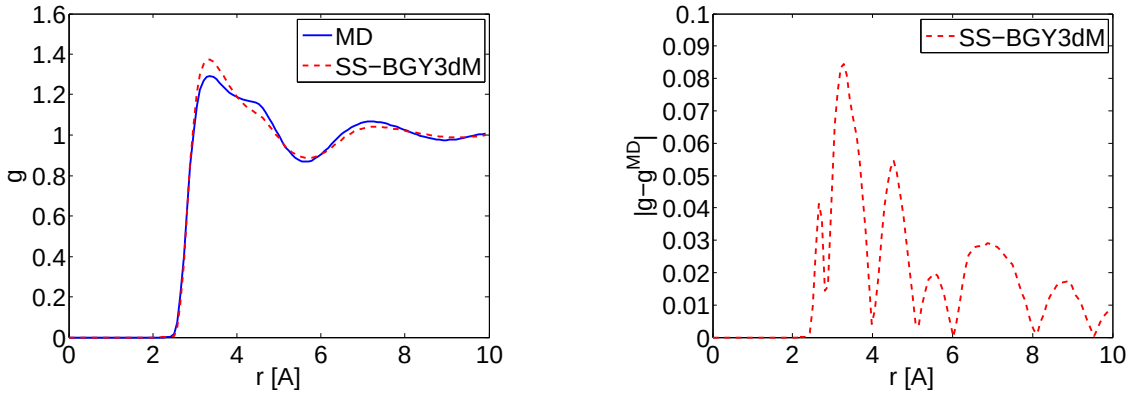

Figure 5.17. *Radial Component of the H-Cl distribution function (Model 1): Comparison (left) and deviation (right) between SS-BGY3dM and molecular dynamics results.*

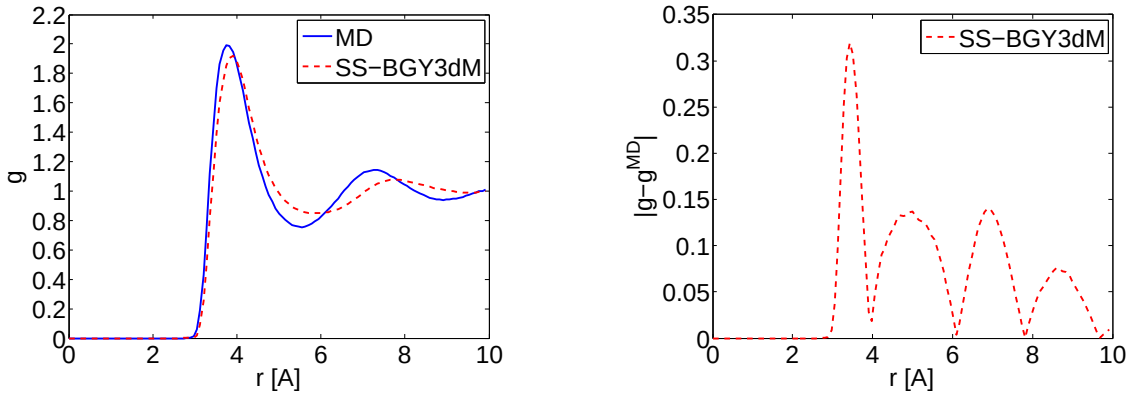

Figure 5.18. *Radial Component of the Cl-Cl distribution function (Model 1): Comparison (left) and deviation (right) between SS-BGY3dM and molecular dynamics results.*

decreased value of σ_{H} considered for Model 2. This leads to more different particle species and, hence, worsens the approximation of the SS-BGY3dM equations. Especially the H-H and H-Cl distribution functions of Model 2 show major deficiencies in the prediction of the position of the first flank of the function. However, similar problems can be observed for the solution of the Ornstein-Zernike based extended RISM equations of Hirata and Rossky in reference [49].

We can conclude that the exact characteristics of the site-site distribution functions are very difficult to approximate, as long as the approximations only comprise pair distribution functions. Obviously, the approximation of the SS-BGY3dM equa-

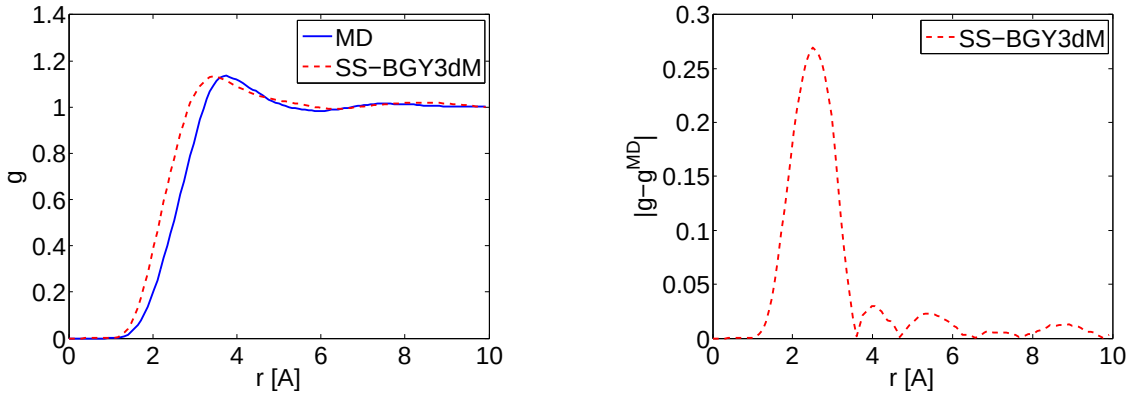

Figure 5.19. *Radial Component of the H-H distribution function (Model 2): Comparison (left) and deviation (right) between SS-BGY3dM and molecular dynamics results.*

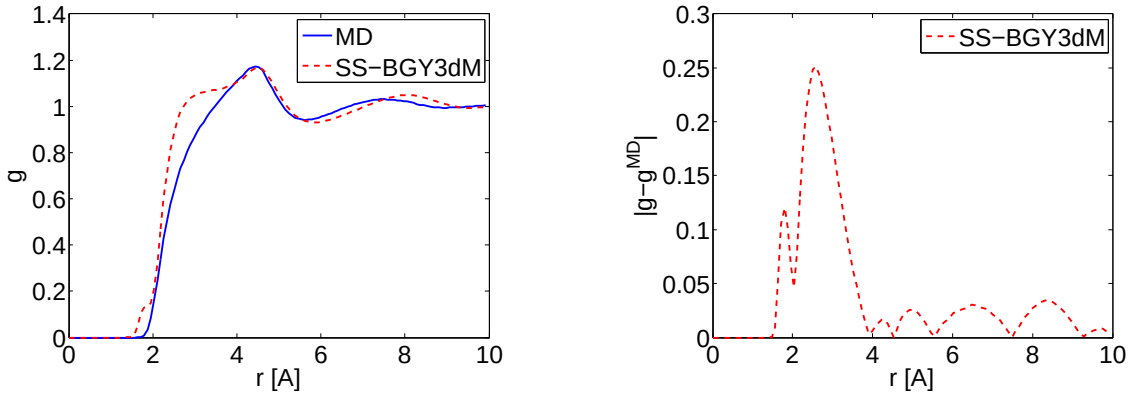

Figure 5.20. *Radial Component of the H-Cl distribution function (Model 2): Comparison (left) and deviation (right) between SS-BGY3dM and molecular dynamics results.*

tions perform better for more similar particle species. All important features of the distribution functions of the HCl-like Model 1 were reproduced. It follows that the general form and especially the modeling of the intramolecular bonds of the SS-BGY3dM equations is validated by the results.

Symmetry Issues

The site-site pair distribution functions $g^{(2)}_{\alpha\gamma}$ are symmetric under exchange of particles, i.e. $g^{(2)}_{\alpha\gamma} = g^{(2)}_{\gamma\alpha}$, since they only depend on the distance of the two sites. It

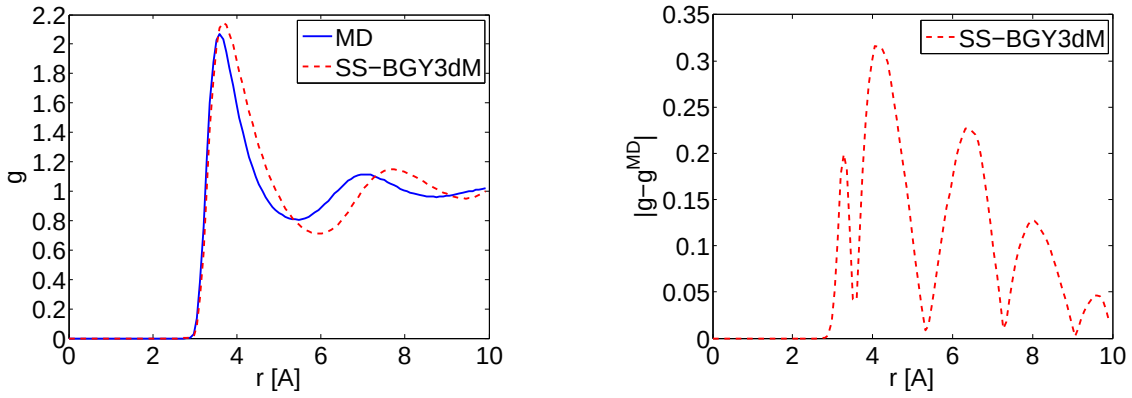

Figure 5.21. *Radial Component of the Cl-Cl distribution function (Model 2): Comparison (left) and deviation (right) between SS-BGY3dM and molecular dynamics results.*

follows that the solutions of the exact equations of the YBG-hierarchy for the site-site pair distribution functions (4.78) are independent of the order of the particle types. If we now consider our two-side model fluid with particle types A and B and change the particle order in the derivation of the SS-BGY3dM equations for the mixed pair distribution function, we get the following equation for $g^{(2)}_{\mathrm{BA}}$ instead of equation (5.66) for $g^{(2)}_{\mathrm{AB}}$:

$$\begin{aligned}
\Delta_{\mathbf{x}_1} u^{(2)}_{\mathrm{BA}}(\mathbf{x}_1,\mathbf{x}_2) = &- \beta\rho_{\mathrm{A}}\nabla_{\mathbf{x}_1}\cdot\int_\Omega \mathbf{F}_{\mathrm{BA}}(\mathbf{x}_1,\mathbf{x}_3) g^{(2)}_{\mathrm{BA}}(\mathbf{x}_1,\mathbf{x}_3) g^{(2)}_{\mathrm{AA}}(\mathbf{x}_2,\mathbf{x}_3)\; d\mathbf{x}_3 \\
&- \beta\rho_{\mathrm{B}}\nabla_{\mathbf{x}_1}\cdot\int_\Omega \mathbf{F}_{\mathrm{BB}}(\mathbf{x}_1,\mathbf{x}_3) g^{(2)}_{\mathrm{BB}}(\mathbf{x}_1,\mathbf{x}_3) g^{(2)}_{\mathrm{BA}}(\mathbf{x}_3,\mathbf{x}_2)\; d\mathbf{x}_3 \\
&- \beta\nabla_{\mathbf{x}_1}\cdot\frac{\int_\Omega \mathbf{F}_{\mathrm{BB}}(\mathbf{x}_1,\mathbf{x}_3)\tilde{g}^{(2)}_{\mathrm{BB;A}}(\mathbf{x}_1,\mathbf{x}_3)\omega_{\mathrm{AB}}(\mathbf{x}_2,\mathbf{x}_3)\; d\mathbf{x}_3}{n^{\mathrm{B}}_{\mathrm{AB}}(\mathbf{x}_1,\mathbf{x}_2)} \\
&- \Delta_{\mathbf{x}_1}\ln\left(\int_\Omega \omega_{\mathrm{AB}}(\mathbf{x}_3,\mathbf{x}_1)\tilde{g}^{(2)}_{\mathrm{AA;B}}(\mathbf{x}_2,\mathbf{x}_3)\; d\mathbf{x}_3\right). \qquad (5.114)
\end{aligned}$$

Please note, that this is not a simple relabeling of indices but a different derivation of the mixed site-site term. This is a consequence of choosing a different term of the transformed Liouville equation (4.67) by exchanging the types of the first two particles.

