The BGY3dM Model for the Approximation of Solvent Densities

Michael Griebel* and Lukas Jager†
Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany.

We present a new approach, the BGY3dM model, for the approximation of solvent densities around solutes of arbitrary shape. Our model is derived directly from the YBG-hierarchy and comprises the famous Kirkwood approximation as closure relation. The molecules of the solvent are modeled as rigid bodies by taking the limit of an infinite restoring force for the intramolecular interactions. Furthermore, short-range potentials as well as the long-range Coulomb interaction are taken into account. The resulting integro-differential equations are efficiently solved by a Picard iteration and a solution of the linearized equations using Fourier transformations. We compare the results obtained from the presented BGY3dM method with results obtained by extensive molecular dynamics simulations for a HCl-like model solvent. Furthermore, we apply the method to carbon disulfide as solvent. The overall performance of the method is promising.

AMS subject classifications: 82-08, 82D15

Keywords: Kirkwood, YBG-Hierarchy, Ornstein-Zernike, Implicit Solvent Model, Fluid Density Approximation

I. INTRODUCTION

The microscopic simulation of molecules such as proteins in solution is a challenging task. An explicit simulation of the entire solute-solvent system is often unfeasible due to the high number of degrees of freedom needed to adequately simulate the solvent effects [1]. Therefore, implicit solvent models have been developed which take the influence of the solvent into consideration without explicitly introducing new degrees of freedom to the system. Most of these implicit solvent models approximate the solvent effects by a continuum model which clearly neglects important local properties of the solvent [2, 3]. Hence, the development of new implicit solvent models which approximate the solvent effects more accurately is a key-topic of current research.

Promising developments were made by the application of the liquid state integral equation theories. Several authors developed methods based on these theories which allow to compute solvent densities around solutes of arbitrary shape for simple monoatomic [4, 5] as well as for molecular solvents [6–14]. It stands out that practically all such 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 a satisfactory accuracy in all situations [6–14]. Additionally, the computational effort involved still makes a repeated evaluation during an extensive solute-solvent simulation unfeasible.

In this paper, we present a new approach, the BGY3dM model, derived directly from the YBG-hierarchy which employs the Kirkwood superposition approximation [15]. We will investigate its properties and benefits for the computation of solvent density distributions around a solute of arbitrary shape. With our new model we are able to treat molecular solvents that interact with the solute by any short-range potential and the long-range Coulomb potential. The solvent molecules are modeled as rigid bodies. A comparison between the BGY3dM method and results obtained by extensive MD simulations demonstrate the quite good performance of the new method.

The remainder of this article is organized as follows: First, we review some related statistical mechanics concepts in Section II before we derive the BGY3dM model in Section III. Some computational details are presented in Section IV. Then, results from the BGY3dM model are compared with results from MD simulations for a HCl-like model solvent and the method is applied to carbon disulfide as solvent in Section V. Finally, we give a short summary and outlook of the presented method in Section VI.

II. STATISTICAL MECHANICS

In a typical solute-solvent system, such as a protein in water, macroscopic properties of the solute are at the center of interest. These properties are usually computed by the approximation of ensemble averages using Monte Carlo (MC) or molecular dynamics (MD) methods which sample the phase space. However, such an explicit simulation of the solvent molecules increases the size of the phase space dramatically which in turn leads to unacceptable long running times of the sampling procedure. Hence, simulations with explicit solvent are often unfeasible with today's computers.

If the microscopic quantity to be averaged does not depend explicitly on the solvent, it is possible to formally split the Hamiltonian of the system into a part only depending on the solute degrees of freedom and a part including the solute-solvent interaction. This latter part can then be integrated over all solvent degrees of freedom to yield an averaged solute potential including the solute-solvent interaction implicitly. This is illustrated in

*Electronic address: griebel@ins.uni-bonn.de
†Electronic address: jager@ins.uni-bonn.de