As already mentioned above, this exchange would have no effect for the exact equations. However, since we introduced approximations in order to be able to solve the equations, the question arises whether the solution of the approximated equations are also symmetric under particle exchange. Numerical tests indicate that this is not the case. As an example, we present the results of the SS-BGY3dM

	MD	SS-BGY3dM			
	$\max g$	$\max g$	$e_{L_2^h}$	$e_{L_\infty^h}$	e_{max}
H-H (AB)	1.32	1.32	3.549_{-6}	6.223_{-2}	0.00
H-Cl (AB)	1.29	1.37	4.385_{-6}	8.698_{-2}	0.08
Cl-Cl (AB)	1.99	1.93	1.491_{-5}	3.194_{-1}	0.06
H-H (BA)	1.32	1.22	5.785_{-6}	1.281_{-1}	0.10
H-Cl (BA)	1.29	1.18	8.335_{-6}	2.645_{-1}	0.12
Cl-Cl (BA)	1.99	1.73	1.741_{-5}	4.562_{-1}	0.26

Table 5.11. *Comparison of SS-BGY3dM (AB) and SS-BGY3dM (BA) with molecular dynamics results for the HCl-like model (Model 1).*

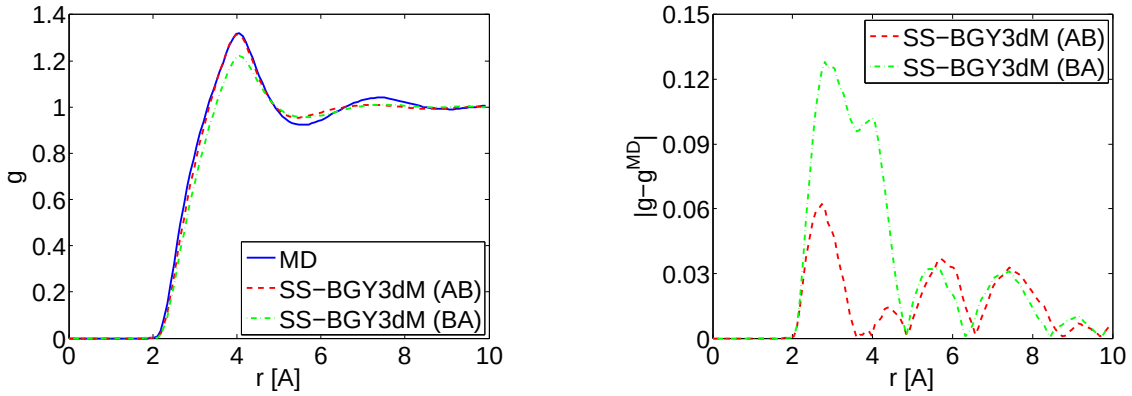

Figure 5.22. *Radial Component of the H-H distribution function (Model 1): Comparison (left) and deviation (right) between SS-BGY3dM (AB), SS-BGY3dM (BA) and molecular dynamics results.*

equations for our HCl-like model (Model 1), where we used equation (5.114) instead of (5.66). The results computed with (5.66) (A=H, B=Cl) are labeled by AB, whereas the new results computed with (5.114) (A=Cl, B=H) are labeled by BA. Table 5.11 shows the results for both ways compared to the molecular dynamics results. Radial plots of the AB and BA results are depicted in Figures 5.22, 5.23 and 5.24.

It is obvious that the results differ considerably for the two approaches to compute the mixed site-site pair distribution function. Since the pair distribution functions with two equal sites (H-H, Cl-Cl) also depend on the mixed function, they differ as well. The bottom line is that the error of the approximation depends on the particle types involved. This is not surprising, since the values of the potential

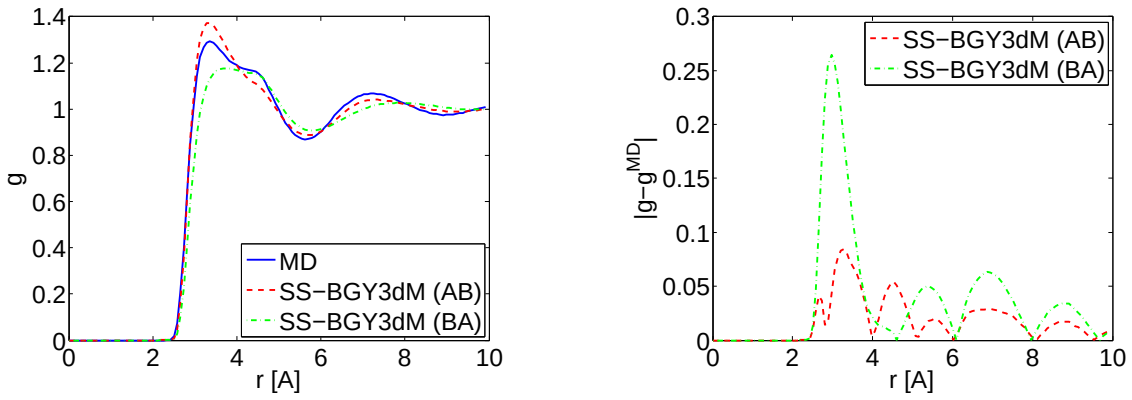

Figure 5.23. *Radial Component of the H-Cl distribution function (Model 1): Comparison (left) and deviation (right) between SS-BGY3dM (AB), SS-BGY3dM (BA) and molecular dynamics results.*

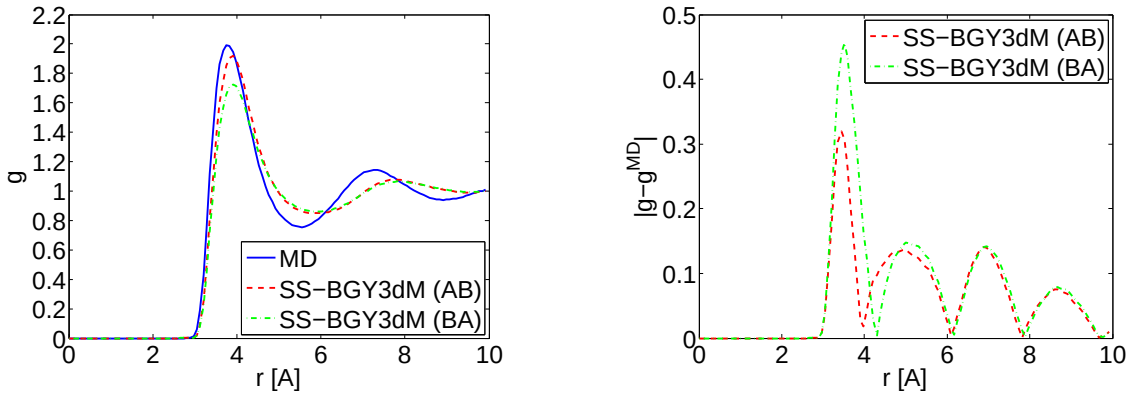

Figure 5.24. *Radial Component of the Cl-Cl distribution function (Model 1): Comparison (left) and deviation (right) between SS-BGY3dM (AB), SS-BGY3dM (BA) and molecular dynamics results.*

parameters of the different particle types affect the relative errors of the different terms. The more similar the particle types become, the smaller is the difference between the AB- and BA-results, which even vanishes for identical particle types. The second observation is that the deviation of the BA-functions compared to the molecular dynamics results is consistently larger than for the AB-functions. Hence, it would have been advantageous to choose the AB-equation in this case. This raises the question how we can choose the order of particle types in advance in order to obtain a better accuracy. This is still an open question. It would require a detailed understanding of how the approximation errors involved in the many different terms

	MD	SS-BGY3dM			
	$\max g$	$\max g$	$e_{L_2^h}$	$e_{L_\infty^h}$	e_{max}
H	1.96	1.56	3.594_{-5}	4.134_{-1}	0.39
Cl	3.39	2.38	9.671_{-5}	1.215_{+0}	1.00

Table 5.12. *Comparison of BGY3dM with molecular dynamics results for the HCl-like solvent around a HCl molecule as solute.*

interact and lead to the error of the solution. For now, we have to compare the results of both ways with the results of a molecular dynamics simulation in order to decide.

5.4.2 Comparison of BGY3dM with Molecular Dynamics

Finally, we test the BGY3dM model with respect to the computation of solvent densities around an arbitrary solute. For this, we again employ the HCl-like model (Model 1) already described in the preceding Section. A single HCl molecule is considered as the solute. It is placed symmetrically along the x_1-axis at the center of the simulation box. The site-site pair distribution functions of the pure solvent, which are required as input of the BGY3dM model, are computed by the SS-BGY3dM model. All other simulation parameters are taken from Section 5.4.1. Details on how the site density distribution is computed by molecular dynamics can be found in Section 5.2.1. In this case, a total of $3.2\ 10^8$ molecular dynamics time steps were necessary in order to reach a satisfactory level of convergence.

The computed site densities and their deviation are depicted in Figure 5.25. The computed error quantities can be found in Table 5.12. The molecular dynamics results still show distinct fluctuations, but all features of the distribution functions have clearly developed. A comparison of the results between the BGY3dM model and molecular dynamics shows a satisfying agreement. The low L_2-errors indicate a good overall approximation. The main peaks and the subsequent oscillation pattern are reproduced with an accuracy that can be expected considering the approximations involved. The main deviation can be observed at the location of the main peaks of the distributions. All other errors are not resolved in the plots due to the fluctuations of the molecular dynamics results. Hence, the L_∞-error is about 0.4 and 1.2 for the hydrogen and the chloride distribution, respectively, and can also be located at the main peaks. The error of the chloride distribution function is considerably larger, as it was also the case for the Cl-Cl pair distribution functions, see Section 5.4.1. Recall that the site-site pair distribution functions required as input of the BGY3dM model have already been computed with the approximate SS-BGY3dM model. Hence, the approximation error enters twice: directly via the

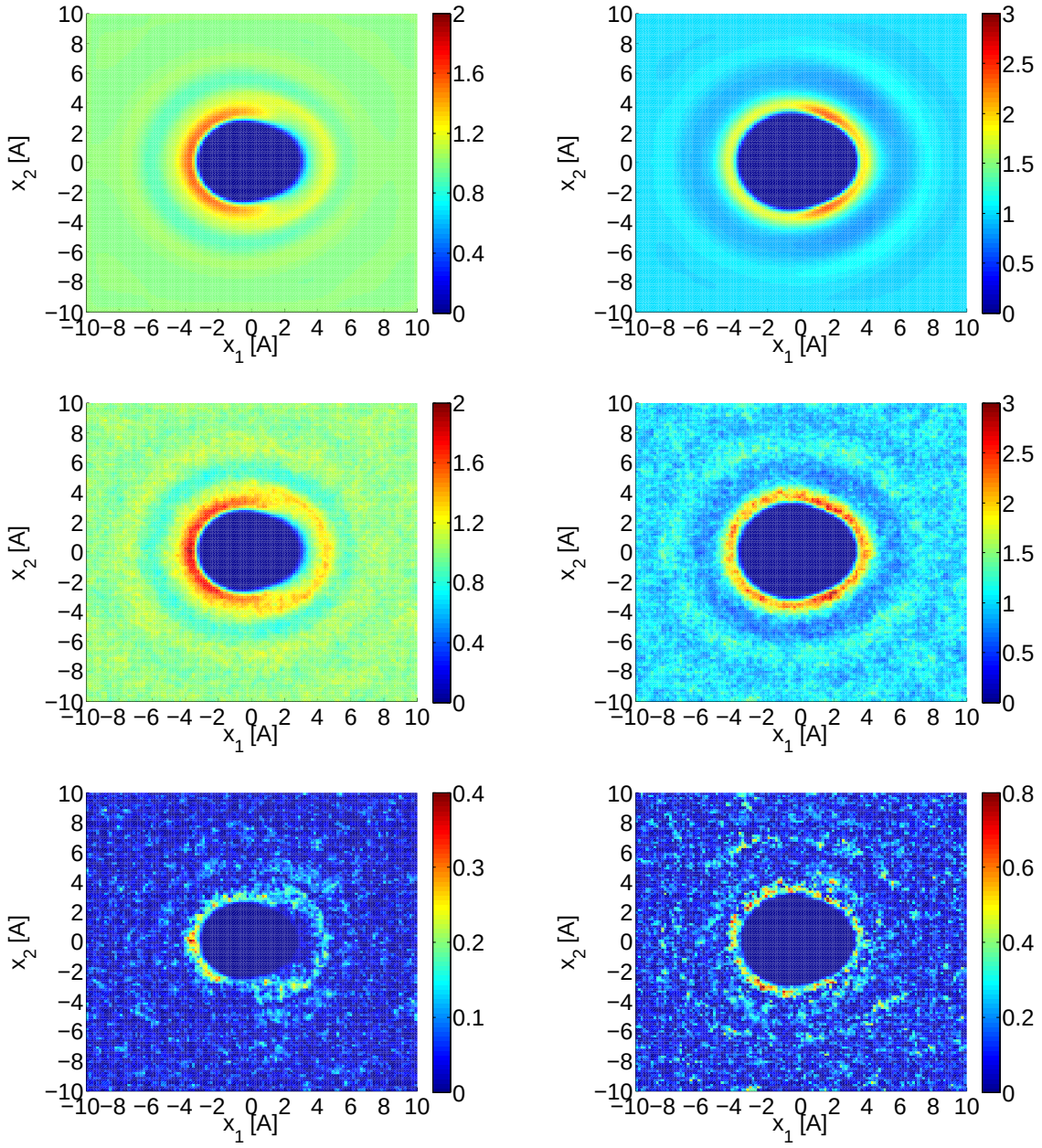

Figure 5.25. *Distribution functions for the HCl-like model around a single HCl solute computed with the BGY3dM model (top) and with molecular dynamics (middle) and deviation between them (bottom) at the $x_3 = 0$ plane. Hydrogen distribution (left) and chloride distribution (right). The solute is not shown.*

approximation involved in the BGY3dM model and by the use of the approximated site-site pair distribution functions computed with the SS-BGY3dM model.