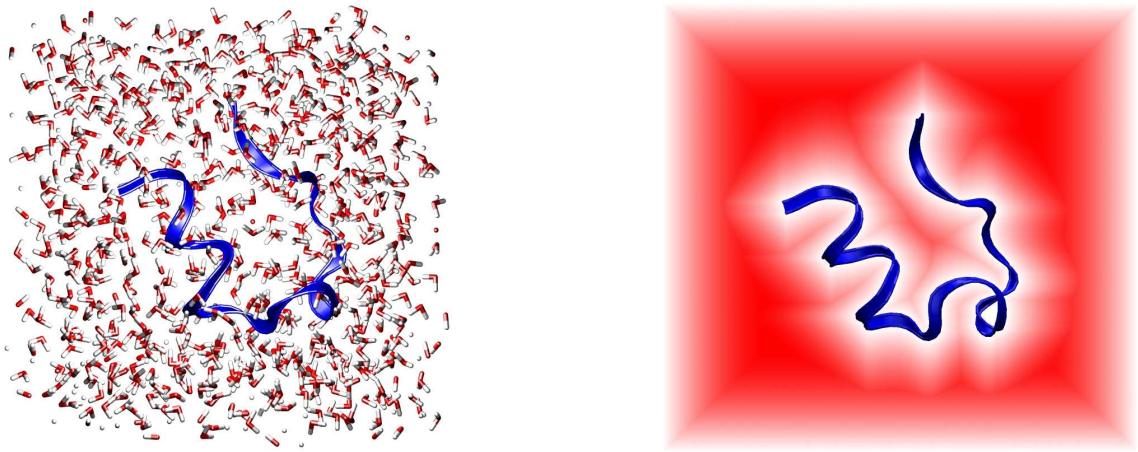

FIG. 1: Left: A protein (blue ribbon) surrounded by H_2O molecules (explicit solvent). Right: The same protein with implicit solvent indicated by the red field.

Figure 1. 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. Instead, the red field of Figure 1 (right) indicates an averaged force field that implicitly incorporates the solvent effects on the solute. Therefore, no explicit solvent molecules are necessary.

A. Potential of Mean Force

For the derivation of such an averaged potential, we 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

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

where $\mathbf{p}^M \in \mathbb{R}^{3N_M}, \mathbf{p}^S \in \mathbb{R}^{3N_S}$ are the momenta, $\mathbf{x}^M \in \Omega_{N_M}, \mathbf{x}^S \in \Omega_{N_S}$ the positions and m_i^M, m_i^S the masses of the solute and the solvent particles, respectively. The sets $\Omega_{N_M} \subset \mathbb{R}^{3N_M}$ and $\Omega_{N_S} \subset \mathbb{R}^{3N_S}$ denote the configurational domains of the solute and the solvent. The potential V describes the interaction between all particles of the system. It can be further split into a part V_M describing the intramolecular interaction of the solute, a part V_S describing the interactions between the solvent molecules and a part V_{MS} which consists of the interactions between the solute and the solvent atoms, i.e.

$$V(\mathbf{x}^M, \mathbf{x}^S) = V_M(\mathbf{x}^M) + V_S(\mathbf{x}^S) + V_{MS}(\mathbf{x}^M, \mathbf{x}^S). \quad (2)$$

Hence, the Hamiltonian can also be written in separated form

$$\begin{aligned} H(\mathbf{p}^M, \mathbf{p}^S, \mathbf{x}^M, \mathbf{x}^S) &= H_M(\mathbf{p}^M, \mathbf{x}^M) + H_S(\mathbf{p}^M, \mathbf{x}^S) \\ &+ H_{MS}(\mathbf{p}^M, \mathbf{p}^S, \mathbf{x}^M, \mathbf{x}^S). \quad (3) \end{aligned}$$

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

$$\begin{aligned} \langle a \rangle &:= C^{-1} \int_{\Omega_N} a(\mathbf{x}_M, \mathbf{x}_S) e^{-\beta V(\mathbf{x}_M, \mathbf{x}_S)} \, d\mathbf{x}_M d\mathbf{x}_S \\ \text{with} \quad C &:= \int_{\Omega_N} e^{-\beta V(\mathbf{x}_M, \mathbf{x}_S)} \, d\mathbf{x}_M d\mathbf{x}_S, \quad (4) \end{aligned}$$

where we assume that the microscopic quantity a does not depend on the momenta. The domain $\Omega_N = \Omega_{N_M} \times \Omega_{N_S}$ denotes the spatial part of the phase space and $\beta = 1/k_B T$ is the inverse temperature with k_B the Boltzmann constant. If we now further assume that a does not depend on the solvent degrees of freedom, the integral can be written as

$$\langle a \rangle = C^{-1} \int_{\Omega_{N_M}} a(\mathbf{x}_M) e^{-\beta V_M(\mathbf{x}_M)} \int_{\Omega_{N_S}} e^{-\beta(V_{MS}(\mathbf{x}_M, \mathbf{x}_S) + V_S(\mathbf{x}_S))} \, d\mathbf{x}_S d\mathbf{x}_M. \quad (5)$$

Hence, the inner integral can be computed separately. The formal integration of this inner part leads directly to the so-called potential of mean force (PMF). It is de-

fined by integrating the Boltzmann factor $e^{-\beta V(\mathbf{x})}$ over the solvent degrees of freedom,

$$\begin{aligned} e^{-\beta V^{PMF}(\mathbf{x}^M)} &= \\ & C_S^{-1} \int_{\Omega_{N_S}} e^{-\beta(V_M(\mathbf{x}^M)+V_S(\mathbf{x}^S)+V_{MS}(\mathbf{x}^M,\mathbf{x}^S))}\, d\mathbf{x}^S \\ & = e^{-\beta V_M(\mathbf{x}^M)} C_S^{-1} \int_{\Omega_{N_S}} e^{-\beta(V_S(\mathbf{x}^S)+V_{MS}(\mathbf{x}^M,\mathbf{x}^S))}\, d\mathbf{x}^S \end{aligned} \tag{6}$$

$$\text{with } C_S := \int_{\Omega_{N_S}} e^{-\beta V_S(\mathbf{x}^S)}\, d\mathbf{x}^S. \tag{7}$$

The PMF can also be written in an additive way as

$$\begin{aligned} V^{PMF}(\mathbf{x}^M) &= V_M(\mathbf{x}^M) \\ & -\frac{1}{\beta}\ln\left(C_S^{-1} \int_{\Omega_{N_S}} e^{-\beta(V_S(\mathbf{x}^S)+V_{MS}(\mathbf{x}^M,\mathbf{x}^S))}\, d\mathbf{x}^S \right) \\ & =: V_M(\mathbf{x}^M) + W(\mathbf{x}^M), \end{aligned} \tag{8}$$

where $W(\mathbf{x}^M)$ is defined by (8) and contains only the energy due to the solute-solvent interaction. It is therefore called the solvation free energy. The Hamiltonian of the reduced system now acts only on the particles of the solute M and reads as

$$H^{PMF}(\mathbf{p}^M,\mathbf{x}^M) = \frac{1}{2}\sum_{i=1}^{N_M} \frac{(p_i^M)^2}{m_i^M} + V^{PMF}(\mathbf{x}^M). \tag{9}$$

Likewise, the microscopic average (5) can be written as

$$\langle a \rangle = C_{PMF}^{-1} \int_{\Omega_{N_M}} a(\mathbf{x}_M) e^{-\beta V^{PMF}(\mathbf{x}_M)}\, d\mathbf{x}_M \tag{10}$$

$$\text{with} \qquad C_{PMF} := \int_{\Omega_{N_M}} e^{-\beta V^{PMF}(\mathbf{x}_M)}\, d\mathbf{x}_M. \tag{11}$$

A comparison of equations (5) and (10) shows that the introduction of the PMF is just a formal transformation. The ensemble averages (4) and (10) 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 (8).

In practice, the integral in (8) can be computed exactly only for very simple systems. Hence, methods have to be employed to approximately compute the PMF. In principle, MD or MC type simulations can be applied. However, due to the high dimension of the phase space they are computationally too costly. Instead, approximations have to be considered that avoid a sampling of the phase space. The most promising approaches are 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, as we will see later.

B. The YBG-Hierarchy

We introduce the concepts of the liquid state integral equation theories starting from the Liouville equation. We follow the derivation given in [16]. The Liouville equation describes the time evolution of the phase space probability distribution $\pi(\mathbf{p}_{(N)},\mathbf{x}_{(N)})$ which represents the probability to find the system at the point $(\mathbf{p}_{(N)},\mathbf{x}_{(N)})$ in phase space. To this end, $\mathbf{p}_{(N)} := (\mathbf{p}_1,\dots,\mathbf{p}_N) \in \mathbb{R}^{3N}$ are again the momenta of the N particles and $\mathbf{x}_{(N)} := (\mathbf{x}_1,\dots,\mathbf{x}_N) \in \Omega_N$ are their positions in the $3N$-dimensional domain Ω_N. We restrict the potential function to be a sum of pairwise terms

$$V(\mathbf{x}_1,\dots,\mathbf{x}_N) = \sum_{i=1}^{N}\sum_{j=i+1}^{N} v(\mathbf{x}_i,\mathbf{x}_j) \tag{12}$$

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

$$\mathbf{F}_{ij} = \mathbf{F}(\mathbf{x}_i,\mathbf{x}_j) = -\nabla_{\mathbf{x}_i} v(\mathbf{x}_i,\mathbf{x}_j). \tag{13}$$