In conclusion, the results are very promising. The BGY3dM model is able to reproduce the important features of the site distribution functions around a solute with satisfying accuracy for the considered model solvent. As indicated by the results for the site-site pair distribution functions, the performance can be worse for different solvents. But still, the approximation of the site densities by the BGY3dM model is very efficient. In order to obtain the results, we have performed 32 molecular dynamics simulations for any site distribution with different initial configurations of the solvent velocities. Each simulation comprised 10^7 time steps and required about 122h of computing time on one processor of the cluster Himalaya [48]. The solution of the BGY3dM model on 32 processors of Himalaya required only 954s for 315 iteration steps, which is two–three orders of magnitudes faster, assuming that the different molecular dynamics simulations can be performed simultaneously.

5.4.3 Summary

We have shown that the BGY3dM model can be solved efficiently and provides reasonable good results when compared to site-site pair distribution functions and site densities computed by a molecular dynamics simulation. The results validate the general form of the derived SS-BGY3dM and BGY3dM equations. In particular, the modeling of the rigid bonds in the solvent molecules and the comprised normalized site-site superposition approximation (NSSA) of Taylor and Lipson [116] in the intramolecular terms have proven to appropriately approximate the structure of the solvent molecules.

Nevertheless, the BGY3dM model does not work well for every solvent model, yet. We have learned in various numerical tests that the method sometimes fails to converge or leads to obviously unphysical results. Problems arise for example for higher densities, lower temperatures or strong interactions. We assume that the reason for this behavior again originates in the nature of the Kirkwood superposition approximation. This approximation is perfectly suited for large separations of particles or in the limit of zero density. However, the error becomes larger for the problematic situations just described. This typically leads to a considerable overestimation of the main peaks of the distribution functions. In combination with the non-linear coupling of several equations for different site density functions, this can yield convergence problems of the non-linear iteration.

Nonetheless, this is not a problem of the Kirkwood approximation alone. The various methods based on the Ornstein-Zernike equation struggle with similar issues. Convergence problems are not reported in the literature, but the accuracy of the computed results often is not satisfactory, see e.g. [49] concerning the XRISM method of Hirata and Rossky [50] for the HCl-like models. A possible way out is

given by the so-called empirical bridge functions as employed by Du, Beglov and Roux [28] or Kovalenko and Hirata [63] in order to improve the results of their methods for water as solvent. These bridge functions can account for deficiencies that are particular for the approximation and the solvent considered. They employ free parameters that have to be fitted by an empirical procedure, such that they lead to an improved accuracy. This would also be possible for our BGY3dM model. However, the empirical adjustment is a rather technical task and is therefore beyond the scope of this thesis.

Compared to the computation of the site densities by a molecular dynamics simulation, our approximate BGY3dM models based on the liquid state integral equation theories exhibit a drastically reduced computation time. The reduction of the computing time has been two–three orders of magnitude for the HCl-like solvent with a HCl molecule as solute. To this end, we even employed a combined molecular dynamics/Monte Carlo method by performing several individual molecular dynamics simulations with different initial configurations. This way, we were able to parallelize the computation using as many processors as there were employed for the solution of the BGY3dM equations. The reduction is indeed smaller as it was the case for monoatomic solvents, but it is still the difference between some minutes and several days. However, we should in fact compare the computational effort needed for the approximation of the potential of mean force (PMF) instead of the solvent densities. Regarding the forces of the PMF, their computation requires a further integral over the domain Ω if the solvent density is known. This can be computed with linear complexity. With respect to the computational effort, the direct approximation of the forces of the PMF by a molecular dynamics simulation is, however, equivalent to the approximation of the solvent density. Hence, our observations concerning the computing time for the approximation of the solvent densities are also valid for the approximation of the PMF.

However, in order to be applicable in an implicit solvent model, the computing times of the approximate models have to be reduced even further. This could be achieved, for example, by exploiting the similarity of the density distributions when the solute configuration changes only slightly, as it would be the case in subsequent time steps of a molecular dynamics simulation. A different discretization of the BGY3dM model may also lead to a more efficient numerical method. However, this is not a trivial task and is therefore postponed to future work.

Chapter 6
Applications

In the preceding chapter we have seen that the SS-BGY3dM and the BGY3dM equations lead to reasonable results for appropriate parameter sets of molecular solvents. In this chapter we are going to present examples of applications of the BGY3dM equations. We will consider a realistic fluid which is employed as solvent in chemical applications. The site distribution functions of the solvent species will be computed and presented for several solutes. From this, we will compute the charge distribution around the solute and discuss some basic properties of the different site and charge distributions.

6.1 Carbon Disulfide as Solvent

We consider carbon disulfide (CS_2) as solvent. Carbon disulfide is a colorless liquid which is mainly used to solve fats, rubber, resins and waxes, among other applications, see e.g. [19]. The CS_2 molecule is linear and has no dipole moment. It is a non-polar solvent. For our numerical computations we employ the model of Zhu et al. [129]. As before, the functional form of the interaction potential is given as a sum of Lennard-Jones and Coulomb terms. The potential between two sites α and γ with $\alpha, \gamma = \mathrm{C}, \mathrm{S}$ is given by

$$\begin{aligned} v_{\alpha\gamma}(r_{\alpha\gamma}) &= v^{LJ}_{\alpha\gamma}(r_{\alpha\gamma}) + v^{C}_{\alpha\gamma}(r_{\alpha\gamma}) \\ &= 4\epsilon_{\alpha\gamma}\left(\left(\frac{\sigma_{\alpha\gamma}}{r_{\alpha\gamma}}\right)^{12} - \left(\frac{\sigma_{\alpha\gamma}}{r_{\alpha\gamma}}\right)^{6}\right) + \epsilon_C \frac{q_\alpha q_\gamma}{r_{\alpha\gamma}} \end{aligned} \tag{6.1}$$

with $r_{\alpha\gamma} = |\mathbf{x}^\alpha - \mathbf{x}^\gamma|$ and $\epsilon_C \approx 331.84 \, \frac{\text{kcal Å}}{\text{mol e}^2}$. The respective parameter values are given in Table 6.1. The linear structure of the CS_2 molecule is shown in Figure 6.1 (left).

The carbon disulfide model is a three-site model. Compared to the BGY3dM equations of a two-site model, as given in (5.59) and (5.60), the equations for a

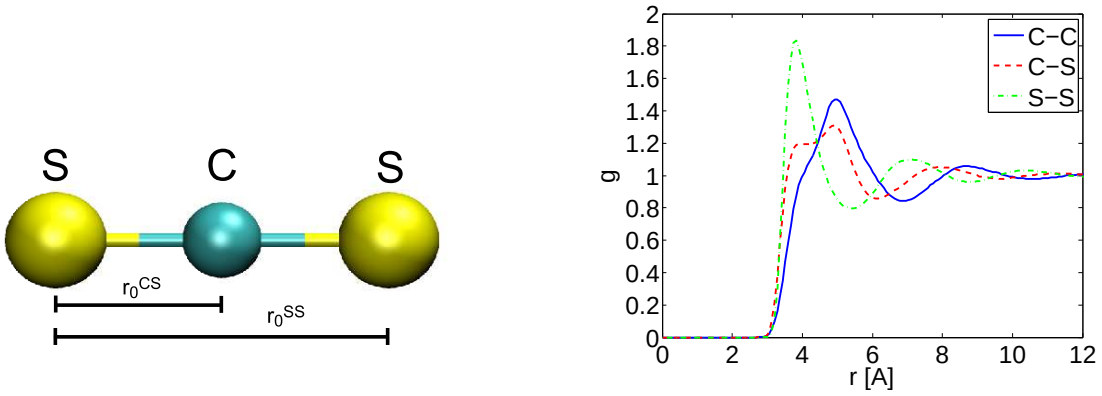

Figure 6.1. *Configuration of the carbon disulfide molecule (left) and the site-site pair distribution functions computed by molecular dynamics (right).*

$m_C = 12.011$ u	$m_S = 32.065$ u	$r_0^{CS} = 1.56$ Å
$q_C = -0.308$ e	$q_S = 0.154$ e	$r_0^{SS} = 3.12$ Å
$\epsilon_C = 0.1013$ kcal/mol	$\epsilon_S = 0.3950$ kcal/mol	
$\sigma_C = 3.200$ Å	$\sigma_S = 3.520$ Å	

Table 6.1. *Parameter values for the carbon disulfide model of Zhu et al. [129].*

three-site model comprise some additional intramolecular terms. However, we will not explicitly state the BGY3dM equations for the three-site model, since they can easily be deduced from the general formula (4.114).

As input for the BGY3dM equations, the site-site pair distribution functions of carbon disulfide are required. Since the SS-BGY3dM equations for the carbon disulfide model lead to unphysical results, we compute these functions by a molecular dynamics simulation of 80 CS_2 molecules at a number density of $\rho = 0.01/\text{Å}^3$ and a temperature of $T = 360K$ with periodic boundary conditions. We have to use a relatively high temperature in order to ensure convergence of the BGY3dM equations. However, the maximum difference between the site-site pair distribution functions at room temperature ($T = 298,15K$) and $T = 360K$ is only about five percent at the first peak of the S-S distribution function. The qualitative characteristics of the functions are exactly conserved. Since the boiling point of real carbon disulfide lies at $T_b = 319.15K$, the considered liquid system corresponds to a closed volume under pressure. The resulting pair distribution functions of the molecular dynamics simulation are depicted in Figure (6.1) (right).

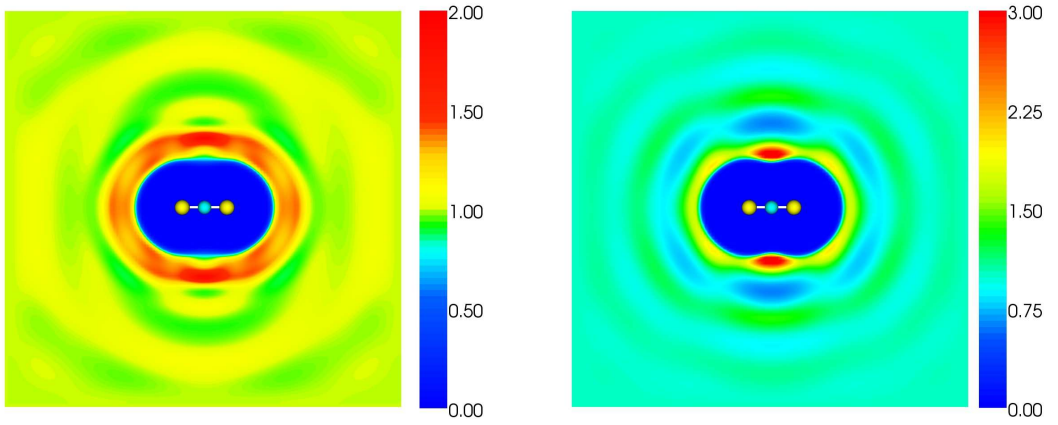

Figure 6.2. *Site distribution functions of carbon disulfide around a CS_2 molecule. Carbon distribution at the $x_3 = 0$ plane (left) and sulfur distribution at the $x_3 = 0$ plane (right).*

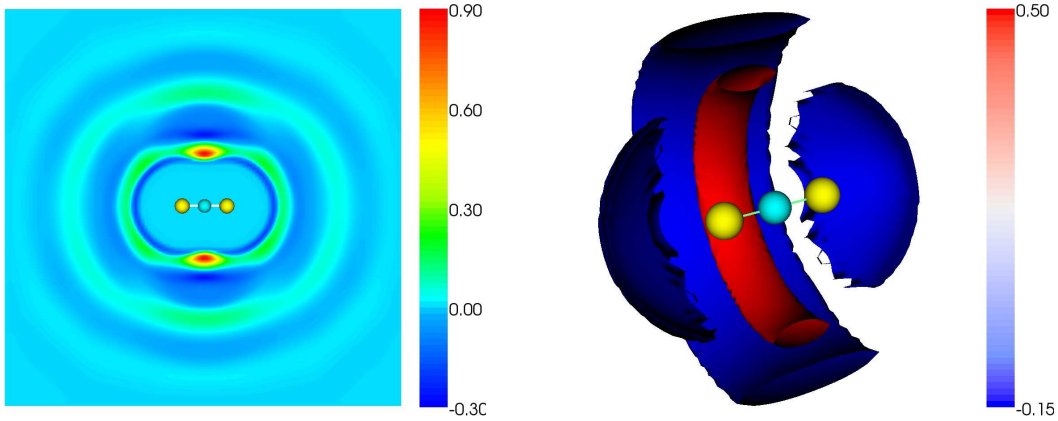

Figure 6.3. *Charge distribution of carbon disulfide around a CS_2 molecule. Cut at the $x_3 = 0$ plane (left) and isosurface plot (right).*

Carbon Disulfide as Solute

As a first example we compute the site distribution functions of carbon disulfide around a single CS_2 molecule as solute. The computational domain is set to $\Omega = [-14\text{Å}, 14\text{Å}]^3$. Figure 6.2 shows the site distributions at the $x_3 = 0$ plane. The carbon distribution function shows a broad maximum around the entire solute molecule. This is a superposition of the van der Waals attraction modeled by the Lennard-Jones potential between the carbon atoms and the solute, and the Coulomb attraction between the solvent carbon and the solute sulfur atoms. The solvent sul-

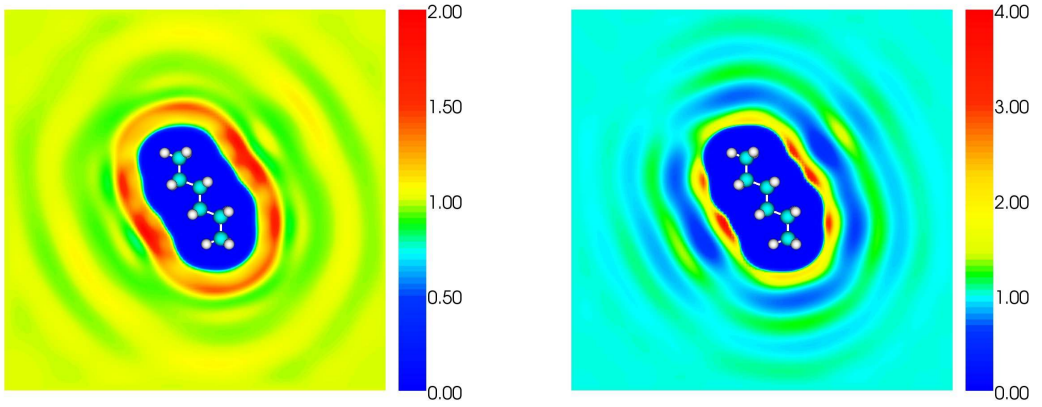

Figure 6.4. *Site distribution functions of carbon disulfide around a hexane molecule. Carbon distribution at the $x_3 = 0$ plane (left) and sulfur distribution at the $x_3 = 0$ plane (right).*

fur distribution shows a sharp peak around the solute carbon particle due to the strong Coulomb interaction between them. This can also be observed in Figure 6.3, where the charge distribution is plotted. The charge distribution for carbon disulfide can be computed from the site distribution functions by

$$g_{charge} = q_{\mathrm{C}} g_{\mathrm{C}} + 2 q_{\mathrm{S}} g_{\mathrm{S}} \tag{6.2}$$

with q_α the charge of site α, $\alpha = \mathrm{C}, \mathrm{S}$. As expected, the plots validate that charges with opposite sign are more likely to be found next to each other than charges with equal sign. Hence, a closed band of high density of positive sulfur atoms evolves around the solute carbon whereas the solvent carbon is more likely to be found next to the solute sulfur atoms.

Hexane as Solute

As second solute we consider a hexane molecule ($\mathrm{C_6H_{14}}$). Hexane is a colorless, highly flammable liquid and a non-polar solvent. It is a common constituent of gasoline and glues. As solvent, it is used to extract oils from crops such as soybeans, flax, peanuts, and safflower seed. It is also used as a cleansing agent in the textile, furniture, shoemaking and printing industries [80]. Hexane has five isomers. We consider the straight chain isomer $\mathrm{CH_3(CH_2)_4CH_3}$. The potential parameters for hexane are taken from the general-purpose force field OPLS [57]. The simulation domain is set to $\Omega = [-16\text{Å}, 16\text{Å}]^3$.

Figure 6.4 shows the carbon and sulfur distributions around hexane at the $x_3 = 0$ plane. In these plots the hexane molecule is raised 2Å above the plane for visualization purposes. As already observed for the carbon disulfide molecule as solute,

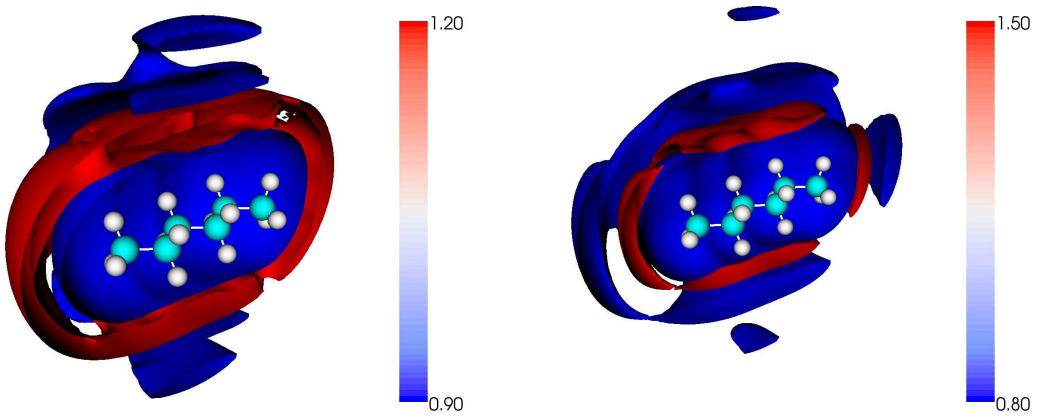

Figure 6.5. *Site distribution functions of carbon disulfide around a hexane molecule. Isosurface plot of carbon distribution (left) and isosurface plot of sulfur distribution (right).*

the carbon distribution has a broad maximum around the solute, whereas the sulfur distribution shows more distinct maxima between the positions of the positively charged hydrogen atoms. At both ends of the hexane chain, the sulfur maxima are considerably smaller. As can also be seen in Figure 6.5 the first maximum of the carbon as well as the sulfur distribution build a nearly closed shell around the entire solute molecule. This is because the hexane molecule is non-polar and the van der Waals interaction modeled by the Lennard-Jones potential is the dominant force in this case.

Methanol as Solute

In contrast to hexane, we now consider methanol as polar solute. Methanol is the simplest alcohol and has the chemical formula CH_3OH. It is a colorless, highly flammable liquid used as a petrol additive, solvent or as antifreeze [78]. Due to the alcohol specific OH-group, methanol is a polar molecule. The oxygen and hydrogen atoms carry strong opposite charges. This makes methanol soluble in water and other polar solvents but insoluble in non-polar solvents as carbon disulfide or hexane. We again employ the OPLS force field [57] for the parameter set of methanol. The computational domain is $\Omega = [-14\text{Å}, 14\text{Å}]^3$ in this case.

In the OPLS force field the hydrogen of the OH-group is modeled as pure charge carrying site without Lennard-Jones interaction. However, for numerical stability of the BGY3dM equations, a hard core is required at the position of any atom. Therefore, we also introduce Lennard-Jones parameters for the oxygen bonded hydrogen and choose $\sigma_H = 3.4\text{Å}$ and $\epsilon_H = 0.03\text{kcal/mol}$. The value of σ_H is artificially high for

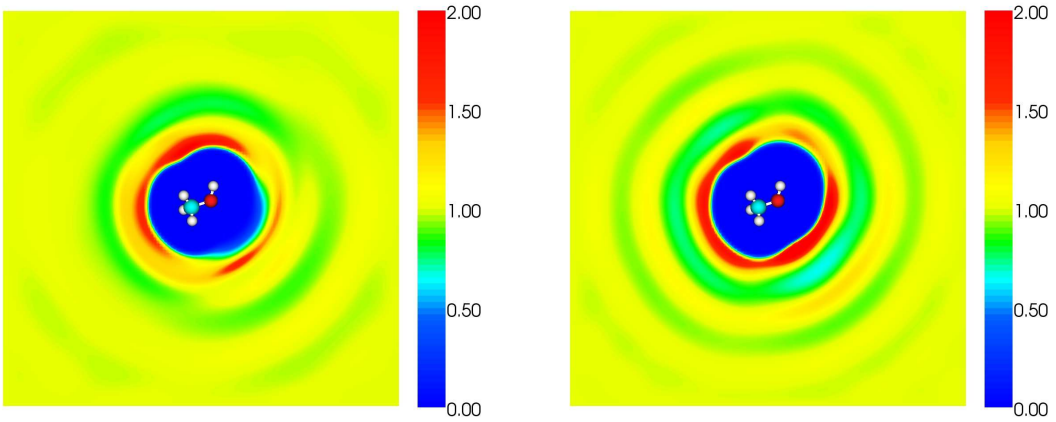

Figure 6.6. *Site distribution functions of carbon disulfide around a methanol molecule. Carbon distribution at the $x_3 = 0$ plane (left) and sulfur distribution at the $x_3 = 0$ plane (right).*

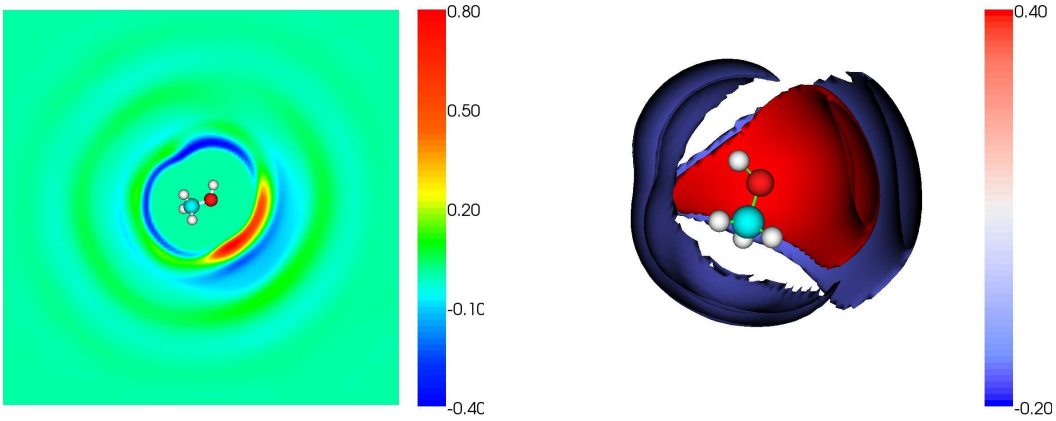

Figure 6.7. *Charge distribution of carbon disulfide around a methanol molecule. Cut at the $x_3 = 0$ plane (left) and isosurface plot (right).*

a hydrogen atom of the OH-group, but it has been required for stable convergence of the BGY3dM equations. Yet, we should motivate this choice. We observed that the carbon density of the CS_2 solvent is overestimated in the neighborhood of strong positively charged particles as the hydrogen atom. This is partly due to a lack of intramolecular coupling in the BGY3dM model. In the intramolecular terms of the BGY3dM model the coupling of the carbon and the sulfur sites is incorporated without considering the relative position of the two sulfur atoms. This is a three-body effect that is neglected by the n-level Kirkwood approximation. Taking the

three-body effect into account would lower the carbon density next to the hydrogen atom, because sulfur has a low density in the vicinity of positive charges. Hence, we choose the high value of σ_{H} in order to compensate for the missing three-body correction.