Then, we can write the Liouville equation as

$$\frac{\partial \pi}{\partial t} = -\sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i}\cdot\frac{\partial \pi}{\partial \mathbf{x}_i} - \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \mathbf{F}_{ij}\cdot\frac{\partial \pi}{\partial \mathbf{p}_i}, \tag{14}$$

where we omit any external force on the particles. Now, the Liouville equation is integrated over $N-n$ positions and momenta and multiplied by the factor $\frac{N!}{(N-n)!}$. Thereby, we consider the definition of the reduced probability density for $n = 1,\dots,N-1$

$$\begin{aligned} \pi^{(n)}(\mathbf{p}_{(n)},\mathbf{x}_{(n)}) &:= \frac{N!}{(N-n)!} \times \\ & \times \int_{\mathbb{R}^{3(N-n)}} \int_{\Omega_{N-n}} \pi(\mathbf{p}_{(N)},\mathbf{x}_{(N)})\, d\mathbf{p}_{(N-n)} d\mathbf{x}_{(N-n)}. \end{aligned} \tag{15}$$

where $d\mathbf{p}_{(N-n)}$ and $d\mathbf{x}_{(N-n)}$ denote $d\mathbf{p}_{n+1}\cdots d\mathbf{p}_N$ and $d\mathbf{x}_{n+1}\cdots d\mathbf{x}_N$, respectively. Taking the symmetry of π under exchange of particles into account we find that

$$\begin{aligned} & \left(\frac{\partial}{\partial t} + \sum_{i=1}^{n} \left[\frac{\mathbf{p}_i}{m_i}\cdot\frac{\partial}{\partial \mathbf{x}_i} + \sum_{j=1,j\neq i}^{n} \mathbf{F}_{ij}\cdot\frac{\partial}{\partial \mathbf{p}_i} \right] \right) \pi^{(n)} \\ &= -\sum_{i=1}^{n} \int_{\mathbb{R}^3}\int_{\Omega} \mathbf{F}_{in+1}\cdot\frac{\partial \pi^{(n+1)}}{\partial \mathbf{p}_i}\, d\mathbf{x}_{n+1} d\mathbf{p}_{n+1} \end{aligned} \tag{16}$$

for $n = 1,\dots,N-1$, where $\Omega := \Omega_1$ is the domain of the system. This set of equations is called the BBGKY-hierarchy after Bogolyubov, Born, Green, Kirkwood and Yvon, see [16].

In the canonical ensemble the (reduced) probability densities can be factorized as

$$\pi^{(n)}(\mathbf{p}_{(n)},\mathbf{x}_{(n)}) = \mathcal{P}^{(n)}(\mathbf{p}_{(n)})\rho^{(n)}(\mathbf{x}_{(n)}) \tag{17}$$

with

$$\mathcal{P}^{(n)}(\mathbf{p}_{(n)}) = \prod_{i=1}^{n} \left(\frac{\beta}{2\pi m_i}\right)^{\frac{d}{2}} e^{-\beta\frac{|\mathbf{p}_i|^2}{2m_i}} \tag{18}$$

only depending on the momenta. If one further notes that

$$\frac{\partial}{\partial \mathbf{p}_i}\mathcal{P}^{(n)}(\mathbf{p}_{(n)}) = -\frac{\beta}{m_i}\mathbf{p}_i\mathcal{P}^{(n)}(\mathbf{p}_{(n)}) \tag{19}$$

and

$$\int_{\mathbb{R}^3} \mathcal{P}^{(n+1)}(\mathbf{p}_{(n+1)})\; d\mathbf{p}_{n+1} = \mathcal{P}^{(n)}(\mathbf{p}_{(n)}), \tag{20}$$

this yields for $n = 1, \ldots, N-1$

$$\begin{aligned}\sum_{i=1}^{n} \mathbf{p}_i \cdot \left(\frac{\partial}{\partial \mathbf{x}_i} - \beta \sum_{j=1, j\neq i}^{n} \mathbf{F}_{ij}\right) \rho^{(n)}(\mathbf{x}_{(n)}) \\ = \beta \sum_{i=1}^{n} \mathbf{p}_i \cdot \int_\Omega \mathbf{F}_{in+1}\rho^{(n+1)}(\mathbf{x}_{(n+1)})\; d\mathbf{x}_{n+1},\end{aligned} \tag{21}$$

where we considered equation (16) at equilibrium, i.e. $\frac{\partial}{\partial t}\pi^{(n)} = 0$. This relation must be independent of the choice of the momenta $\mathbf{p}_i$. Hence, it must hold term by term, which leads us for $n = 1, \ldots, N-1$ to the YBG-hierarchy (Yvon, Born, Green) [16],

$$\begin{aligned}\nabla_{\mathbf{x}_1} g^{(n)}(\mathbf{x}_{(n)}) \;&=\; \beta \sum_{i=2}^{n} \mathbf{F}_{1i} g^{(n)}(\mathbf{x}_{(n)}) \\ &\quad + \beta\bar{\rho} \int_\Omega \mathbf{F}_{1n+1} g^{(n+1)}(\mathbf{x}_{(n+1)})\; d\mathbf{x}_{n+1}.\end{aligned} \tag{22}$$

Here, we introduced the (reduced) distribution functions $g^{(n)} := \frac{1}{\bar{\rho}^n}\rho^{(n)}$ with $\bar{\rho}$ the overall density of the fluid.