Figure 6.6 shows the carbon and sulfur distributions around methanol at the $x_3 = 0$ plane. Again, the methanol molecule is raised 2Å above the plane for visualization purposes. It is obvious that the strong Coulomb interaction highly influences the behavior of the distribution functions. The negatively charged solvent carbons are more likely to be found in the vicinity of the positive solute hydrogens, whereas the solvent sulfur atoms are dominantly attracted by the negative oxygen site. This is even more definite in the plots of the charge distributions of Figure 6.7. We also notice a negatively charged cloud behind the strong positive sulfur peak next to the solute oxygen atom. This charge minimum forms partly due to the intramolecular bond between carbon and sulfur, but also due to the intermolecular attraction of different solvent atoms. The whole picture reveals the well-known fact that charges tend to neutralize each other. Hence, the net forces on a particle in a fluid at equilibrium are exerted only by nearby particles although the long-range Coulomb potential is involved.

Butyric Acid

As a last example, we compute the solvent density around butyric acid. Butyric acid is a carboxylic acid with structural formula $C_4H_8O_2$. It belongs to the so-called fatty acids which are in their form of esters constituents of all kinds of animal fats and plant oils. As esters, the fatty acids are bonded to a backbone structure as e.g. glycerol. Butyric acid is commonly known, since it is responsible for the unpleasant odor of rancid butter. For more details about butyric acid see e.g. [18]. We again use the OPLS [57] force field for the potential parameter set of butyric acid. The computational domain is set to $\Omega = [-16\text{Å}, 16\text{Å}]^3$.

Similar to methanol, butyric acid has a charged functional group, the COOH-group. Hence, we are again faced with the problem that the BGY3dM equations tend to overestimate the solvent carbon density next to strong positively charged particles. In order to account for this, we again choose a relatively large value for σ_H in the COOH-group, i.e., we set $\sigma_H = 3.4\text{Å}$ and $\epsilon_H = 0.03\text{kcal/mol}$ in this case. Additionally, we have to damp the Coulomb force by a factor of 0.9 in order to reach convergence of the BGY3dM equations. That corresponds to a reduction of all partial charges by a factor of $\sqrt{0.9} \approx 0.95$.

The results for the carbon and sulfur distribution functions around butyric acid are depicted in Figure 6.8. As before, the solute molecule is raised 2Å above the plane. The strong repulsion between the solvent carbon and solute oxygen atoms is evident. On the other hand, large peaks of the sulfur density can be found in

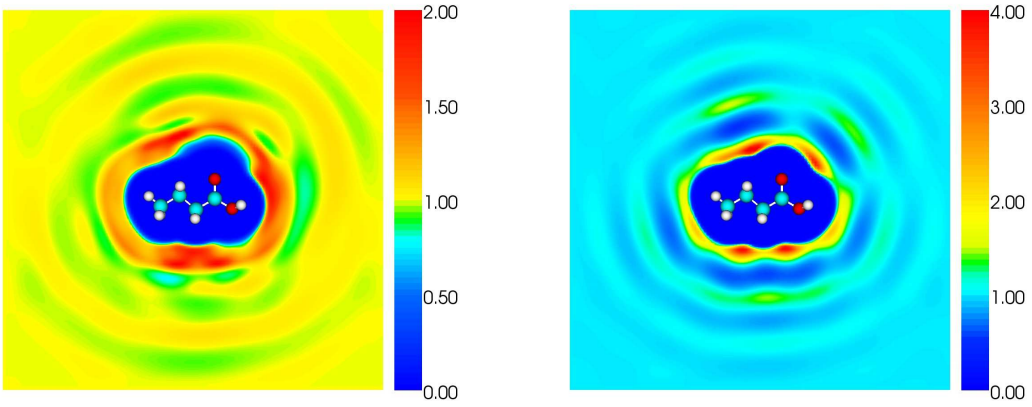

Figure 6.8. *Site distribution functions of carbon disulfide around butyric acid. Carbon distribution at the $x_3 = 0$ plane (left) and sulfur distribution at the $x_3 = 0$ plane (right).*

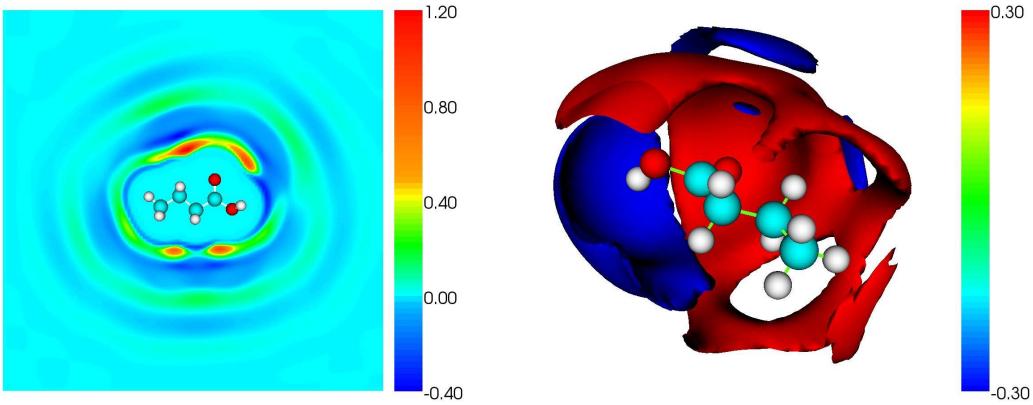

Figure 6.9. *Charge distribution of carbon disulfide around butyric acid. Cut at the $x_3 = 0$ plane (left) and isosurface plot (right).*

the vicinity of the solute oxygen atoms. The same behavior can be observed for the charge distributions of Figure 6.9. The strong positive charge of the hydrogen is neutralized by a negative charge cloud of the solvent, whereas the oxygen atoms are surrounded by positive solvent charges. This reveals a similar behavior as for methanol as solute. Strong charges of the solute are neutralized by opposite charge clouds of the solvent. Hence, the charge distribution of the solvent rapidly decays to zero for large distances to the solute.

Summary

We presented the density profiles of the solvent sites of carbon disulfide around several solutes. Homogeneous solutes as e.g. hexane lead to closed shells of high carbon and sulfur densities around the entire solute. Here, the van der Waals interaction dominates. The density profiles for solute molecules with stronger partial charges, as e.g. methanol and butyric acid, exhibit, however, distinct regions of higher carbon or sulfur density according to their charge distributions. Especially the solvent charge distributions revealed important properties of charged systems. Typical features of Coulomb systems, where the attraction of opposite-signed charges and the rejection of equal-signed charges lead to the neutralization of local charge distributions, can be observed. Hence, the solvent site distribution functions are of short range, even though the potential function is not. As already pointed out in Section 5.3.3, this can also be proved rigorously, see [1]. The numerical solution of our BGY3dM model showed that this important property is reproduced even for complex distribution functions in three-dimensions.

In summary, the BGY3dM equations are able to compute reasonable site density distributions around complex solute molecules in three dimensions. These site distributions can be used to compute the potential of mean force, i.e. the free energy of the solute-solvent system and the mean force exerted on the solute by the solvent. The BGY3dM method is, however, not yet stable with respect to the parameter sets of the interaction potentials involved. Certain combinations, in particular strong attractions between different particle species, do not lead to satisfactory results. In order to account for this, we slightly altered the potential parameters of the respective particle species. A different approach would be to improve the approximations contained in the BGY3dM equations by empirical functions that can be adapted to the solvent under consideration. This approach would be similar to the bridge functions of Beglov, Roux et al. [12,28] and Kovalenko and Hirata [63] employed for the computation of the site distribution functions of water. A further investigation along these lines is, however, beyond the scope of this thesis.

Chapter 7
Conclusions

We presented and investigated a novel computational model for the simulation of solute-solvent systems based on the YBG-hierarchy. The presented model allows for the efficient approximation of the solvent density around a solute of arbitrary shape. The solvent density can be used to compute the potential of mean force (PMF) which in turn allows for an incorporation of the solvent effects without the need to explicitly include solvent molecules into the simulation. This way, a more efficient simulation of the entire solute-solvent system becomes possible. Existing implicit solvent models approximate the PMF very poorly. Hence, it is important to develop methods that improve the accuracy of the implicit solvent models while keeping the computational effort tractable for repeated evaluation.

Promising developments were made by the application of the liquid state integral equation theories. Several authors developed methods based on these theories that can compute solvent densities around solutes of arbitrary shape. It stands out that practically all methods found in the literature are based on the Ornstein-Zernike equation and mostly employ the hypernetted chain (HNC) closure. However, these methods do not lead to satisfactory results in all situations. Additionally, the computational effort involved still makes a repeated evaluation during an extensive solute-solvent simulation unfeasible. To overcome these issues we pursued a different approach. We derived our BGY3d and BGY3dM models starting from the YBG-hierarchy together with the Kirkwood superposition approximation. Beside the investigation of the Kirkwood approximation which has never been applied in this context, the development of our method based on the YBG-hierarchy enabled us to investigate a new numerical algorithm which indeed proved to be superior with respect to the computational costs in the case of monoatomic solvents. With respect to accuracy our BGY3d and BGY3dM models yield to almost identical results when compared to methods based on the Ornstein-Zernike equation from the literature.

We derived our BGY3d model employing the Kirkwood approximation in order to compute the solvent density of simple monoatomic solvents around arbitrary

solutes. We are able to efficiently solve the non-linear BGY3d equation with full three-dimensional resolution by means of a fixed point iteration. In each step of this iteration a Poisson problem is solved by diagonal scaling in Fourier space. The pair-distribution functions which are required as input of the BGY3d equation are computed beforehand by the Born-Green equation which can be seen as a special case of the BGY3d equation. We compared the results to those obtained by the Ornstein-Zernike based 3d-HNC method of Beglov and Roux [10] and to distribution functions computed by molecular dynamics simulations. We found that our approach gave a similar overall accuracy as the 3d-HNC method. A more detailed analysis shows that our BGY3d method is superior in the prediction of the height and position of the first peak of the distribution functions, whereas the 3d-HNC method leads to a better agreement of the subsequent oscillation pattern. Moreover, the BGY3d model proved to be superior to the 3d-HNC model concerning the computational costs. This is due to the considerably reduced number of non-linear iteration steps of our method. Compared to the molecular dynamics simulation the computing time for our BGY3d method was four orders of magnitude smaller, which is a substantial gain with respect to computational effort.

In order to consider more realistic systems, we extended our model to molecular solvents. To this end, the molecular BGY3d (BGY3dM) model comprises terms for the intermolecular interactions as well as for the intramolecular interactions. The intramolecular terms were derived to model rigid bonds by taking the limit of an infinite restoring force between two bonded particles. This way, the solvent molecules are represented as rigid bodies. As before, the intermolecular terms incorporate the Kirkwood superposition approximation. This approximation, however, is not appropriate for the intramolecular terms. Here, different approximations have to be employed in order to ensure the correct asymptotic behavior. For this, we incorporated a slightly modified version of the normalized site-site superposition approximation (NSSA) of Taylor and Lipson [116]. The intramolecular terms together with the NSSA approximation turned out to model the rigid bonds of the solvent molecules very well. As a further advance of the method, the optimal approximation for the intramolecular terms as derived by Attard [4] could be incorporated.