The YBG- and the BBGKY-hierarchy are not immediately useful, since they relate for $n = 1, \ldots, N-1$ the unknown functions $\rho^{(n)}$ and $g^{(n)}$ to $\rho^{(n+1)}$ and $g^{(n+1)}$, respectively. In order to solve (22) for $g^{(n)}$ the solution $g^{(n+1)}$ for $n+1$ is needed, and so on. This ultimately requires the knowledge of $g^{(N)}$ and thereby the solution of the overall Liouville equation which is too costly to compute. To this end, for a truncation of the YBG-hierarchy a closure relation between $g^{(n+1)}$ and $g^{(n)}$ is required. Here, the case $n = 2$ is the best investigated one, see [16]. The application of the Kirkwood superposition approximation [15]

$$g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}(\mathbf{x}_2, \mathbf{x}_3), \tag{23}$$

for $n = 2$ yields the Born-Green equation

$$\begin{aligned}&\nabla_{\mathbf{x}_1}\left(\ln(g^{(2)}(\mathbf{x}_1, \mathbf{x}_2)) + \beta v(\mathbf{x}_1, \mathbf{x}_2)\right) \\ &\qquad = \beta\bar{\rho} \int_\Omega \mathbf{F}_{13}\; g^{(2)}(\mathbf{x}_1, \mathbf{x}_3)\left(g^{(2)}(\mathbf{x}_2, \mathbf{x}_3) - 1\right)\; d\mathbf{x}_3.\end{aligned} \tag{24}$$

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 $\bar{\rho}$ the results are in good agreement with those obtained by MC or MD methods or analytical results in the case of a hard sphere fluid [16]. Finding better closures for (22) is a very challenging task. According to Meeron [17] and Salpeter [18] the triplet correlation function can be expressed formally exact as

$$\begin{aligned}g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \;&=\; g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}(\mathbf{x}_2, \mathbf{x}_3) \times \\ &\quad \times \exp(\tau(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3), \bar{\rho})\end{aligned} \tag{25}$$

with $\tau(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \bar{\rho}) = \sum_{n=1}^{\infty} \bar{\rho}^n \delta_{n+3}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$. 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 [16] or [17, 18] for details. The coefficients δ_4 and δ_5 were computed for a Lennard-Jones fluid in [19, 20], 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 [21] for the quadruplet distribution function in the case $n = 3$:

$$g^{(4)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) \approx \frac{g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4) g^{(3)}(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4) g^{(3)}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)}{g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) g^{(2)}(\mathbf{x}_1, \mathbf{x}_3) g^{(2)}(\mathbf{x}_1, \mathbf{x}_4) g^{(2)}(\mathbf{x}_2, \mathbf{x}_3) g^{(2)}(\mathbf{x}_2, \mathbf{x}_4) g^{(2)}(\mathbf{x}_3, \mathbf{x}_4)}. \tag{26}$$

Applied to (22) for $n = 3$ this gives a relation for the triplet distribution function $g^{(3)}$, called BGY2 equation [22, 23]. Lee et al. [22, 23] computed $g^{(3)}$ for a hard sphere fluid. The results were significantly better than those obtained with the Kirkwood approximation for $n = 2$. They showed that the BGY2 theory is superior to the closure (25) truncated after δ_5. But to our knowledge the BGY2 theory has never been applied to other potential functions than the hard sphere potential.

C. Approximation of the PMF

The integral equation theories have been developed to compute macroscopic properties of fluids without explicitly performing the integration over the full phase space.