Beside the short-range Lennard-Jones potential, we also considered the long-range Coulomb interaction. For this, we introduced a splitting of the Coulomb potential into a singular short-range part and a smooth long-range part. The short-range part is processed in exactly the same way as the Lennard-Jones potential. The long-range part has fast decaying analytic Fourier components and is therefore directly inserted in Fourier space. Nevertheless, the inverse Fourier transform of this long-range part leads to undesirable boundary conditions that have to be corrected. The correction comprises the solution of an additional Laplace problem. It can efficiently be solved by a finite difference scheme with an iterative GMRES solver. Finally, we also derived the site-site BGY3dM (SS-BGY3dM) equations in order

to compute the site-site pair distribution functions of the pure solvent which are required as input of the BGY3dM model.

A comparison of the results computed by the (SS-)BGY3dM model and by molecular dynamics showed a similar performance as in the case of monoatomic solvents. All important characteristics of the site-site pair distribution functions and the site density distributions are reproduced. Hence, the general form of the (SS-)BGY3dM model including the modeling of the intramolecular bonds is validated by the results. Likewise, the application of the BGY3dM model to the site density computation of carbon disulfide as solvent around several realistic solutes lead to reasonable density and charge distributions. The reduction of the computational effort compared to a molecular dynamics simulation for the computation of the site densities is substantial even in the case of complex molecular solvents. For a two-site model of a HCl-like solvent the BGY3dM method computed the results about two-three orders of magnitudes faster. This is an important step in the development of an implicit solvent model that allows repeated evaluations of an accurate approximation to the PMF.

It is a well-known problem that methods based on the liquid state integral equation theories do not lead to satisfactory results in all situations [47]. This is due to the approximations involved. In general, all methods perform worse with increasing density, decreasing temperature or for strong interactions between two particle species. Especially the computation of densities for water as solvent is very challenging due to the strong Coulomb interaction between the oxygen and hydrogen sites of the H_2O molecule. Water is the most important fluid in general and as a solvent. It is known to be a very complex fluid and there exist many different empirical water models intended for microscopic simulation. But none of these models is able to reproduce all of the properties of water with appropriate accuracy. Hence, the model is usually chosen such that it reproduces well only one property that is assumed to be important for the respective application. Most general purpose force fields such as CHARMM [16], AMBER [23] and OPLS [57] employ the TIP3P [55] and TIP4P [56] water models for the simulation of biomolecular systems. Hence, the application of the integral equation theories to the TIP3P water model is an important test. Both the 3d-RISM-HNC method of Du, Beglov and Roux [28] and the 3d-HNC-PLHNC method of Kovalenko and Hirata [63] did not lead to accurate results for TIP3P water. Hence, the methods were extended by so-called empirical bridge functions in order to improve their accuracy. These bridge functions account for the overestimation of water ordering around a hydrophobic solute.

Our experiments with the (SS-)BGY3dM model revealed that its application to TIP3P water is also very challenging. It is known that the Kirkwood approximation can cause an overestimation of the first peak of the pair distribution or site density function [3] for high densities or strong interactions. This is a result of the neglected three-particle interaction. Due to this approximation error the BGY3dM and SS-

BGY3dM models do not lead to satisfactory results or even do not converge in certain situations. One approach to overcome this problem would be the introduction of empirical corrections as it was described for the Ornstein-Zernike based methods [28, 63]. By this, one can account for deficiencies which are characteristic for the employed approximation and the respective application. This has already been suggested by Attard [3] concerning the Kirkwood approximation. It would however require an extensive adjustment of the empirical functions and is therefore beyond the scope of this thesis.

The major advantage of an implicit solvent model based on the integral equation theories is the enormous reduction of the computational effort needed in order to include the solute-solvent effects in a microscopic simulation. However, a considerable reduction of the computing time would still be necessary for any integral equation method in order to allow for repeated evaluation of the potential of mean force during a Monte Carlo or molecular dynamics simulation of a complex solute-solvent system. Concerning this application they still cannot replace the classical implicit solvent models, which approximate the potential of mean force very roughly. In order to enable an even more efficient numerical solution of our BGY3d and BGY3dM models two approaches are conceivable: First, a finite element discretization of the BGY3d or BGY3dM equations could yield a reduction of the degrees of freedom and thereby a decrease of the computational effort. But it is important to keep in mind that the discretization should allow for an efficient evaluation of the convolution integrals appearing in both models. This could be achieved for example by a low-rank approximation of the convolution matrices as it has been employed by Fedorov et al. [33] in combination with a wavelet basis for the solution of the radial symmetric Ornstein-Zernike equation. Moreover, an efficient representation of the occurring matrices could also enable the application of Newton's method in order to solve the non-linear equations. This would improve the convergence of the non-linear iteration and thereby the computing time. Our current discretization would however lead to full Jacobi matrices, which is acceptable only on a coarse grid resolution. Hence, Newton's method could only serve as a coarse level solver in a multi-grid like scheme or special matrix compression techniques would have to be employed.

In summary, we developed our BGY3dM model based on the YBG-hierarchy and showed that it is able to efficiently compute the solvent densities of complex molecular solvents around solutes of arbitrary shape with good accuracy. By this, it is possible to compute the potential of mean force for the solute-solvent system and thereby to consider the solvent effects on the solute implicitly. For the first time we considered the Kirkwood approximation in this context. When compared to an Ornstein-Zernike based method from the literature, our numerical algorithm proved to compute the solvent densities in a highly efficient way. The accuracy of the results can be improved by empirical corrections in the same way as it is done for

the Ornstein-Zernike based methods. Hence, the BGY3dM model has demonstrated to be applicable to any complex solute-solvent system. However, further research is necessary in order to even more reduce the computational effort required for the numerical solution of the BGY3dM model such that it can be applied as implicit solvent model during a large-scale realistic biomolecular simulation.

Appendix A

Convolution of Spherical Symmetric Functions

We want to compute the convolution of two spherical symmetric functions,

$$f_1, f_2 : \mathbb{R}^3 \to \mathbb{R},$$

$$(f_1 * f_2)(\mathbf{x}) = \int_{\mathbb{R}^3} f_1(|\mathbf{x}'|) f_2(|\mathbf{x} - \mathbf{x}'|)\ d\mathbf{x}'. \tag{A.1}$$

According to the convolution theorem, the convolution can be computed by means of the Fourier transform. It is

$$(f_1 * f_2) = \mathcal{F}_3^{-1}\left(\mathcal{F}_3(f_1)\mathcal{F}_3(f_2)\right) \tag{A.2}$$

with the Fourier transform and its inverse in 3 dimensions,

$$\tilde{f}(\mathbf{k}) = \mathcal{F}_3(f)(\mathbf{k}) = \int_{\mathbb{R}^3} f(\mathbf{x}) e^{-2\pi\imath\mathbf{k}\cdot\mathbf{x}}\ d\mathbf{x}, \tag{A.3}$$

$$f(\mathbf{x}) = \mathcal{F}_3^{-1}(\tilde{f})(\mathbf{x}) = \int_{\mathbb{R}^3} \tilde{f}(\mathbf{k}) e^{2\pi\imath\mathbf{k}\cdot\mathbf{x}}\ d\mathbf{k}. \tag{A.4}$$

The three-dimensional Fourier transform of a spherical symmetric function can be further simplified. Due to the invariance of integral (A.3) under rotation, it is sufficient to consider the case $\mathbf{k} = (0, 0, k)$. Then we have in polar coordinates $\mathbf{k} \cdot \mathbf{x} = kr \sin\theta$ with $\mathbf{x} = (r\cos\varphi\cos\theta, r\sin\varphi\cos\theta, r\sin\theta)$ and (A.3) simplifies to

$$\begin{aligned}
\mathcal{F}_3(f)(k) &= \int_0^\infty \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{-\pi}^{\pi} f(r) e^{-2\pi\imath kr\sin\theta} r^2 \cos\theta\ d\varphi d\theta dr \\
&= 2\pi \int_0^\infty \frac{f(r)r^2}{-2\pi\imath kr} \left[e^{-2\pi\imath kr\sin\theta}\right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\ dr \\
&= \frac{2}{k}\int_0^\infty f(r) r \sin(2\pi kr)\ dr \qquad (A.5) \\
&=: \mathcal{FB}(f)(k) \qquad (A.6)
\end{aligned}$$

The analogous simplification holds for the inverse Fourier transform, (A.4)

$$\mathcal{F}_3^{-1}(\tilde{f})(r) = \frac{2}{r}\int_0^\infty \tilde{f}(k)k\sin(2\pi kr)\ dk = \mathcal{FB}^{-1}(\tilde{f})(r). \tag{A.7}$$

The transforms (A.5) and (A.7) are called the Fourier-Bessel transforms. The radial component of the convolution (A.2) can now be easily computed as

$$(f_1 * f_2) = \mathcal{FB}^{-1}\left(\mathcal{FB}(f_1)\cdot\mathcal{FB}(f_2)\right). \tag{A.8}$$

Appendix B

Transformations of the Ornstein-Zernike Equation

In order to solve the Ornstein-Zernike equation some transformations are very useful. For a homogeneous fluid, the Ornstein-Zernike equation (2.56) can be written as

$$h(r) = c(r) + \rho \int_\Omega c(|\mathbf{r} - \mathbf{r}'|)h(r')\ d\mathbf{r}' \tag{B.1}$$

with $r = |\mathbf{r}|$ and $r' = |\mathbf{r}'|$. The integral is a convolution integral. Hence, we can use the convolution theorem which states that

$$\mathcal{F}_3(c * h) = \mathcal{F}_3(c)\mathcal{F}_3(h) \tag{B.2}$$

with $\mathcal{F}_3$ the three-dimensional Fourier transform. The asterix $*$ denotes the convolution

$$(c * h)(r) = \int_\Omega c(|\mathbf{r} - \mathbf{r}'|)h(r')\ d\mathbf{r}'. \tag{B.3}$$

Since the functions c and h are spherically symmetric, the three-dimensional Fourier transform reduces to the Fourier-Bessel transform

$$\mathcal{FB}(c)(k) = \frac{2}{k} \int_0^\infty c(r) r \sin(2\pi k r)\ dr, \tag{B.4}$$

i.e., the radial component of $\mathcal{F}_3(c)(\mathbf{k})$ can be computed by a one-dimensional integral, i.e. $\mathcal{F}_3(c)(|\mathbf{k}|) = \mathcal{FB}(c)(k)$. For a detailed derivation of the Fourier-Bessel transform see Appendix A.

The Ornstein-Zernike equation (B.1) has a nice algebraic form in Fourier space and is therefore often written in this form,

$$\hat{h}(k) = \hat{c}(k) + \rho \hat{c}(k)\hat{h}(k), \tag{B.5}$$

where we write $\hat{h} = \mathcal{F}_3(h)$. It is sometimes advantageous to use $\gamma = h - c$ in (B.5) instead of h,

$$\hat{\gamma}(k) = \frac{\rho \hat{c}^2(k)}{1 - \rho \hat{c}(k)}. \tag{B.6}$$

Together with either the HNC

$$c(r) = e^{-\beta v(r)+\gamma(r)} - \gamma(r) - 1 \tag{B.7}$$

or the PY closure

$$c(r) = e^{-\beta v(r)}(1 + \gamma(r)) - \gamma(r) - 1, \tag{B.8}$$

equation (B.6) can be solved very efficiently by an iteration scheme, see [47] for details.

Bibliography

[1] A. Alastuey and P. A. Martin, *Decay of Corrleations in Classical Fluids with Long-Range Forces*, J. Stat. Phys., 39 (1985), pp. 405–426.

[2] M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*, Oxford Science Publications, Oxford, 1987.

[3] P. Attard, *An Improved Kirkwood Superposition Approximation for Three Atoms in Rolling Contact*, Mol. Phys., 74 (1991), pp. 547–552.

[4] ——, *Polymer Born-Green-Yvon Equation with Proper Triplet Superposition Approximation. Results for Hard-Sphere Chains*, J. Chem. Phys., 102 (1995), pp. 5411–5426.

[5] S. Balay, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang, *PETSc users manual*, Tech. Rep. ANL-95/11 - Revision 2.1.5, Argonne National Laboratory, 2004.

[6] S. Balay, K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang, *PETSc Web page*, 2001. http://www.mcs.anl.gov/petsc.

[7] S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith, *Efficient management of parallelism in object oriented numerical software libraries*, in Modern Software Tools in Scientific Computing, E. Arge, A. M. Bruaset, and H. P. Langtangen, eds., Birkhäuser Press, 1997, pp. 163–202.

[8] R. Balescu, *Statistical Dynamics, Matter out of Equilibrium*, Imperial College Press, Imperial College, London, 1997.

[9] D. Beglov and B. Roux, *Finite Representation of an Infinite Bulk System*, J. Chem. Phys., 100 (1994), pp. 9050–9063.

[10] ——, *Numerical Solution of the Hypernetted Chain Equation for a Solute of Arbitrary Geometry in Three Dimensions*, J. Chem. Phys., 103 (1995), pp. 360–364.

[11] ——, *Solvation of Complex Molecules in a Polar Liquid: An Integral Equation Theory*, J. Chem. Phys., 104 (1996), pp. 8678–8689.

[12] ——, *An Integral Equation to Describe the Solvation of Polar Molecules in Liquid Water*, J. Phys. Chem. B, 101 (1997), pp. 7821–7826.

[13] R. Bellman, *Dynamic Programming*, University Press, Princeton, 1957.

[14] L. Blum, *Invariant Expansion. II. The Ornstein-Zernike Equation for Non-spherical Molecules and an Extended Solution to the Mean Spherical Model*, J. Chem. Phys., 57 (1972), pp. 1862–1869.

[15] L. Blum and A. J. Torruella, *Invariant Expansion for Two-Body Correlations: Thermodynamic Functions, Scattering, and the Ornstein-Zernike Equation*, J. Chem. Phys., 56 (1972), pp. 303–310.

[16] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations*, J. Comp. Chem., 4 (1983), pp. 187–217.

[17] C. L. Brooks III and M. Karplus, *Deformable Stochastic Boundaries in Molecular Dynamics*, J. Chem. Phys., 79 (1983), pp. 6312–6325.

[18] *Butyric Acid, in Wikipedia, The Free Encyclopedia*, July 2007. http://en.wikipedia.org/wiki/Butyric_acid.

[19] *Carbon Disulfide Fact Sheet*, July 2007. http://www.npi.gov.au/database/substance-info/profiles/18.html.

[20] D. Chandler and H. C. Andersen, *Optimized Cluster Expansion for Classical Fluids. II. Theory of Molecular Liquids*, J. Chem. Phys., 57 (1972), pp. 1930–1937.

[21] D. Chandler, J. D. McCoy, and S. J. Singer, *Density Functional Theory of Nonuniform Polyatomic Systems. I. General Formulation*, J. Chem. Phys., 85 (1986), pp. 5971–5976.

[22] M. Connolly, *Analytical Molecular Surface Calculation*, J. Appl. Cryst., 16 (1983), pp. 548–558.

[23] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. J. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules*, J. Am. Chem. Soc., 117 (1995), pp. 5179–5197.

[24] C. M. Cortis, P. J. Rossky, and R. A. Friesner, *A Three-Dimensional Reduction of the Ornstein-Zernike Equation for Molecular Fluids*, J. Chem. Phys., 107 (1997), pp. 6400–6414.

[25] P. T. Cummings and G. Stell, *Interaction Site Models for Molecular Fluids*, Mol. Phys., 46 (1982), pp. 383–426.

[26] T. Darden, D. York, and L. Pedersen, *Particle Mesh Ewald: An nlog(n) Method for Ewald Sums in Large Systems*, J. Chem. Phys., (1993).

[27] J. P. Donley, J. G. Curro, and J. D. McCoy, *A Density Functional Theory for Pair Correlation Functions in Molecular Liquids*, J. Chem. Phys., 101 (1994), pp. 3205–3215.

[28] Q. Du, D. Beglov, and B. Roux, *Solvation Free Energy of Polar and Nonpolar Molecules in Water: An Extended Interaction Site Integral Equation Theory in Three Dimensions*, J. Phys. Chem., 104 (2000), pp. 796–805.

[29] H. Edelsbrunner, *The Union of Balls and its Dual Shape*, Discrete Comput. Geom., 13 (1995), pp. 415–440.

[30] H. Edelsbrunner and E. Mucke, *Simulation of Simplicity: A Technique to Cope with Degenerate Cases in Geometric Algorithms*, ACM Trans. Graphics, 9 (1990), pp. 66–104.

[31] F. Eisenhaber and P. Argos, *Improved Strategy in Analytic Surface Calculation for Molecular Systems: Handling of Singularities and Computational Efficiency*, J. Comput. Chem., 14 (1993), pp. 1272–1280.

[32] B. C. Eu and H. H. Gan, *Integral Equations of the Correlation Functions for Polymeric Liquids*, J. Chem. Phys., 99 (1993), pp. 4084–4102.

[33] M. V. Fedorov, H.-J. Flad, G. N. Chuev, L. Grasedyck, and B. N. Khoromskij, *A Structured Low-Rank Wavelet Solver for the Ornstein-Zernike Integral Equation*, Computing, 80 (2007), pp. 47–73.

[34] P. Ferrara, J. Apostolakis, and A. Caflisch, *Evaluation of a Fast Implicit Solvent Model for Molecular Dynamics Simulations*, 2001, 46 (Proteins), pp. 24–33.

[35] I. Z. Fisher and B. I. Kopeliovich, *On a Refinement of the Superposition Approximation in the Theory of Fluids*, Dokl. Akad. Nauk. SSSR, 133 (1960), p. 81.

[36] R. Fraczkiewicz and W. Braun, *Exact and Efficient Analytical Calculation of the Accessible Surface Area and their Gradient for Macromolecules*, J. Comput. Chem., 19 (1998), pp. 319–333.

[37] P. H. Fries, W. Kunz, P. Calmettes, and P. Turq, *Molecular Solvent Model for a Cryptate Solution in Acetonitrile: A Hypernetted Chain Study*, J. Chem. Phys., 101 (1994), pp. 554–577.

[38] P. H. Fries and G. N. Patey, *The Solution of the Hypernetted-Chain Approximation for Fluids of Nonspherical Particles. A general Method with Application to Dipolar Hard Spheres*, J. Chem. Phys., 82 (1985), pp. 429–440.

[39] M. Frigo and S. G. Johnson, *The Design and Implementation of FFTW3*, in Special Issue on Program Generation, Optimization, and Platform Adaptation, vol. 93 (2) of Proceedings of the IEEE, 2005, pp. 216–231.

[40] H. H. Gan and B. C. Eu, *Application of the Integral Equation Theory of Polymers: Distribution Function, Chemical Potential, and Mean Expansion Coefficient*, J. Chem. Phys., 99 (1993), pp. 4103–4111.

[41] ——, *Self-Consistent Field Equations in the Distribution Function Theory of Polymeric Liquids*, Journal of Polymeric Science: Part B: Polymer Physics, 33 (1995), pp. 2319–2329.

[42] ——, *Polymer Kirkwood Integral Equations: Structure and Equation of State of Polymeric Liquids*, AIChE Journal: Materials, Interfaces, and Electrochemical Phenomena, 42 (1996), pp. 2960–2966.

[43] ——, *Integral Equation Theory of Single-Chain Polymers: Comparison with Simulation Data for Hard-Sphere and Square-Well Chains*, J. Chem. Phys., 110 (1999), pp. 3235–3240.

[44] V. Gogonea and E. Osawa, *An Improved Algorithm for the Analytical Computation of Solvent-excluded Volume. the Treatment of Singularities in Solvent Accessible Surface Area and Volume Functions*, J. Comput. Chem., 16 (1995), pp. 817–842.

[45] C. G. Gray and K. E. Gubbins, *Fundamentals*, vol. 1 of Theory of Molecular Fluids, Clarendon Press, Oxford, 1984.

[46] M. Griebel, S. Knapek, and G. Zumbusch, *Numerical Simulation in Molecular Dynamics. Numerics, Algorithms, Parallelization, Applications*, Texts in Computational Science and Engineering, Springer, 2007.

[47] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Academic Press, London, 2. ed., 1986.

[48] *Himalaya, High-Performance Cluster Computers at the INS and SFB 611, University of Bonn*, July 2007. http://wissrech.ins.uni-bonn.de/research/himalaya/.

[49] F. Hirata, B. M. Pettitt, and P. J. Rossky, *Application of an Extended RISM Equation to Dipolar and Quadrupolar Fluids*, J. Chem. Phys., 77 (1982), pp. 509–520.

[50] F. Hirata and J. Rossky, *An Extended RISM Equation for Molecular Polar Fluids*, Chem. Phys. Lett., 83 (1981), pp. 329–334.

[51] F. Hirata, P. J. Rossky, and B. M. Pettitt, *The Interionic Potential of Mean Force in a Molecular Polar Solvent from an Extended RISM Equation*, J. Chem. Phys., 78 (1983), pp. 4133–4144.

[52] R. Hockney and J. Eastwood, *Computer Simulations using Particles*, IOP Publishing Ltd., London, 1988.

[53] M. Ikeguchi and J. Doi, *Direct Numerical Solution of the Ornstein-Zernike Integral Equation and Spatial Distribution of Water around Hydrophobic Molecules*, J. Chem. Phys., 103 (1995), pp. 5011–5017.

[54] B. Jayaram, D. Sprous, and D. L. Beveridge, *Solvation Free Energy of Biomacromolecules: Parameters for a Modified Generalized Born Model Consistent with the AMBER Force Field*, J. Phys. Chem. B, 102 (1998), pp. 9571–9576.

[55] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *Comparison of Simple Potential Functions for Simulating Liquid Water*, J. Chem. Phys., 79 (1983), pp. 926–935.

[56] W. L. Jorgensen and J. D. Madura, *Temperature and Size Dependence for Monte Carlo Simulations of TIP4P Water*, Mol. Phys., 56 (1985), pp. 1381–1392.

[57] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids*, J. Am. Chem. Soc., 118 (1996), pp. 11225–11236.

[58] J. Kerins, L. E. Scriven, and H. T. Davis, *Correlation Functions in Subcritical Fluid*, Adv. Chem. Phys., 65 (1986), pp. 215–279.

[59] J. G. Kirkwood, *Statistical Mechanics of Fluid Mixtures*, J. Chem. Phys., 3 (1935), pp. 300–313.

[60] K. Königsberger, *Analysis 2*, Springer, Heidelberg, Germany, 1993.

[61] A. Kovalenko and F. Hirata, *Three-Dimensional Density Profiles of Water in Contact with a Solute of Arbitrary Shape: a RISM Approach*, Chem. Phys. Lett., 290 (1998), pp. 237–244.

[62] ——, *Potential of Mean Force between Two Molecular Ions in a Polar Molecular Solvent: A Study by the Three-Dimensional Reference Interaction Site Model*, J. Phys. Chem. B, 103 (1999), pp. 7942–7957.

[63] ——, *Hydration Free Energy of Hydrophobic Solutes Studied by a Reference Interaction Site Model with a Repulsive Bridge Correction and a Thermodynamic Perturbation Method*, J. Chem. Phys., 113 (2000), pp. 2793–2805.

[64] ——, *Hydration Structure and Stability of Met-enkephalin Studied by a Three-Dimensional Reference Interaction Site Model with a Repulsive Bridge Correction and a Thermodynamic Perturbation Method*, J. Chem. Phys., 113 (2000), pp. 9830–9836.

[65] ——, *Potentials of Mean Force of Simple Ions in Ambient Aqueous Solution. I. Three-Dimensional Reference Interaction Site Model Approach*, J. Chem. Phys., 112 (2000), pp. 10391–10402.

[66] ——, *Potentials of Mean Force of Simple Ions in Ambient Aqueous Solution. II. Solvation Structure from the Three-Dimensional Reference Interaction Site Model Approach, and Comparison with Simulations*, J. Chem. Phys., 112 (2000), pp. 10403–10417.

[67] A. Kovalenko and T. N. Truong, *Thermochemistry of Solvation: A Self-Consistent Three-Dimensional Reference Interaction Site Model Approach*, J. Chem. Phys., 113 (2000), pp. 7458–7470.

[68] J. L. Lebowitz and J. K. Percus, *Statistical Thermodynamics of Nonuniform Fluids*, J. Math. Phys., 4 (1963), pp. 116–123.