Similarly, they can be applied to approximate the PMF, which also can be represented by means of a reduced probability distribution. To see this, we consider the computation of the forces $-\nabla V^{PMF}$ of the 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,

$$\nabla_{\mathbf{x}^M} V^{PMF}(\mathbf{x}^M) = \langle \nabla_{\mathbf{x}^M} V(\mathbf{x}^M, \mathbf{x}^S) \rangle_{(\mathbf{x}^M)} \tag{27}$$

with

$$\langle a(\mathbf{x}^M, \mathbf{x}^S) \rangle_{(\mathbf{x}^M)} := C_{MS}^{-1} \int_{\Omega_{N_S}} a(\mathbf{x}^M, \mathbf{x}^S) e^{-\beta V(\mathbf{x}^M, \mathbf{x}^S)} \, d\mathbf{x}^S \tag{28}$$

and

$$C_{MS} := \int_{\Omega_{N_S}} e^{-\beta V(\mathbf{x}^M, \mathbf{x}^S)} \, d\mathbf{x}^S. \tag{29}$$

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

$$V_{MS}(\mathbf{x}^M, \mathbf{x}^S) = \sum_{i=1}^{N_M} \sum_{j=1}^{N_S} v_{MS}(\mathbf{x}_i^M, \mathbf{x}_j^S), \tag{30}$$

we can transform (27) and obtain

$$\begin{aligned} \nabla_{\mathbf{x}^M} V^{PMF}(\mathbf{x}^M) \; &= \; \nabla_{\mathbf{x}^M} V_M(\mathbf{x}^M) + \nabla_{\mathbf{x}^M} W(\mathbf{x}^M) \\ &= \nabla_{\mathbf{x}^M} V_M(\mathbf{x}^M) + \sum_{i=1}^{N_M} \sum_{j=1}^{N_S} \langle \nabla_{\mathbf{x}_M} v_{MS}(\mathbf{x}_i^M, \mathbf{x}_j^S) \rangle_{(\mathbf{x}_M)} \\ &= \nabla_{\mathbf{x}^M} V_M(\mathbf{x}^M) \\ &\quad + \sum_{i=1}^{N_M} \int_{\Omega} \nabla_{\mathbf{x}_M} v_{MS}(\mathbf{x}_i^M, \mathbf{r}) \rho^{(N_M+1)}(\mathbf{r}|\mathbf{x}^M) \, d\mathbf{r} \end{aligned} \tag{31}$$

with the conditional probability

$$\rho^{(N_M+1)}(\mathbf{r}|\mathbf{x}^M) = \frac{\rho^{(N_M+1)}(\mathbf{r}, \mathbf{x}^M)}{\rho^{(N_M)}(\mathbf{x}^M)} \tag{32}$$

and $\Omega \subset \mathbb{R}^3$ the domain of the system. In this derivation, we used the definition of reduced probability functions (15) for the integral (28) as well as the normalization C_{MS}, which is actually a function of the solute coordinates. This conditional probability can be identified with the average solvent density

$$\left\langle \sum_{j=1}^{N_S} \delta(\mathbf{r} - \mathbf{x}_j^S) \right\rangle_{(\mathbf{x}^M)} = \rho^{(N_M+1)}(\mathbf{r}|\mathbf{x}^M), \tag{33}$$

where $\mathbf{x}_i^S$ are the individual solvent particles.

Hence, we can compute the forces of the PMF by an integral over the three-dimensional domain if we know the average solvent density around the solute with configuration $\mathbf{x}^M$. The goal is therefore the efficient approximation of the solvent density. This can be achieved by applying the liquid state integral equation theories. To this end, concerning this application to solute-solvent systems, the literature is largely focused on methods based on the Ornstein-Zernike equation. The Ornstein-Zernike equation is an integral equation defining the direct correlation function. Similar to the equations from the YBG-hierarchy it can only be solved by assuming additional approximations, as e.g. the hypernetted-chain (HNC) or the Percus-Yevick (PY) closures. Its popularity in the literature may be caused by the fact that it can more easily be reduced to a one-dimensional equation in the case of rotational symmetry and that it has an algebraic form in Fourier space. Hence, its numerical solution is less costly at least in the case of rotational symmetry. Ikeguchi and Doi [4] and Beglov and Roux [5] have employed the Ornstein-Zernike equation together with the HNC and PY closure for the computation of the density of a simple monoatomic solvent around solutes of arbitrary shape. Kovalenko, Hirata et al. [6–12] and Beglov, Roux et al. [13, 14] have extended the methods 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 [13] 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 needed to improve the agreement between theory and experimental data.

To our knowledge, methods based directly on the YBG-hierarchy have never been considered for the computation of solvent densities in solute-solvent systems. Only methods that are related to the computation of pair distribution functions of pure molecular fluids were developed in the field of polymeric fluids. To this end, Eu and Gan [24], Taylor and Lipson [25] and Attard [26] have derived equations based on the YBG-hierarchy that have been quite successfully applied to several polymer models [27–35]. 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.