[69] B. Lee and F. M. Richards, *The Interpretation of Protein Structures: Estimation of Static Accessibility*, J. Mol. Biol., 55 (1971), pp. 379–400.

[70] Y.-T. Lee, F. H. Ree, and T. Ree, *Distribtution Function of Classical Fluids of Hard Spheres. I*, J. Chem. Phys., 48 (1968), pp. 3506–3515.

[71] S. Legrand and K. Merz, *Rapid Approximation to Molecular Surface Area via the use of Boolean Logic and Lookup Tables*, J. Comp. Chem., 14 (1993), pp. 349–352.

[72] B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schütte, and R. Skeel, eds., *New Algorithms for Macromolecular Simulation*, vol. 49 of Lecture Notes in Computational Science and Engineering, Springer, Heidelberg, Germany, 2006.

[73] J. Liang, H. Edelsbrunner, P. Fu, P. Sudhakar, and S. Subramaniam, *Analytical Shape Computation of Macromolecules i and ii*, Proteins, 33 (1998), pp. 1–17, 18–29.

[74] J. E. G. Lipson, *A Born-Green-Yvon Integral Equation Treatment of Incompressible Lattice Mixtures*, J. Chem. Phys., 96 (1992), pp. 1418–1425.

[75] J. E. G. Lipson and S. S. Andrews, *A Born-Green-Yvon Integral Equation Treatment of Compressible Fluid*, J. Chem. Phys., 96 (1992), pp. 1426–1434.

[76] L. Lue and D. Blankschtein, *Liquid-State Theory of Hydrocarbon-Water Systems: Application to Methane, Ethane, and Propane*, J. Phys. Chem., 96 (1992), pp. 8582–8594.

[77] E. Meeron, *Series Expansion of Distribution Functions in Multicomponent Fluid Systems*, J. Chem. Phys., 27 (1957), pp. 1238–1246.

[78] *Methanol Fact Sheet*, July 2007. http://www.npi.gov.au/database/substance-info/profiles/54.html.

[79] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A. Teller, *Equation of State Calculations by Fast Computing Machines*, J. Chem. Phys., 21 (1953), pp. 1087–1092.

[80] *n-Hexane Fact Sheet*, July 2007. http://www.npi.gov.au/database/substance-info/profiles/47.html.

[81] M. Nina, D. Beglov, and B. Roux, *Atomic Radii for Continuum Electrostatics Calculations Based on Molecular Dynamics Free Energy Simulations*, J. Phys. Chem. B, 101 (1997), pp. 5239–5248.

[82] L. S. Ornstein and F. Zernike, *Accidental Deviations of Density and Opalescence at the Critical Point in a Single Substance*, Proc. Acad. Sci. Amsterdam, 17 (1914), pp. 793–806.

[83] J. Perkyns and B. M. Pettitt, *A Site-Site Theory for Finite Concentration Saline Solutions*, J. Chem. Phys., 97 (1992), pp. 7656–7666.

[84] B. M. Pettitt and M. Karplus, *The Potential of Mean Force between Polyatomic Molecules in Polar Molecular Solvents*, J. Chem. Phys., 83 (1985), pp. 781–789.

[85] ——, *The Structure of Water Surrounding a Peptide: A Theoretical Approach*, Chem. Phys. Lett., 136 (1987), pp. 383–386.

[86] B. M. Pettitt, M. Karplus, and P. J. Rossky, *Inetgral Equation Model for Aqueous Solvation of Polyatomic Solutes: Application to the Determination of the Free Energy Surface of the Internal Motion of Biomolecules*, J. Phys. Chem., 90 (1986), pp. 6335–6345.

[87] B. M. Pettitt and P. J. Rossky, *Integral Equation Predictions of Liquid State Structure for Waterlike Intermolecular Potentials*, J. Chem. Phys., 77 (1982), pp. 1451–1457.

[88] ——, *The Contribution of Hydrogen Bonding to the Structure of Liquid Methanol*, J. Chem. Phys., 78 (1983), pp. 7296–7299.

[89] ——, *Alkali Halides in Water: Ion-Solvent Sorrelations and Ion-Ion Potentials of Mean Force at Infinite Dilution*, J. Chem. Phys., 84 (1986), pp. 5836–5844.

[90] R. A. Pierotti, *A Scaled Particle Theory of Aqueous and Nonaqueous Solutions*, Chem. Rev., 76 (1976), pp. 717–726.

[91] J. Ponder, *http://dasher.wustl.edu/tinker/*.

[92] N. V. Prabhu, J. S. Perkyns, H. D. Blatt, P. E. Smith, and B. M. Pettitt, *Comparison of the Potential of Mean Force for Alanine Tetrapeptide between Integral Equation Theory and Simulation*, Biophysical Chemistry, 78 (1999), pp. 113–126.

[93] G. Reddy, C. P. Lawrence, J. L. Skinner, and A. Yethiraj, *Liquid State Theories for the Structure of Water*, J. Chem. Phys., 119 (2003), pp. 13012–13016.

[94] F. H. Ree, Y.-T. Lee, and T. Ree, *Distribtution Function of Classical Fluids of Hard Spheres. II*, J. Chem. Phys., 55 (1971), pp. 234–245.

[95] H. Reiss, *Superposition Approximations from a Variation Principle*, J. Stat. Phys., 6 (1972), pp. 39–47.

[96] H. Reiss, H. L. Frisch, and J. L. Lebowitz, *Statistical Mechanics of Rigid Spheres*, J. Chem. Phys., 31 (1959), pp. 369–380.

[97] S. A. Rice and J. Lekner, *On the Equation of State of the Rigid-Sphere Fluid*, J. Chem. Phys., 42 (1965), pp. 3559–3565.

[98] S. A. Rice and D. A. Young, *Equation of State of a Monatomic Fluid with 6-12 Potential*, Discuss. Faraday Soc., 43 (1967), pp. 16–25.

[99] C. Richardi, J. adn Millot and P. H. Fries, *A Molecular Ornstien-Zernike Study of Popular Models for Water and Methanol*, J. Chem. Phys., 110 (1998), pp. 1138–1147.

[100] J. Richardi, P. H. Fries, R. Fischer, S. Rast, and H. Krienke, *Liquid Acetone and Chloroform: A Comparison between Monte Carlo Simulation, Molecular Ornstein-Zernike Theory, and Site-Site Ornstein-Zernike Theory*, Mol. Phys., 93 (1998), pp. 925–938.

[101] T. Richmond, *Solvent Accessible Surface Area and Excluded Volume in Proteins. Analytical Equations for Overlapping Spheres and Implications for the Hydrophobic Effect*, J. Mol. Biol., 178 (1984), pp. 63–89.

[102] M. Rieth and W. Schommers, eds., *Handbook of Theoretical and Computational Nanotechnology*, American Scientific Publishers, 2006.

[103] B. Roux, *Implicit Solvent Models*, in Computational Biophysics, O. Becker, A. D. MacKerrel, B. Roux, and M. Watanabe, eds., Marcel Dekker Inc, New York, 2001.

[104] B. Roux and T. Simonson, *Implicit Solvent Models*, Biophys. Chem., 78 (1999), pp. 1–20.

[105] E. E. Salpeter, *On Mayer's Theory of Cluster Expansions*, Ann. Phys., 5 (1958), pp. 183–223.

[106] T. Schlick and H. H. Gan, eds., *Computational Methods for Macromolecules: Challenges and Applications*, vol. 24 of Lecture Notes in Computational Science and Engineering, Springer, Heidelberg, Germany, 2002.

[107] A. Shrake and J. Rupley, *Environment and Exposure to Solvent of Protein Atoms in Lysozyme and Insulin*, J. Mol. Biol., 79 (1973), pp. 351–371.

[108] A. Singer, *Maximum Entropy Formulation of the Kirkwood Superposition Approximation*, J. Chem. Phys., 121 (2004), pp. 3657–3666.

[109] M. Sippl, *Calculation of Conformation Ensembles from Potentials of Mean Force*, J. Mol. Biol., 213 (1990), pp. 859–883.

[110] D. Sitkoff, K. A. Sharp, and B. Honig, *Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models*, J. Phys. Chem., 98 (1994), pp. 1978–1988.

[111] W. C. Still, A. Tempczyk, and R. C. Hawley, *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics*, J. Am. Chem. Soc., 112 (1990), pp. 6127–6129.

[112] F. Stillinger, *Structure in Aqueous Solutions of Nonpolar Solutes from the Standpoint of Scaled-Particle Theory*, J. Solution Chem., 2 (1973), pp. 141–158.

[113] T. Sumi and F. Hirata, *A Density-Functional Theory for Polymer Liquids based on the Interaction Site Model*, J. Chem. Phys., 118 (2003), pp. 2431–2442.

[114] T. Sumi, T. Imai, and F. Hirata, *Integral Equations for Molecular Fluids based on the Interaction Site Model: Density-Functional Formulation*, J. Chem. Phys., 115 (2001), pp. 6653–6662.

[115] T. Sumi and H. Sekino, *An Interaction Site Model Integral Equation Study of Molecular Fluids Explicitly Considering the Molecular Orientation*, J. Chem. Phys., (2006).

[116] M. P. Taylor and J. E. G. Lipson, *A Site-Site Born-Green-Yvon Equation for Hard Sphere Dimers*, J. Chem. Phys., 100 (1994), pp. 519–527.

[117] ——, *A Born-Green-Yvon Equation for Flexible Chain-Molecule Fluids. II. Applications to Hard-Sphere Polymers*, J. Chem. Phys., 102 (1995), pp. 6272–6279.

[118] ——, *A Born-Green-Yvon Equation for Flexible Chain-Molecule Fluids. I. General Formalism and Numerical Results for Short Hard-Sphere Chains*, J. Chem. Phys., 102 (1995), pp. 2118–2125.

[119] ——, *Collapse of a Polymer Chain: A Born-Green-Yvon Integral Equation Study*, J. Chem. Phys., 104 (1996), pp. 4835–4841.

[120] ——, *A Born-Green-Yvon Integral Equation Theory of Self-Interacting Lattice Polymers*, J. Chem. Phys., 109 (1998), pp. 7583–7590.

[121] M. P. Taylor, J. Luettmer-Strathmann, and J. E. G. Lipson, *Structure and Phase Behavior of Square-Well Dimer Fluids*, J. Chem. Phys., 114 (2001), pp. 5654–5662.

[122] M. P. Taylor, J. L. Mar, and J. E. G. Lipson, *Collapse of a Ring Polymer: Comparison of Monte Carlo and Born-Green-Yvon Integral Equation Results*, 1997, 106 (J. Chem. Phys.), pp. 5181–5188.

[123] *Tremolo - A Parallel Molecular Dynamics Software Package*. http://wissrech.iam.uni-bonn.de/research/projects/tremolo/, 2007.

[124] B. Von Freyberg, T. Richmond, and W. Braun, *Surface Area included in Energy Refinements of Proteins: a Comparative Study on Atomic Solvation Parameters*, J. Mol. Biol., (1993).

[125] R. Wawak, K. Gibson, and H. Scheraga, *Gradient Discontinuities in Calculations involving Molecular Surface Area*, J. Math. Chem., 15 (1994), pp. 207–232.

[126] S. G. Whittington and L. G. Dunfield, *A Born-Green-Yvon Treatment of Polymers with Excluded Volume*, J. Phys. A, 6 (1973), pp. 484–489.

[127] A. Yethiraj, H. Fynewever, and C.-Y. Shew, *Density Functional Theory for Pair Correlation Functions in Polymeric Liquids*, J. Chem. Phys., 114 (2001), pp. 4323–4330.

[128] R. J. Zauhar and R. S. Morgan, *A New Method for Computing the Macromolecular Electric Potential*, J. Mol. Biol., 186 (1985), pp. 815–820.

[129] S.-B. Zhu, J. Lee, and G. W. Robinson, *Molecular Dynamics Simulation of Liquid Carbon Disulphide with a Harmonic Intramolecular Potential*, Mol. Phys., 65 (1988), pp. 65–75.

[130] D. A. Zichi and P. J. Rossky, *Molecular Conformational Equilibria in Liquids*, J. Chem. Phys., 84 (1986), pp. 1712–1723.