III. THE BGY3DM MODEL

In the following, we present our BGY3dM model for the approximation of solvent densities around an arbitrary solute. To this end, we first derive the site-site BGY3dM equations for the computation of pair distribution functions of the pure solvent, since these pair distribution functions are needed as input for the BGY3dM equations, which will be described afterwards. Both the SS-BGY3dM and the BGY3dM model represent equations from two distinct YBG-hierarchies, i.e. that for pure

molecular fluids and that for a solute immersed into a molecular fluid, respectively. These hierarchies are obtained by integrating the corresponding Liouville equations similar to the monoatomic case as described in Section II B. The application of approximations for the distribution functions then yields a closed set of equations, which can be solved numerically.

A. The Site-Site BGY3dM Equations

First, we derive the BGY equation, i.e. we consider the case $n = 2$, for a molecular fluid. For this, we assume that the molecules of the fluid consist of s not necessarily different particle species. From the literature, the BGY equation for a simple mixture of s different particle species is known [36]. It reads as

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \;&=\; \beta \mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \\ &+\beta \sum_{\eta=1}^{s} \bar{\rho}_\eta \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_3^\eta)\, d\mathbf{x}_3^\eta \quad (34)\end{aligned}$$

for any pair $\alpha, \gamma = 1, \ldots, s$. Here, $g^{(2)}_{\alpha\gamma}$ and $g^{(3)}_{\alpha\gamma\eta}$ are the pair and triplet distribution functions for particles of species α, γ and η, respectively, $\mathbf{F}_{\alpha\gamma} = -\nabla_{\mathbf{x}_1^\alpha} v_{\alpha\gamma}$ denotes the force between particles of species α and γ and $\bar{\rho}_\alpha$ is, as already stated above, the number density of particle species α. For clarity the position vectors have a superscript also indicating their respective particle species, e.g. $\mathbf{x}_1^\alpha \in \mathbb{R}^3$.

We now consider molecular solvents. Therefore, we set $\bar{\rho}_S := \bar{\rho}_1 = \bar{\rho}_2 = \ldots = \bar{\rho}_s$. Then, we build molecules which contain exactly one particle from every not necessarily different species. By this, we model a molecular fluid with number density $\bar{\rho}_S$, where the molecules consist of s particles. When we now derive the BGY equation for this molecular fluid, we have to distinguish between intramolecular and intermolecular interactions. The intramolecular forces are indicated by the superscript i, i.e., we write $\mathbf{F}^i_{\alpha\gamma}$ and $v^i_{\alpha\gamma}$ for the forces and the pair potential, respectively. Similar to the forces, we now have different types of distribution functions, which depend on how many of the particles belong to the same molecule. We indicate this dependency by the indices of the corresponding position vectors, i.e. $\mathbf{x}_j^\alpha$ denotes the position of particle α of molecule j. Hence, $g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma)$ denotes the intramolecular pair distribution function between particle α and γ of the same molecule, whereas $g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma)$ denotes the intermolecular pair distribution function between particles α and γ of different molecules. In this notation we can write the BGY equation for the intermolecular pair distribution functions ($n = 2$) of a molecular fluid as

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \;&=\; \beta \mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \quad (35)\\ &+\beta \sum_{\eta=1}^{s} \bar{\rho}_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_3^\eta)\, d\mathbf{x}_3^\eta \\ &+\beta \sum_{\eta=1,\eta\neq\alpha}^{s} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta \\ &+\beta \sum_{\eta=1,\eta\neq\gamma}^{s} \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta,\end{aligned}$$

and the BGY equation for the intramolecular pair distribution functions as

$$\begin{aligned}\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) \;&=\; \beta \mathbf{F}^i_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) \quad (36)\\ &+\beta \sum_{\eta=1}^{s} \bar{\rho}_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma, \mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta \\ &+\beta \sum_{\eta=1,\eta\neq\alpha,\eta\neq\gamma}^{s} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta.\end{aligned}$$

As for a simple mixture, the equations are obtained by integrating the Liouville equation over $N-2$ particles, i.e., we choose the case $n = 2$ of the YBG-hierarchy for molecular fluids. To this end, one has to take into account that particles can now belong to different molecules. This additional distinguishability results in the last two lines of equation (35) and the last line of equation (36) which represent the intramolecular coupling within the molecules. The first lines are identical to those of a simple mixture as in equation (34). Recall that all particle species in the molecular fluid have the same number density $\bar{\rho}_S$.

In order to facilitate the numerical solution of the final equations, the molecules are modeled as rigid bodies. To this end, we introduce a harmonic potential as intramolecular interaction

$$v^i(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa) = \kappa (r_1^{\alpha\gamma} - r_0^{\alpha\gamma})^2, \qquad \forall\, \alpha \neq \gamma \quad (37)$$

with $r_1^{\alpha\gamma} = |\mathbf{x}_1^\alpha - \mathbf{x}_1^\gamma|$. Here, $r_0^{\alpha\gamma}$ denotes 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 (37) does not yet lead to fixed distances within the molecule, but allows fluctuations around the desired distances $r_0^{\alpha\gamma}$. Hence, we investigate the limit case where the constant κ goes to infinity, i.e., we consider $\lim_{\kappa\to\infty} v^i(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa)$. Here, κ determines the strength of the force that constrains two particles to their desired distance. We examine equation (36), which determines the intramolecular pair distribution functions, and assume that, in this limit, the solution of equation (36) is strongly dominated by the first term of the right hand side, and that all integral terms can be neglected, i.e., for $\kappa \to \infty$ we have

$$\nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa) = \beta \mathbf{F}^i_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa). \quad (38)$$

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 limit version of $g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma)$ which will not depend on this parameter. The solution of (38) is

$$\begin{aligned} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa) &= \sqrt{\frac{4\beta\kappa}{\pi}} \frac{1}{4\pi (r_0^{\alpha\gamma})^2} e^{-\beta v^i(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma; \kappa)} \quad (39) \\ &= \sqrt{\frac{4\beta\kappa}{\pi}} \frac{1}{4\pi (r_0^{\alpha\gamma})^2} e^{-\beta\kappa (r_1^{\alpha\gamma} - r_0^{\alpha\gamma})^2}. \end{aligned}$$

The factor $\sqrt{\frac{4\beta\kappa}{\pi}} \frac{1}{4\pi (r_0^{\alpha\gamma})^2}$ is chosen such that the intramolecular pair distribution function obeys the correct normalization condition

$$\lim_{\kappa\to\infty} \int_\Omega g^{(2)}_{\alpha\gamma}(\mathbf{r}_1^{\alpha\gamma}; \kappa)\, d\mathbf{r}_1^{\alpha\gamma} = 1 \qquad (40)$$

with $\mathbf{r}_1^{\alpha\gamma} = \mathbf{x}_1^\alpha - \mathbf{x}_1^\gamma$. With this choice, we find for the convolution with an arbitrary function f

$$\begin{aligned} &\int_\Omega f(\mathbf{r}_1^{\alpha\gamma} - \mathbf{r}') g^{(2)}_{\alpha\gamma}(\mathbf{r}_1^{\alpha\gamma})\, d\mathbf{r}_1^{\alpha\gamma} \\ &= \lim_{\kappa\to\infty} \int_\Omega f(\mathbf{r}_1^{\alpha\gamma} - \mathbf{r}') g^{(2)}_{\alpha\gamma}(\mathbf{r}_1^{\alpha\gamma}; \kappa)\, d\mathbf{r}_1^{\alpha\gamma} \\ &= \lim_{\kappa\to\infty} \int_\Omega f(\mathbf{r}_1^{\alpha\gamma} - \mathbf{r}') \sqrt{\frac{4\beta\kappa}{\pi}} \frac{e^{-\beta\kappa (r_1^{\alpha\gamma} - r_0^{\alpha\gamma})^2}}{4\pi (r_0^{\alpha\gamma})^2}\, d\mathbf{r}_1^{\alpha\gamma} \\ &= \int_\Omega f(\mathbf{r}_1^{\alpha\gamma} - \mathbf{r}') \frac{\delta(r_1^{\alpha\gamma} - r_0^{\alpha\gamma})}{4\pi (r_0^{\alpha\gamma})^2}\, d\mathbf{r}_1^{\alpha\gamma}. \qquad (41) \end{aligned}$$

Equation (41) represents the definition of the delta distribution as the limit of a Dirac sequence. The result is 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. Consequently, we know all *intra*molecular pair distribution functions and set

$$g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\gamma) = \frac{\delta(r_1^{\alpha\gamma} - r_0^{\alpha\gamma})}{4\pi (r_0^{\alpha\gamma})^2}, \qquad \forall\ \alpha \neq \gamma. \qquad (42)$$

Now, to solve equation (35) for the *inter*molecular pair distribution functions, a closure relation is needed. For this, we approximate the triplet distribution functions by products of functions that depend only on two of the three particle coordinates. In the case where all three particles belong to different molecules we insert the famous Kirkwood approximation [15]

$$g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_3^\eta) \approx g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_3^\eta). \qquad (43)$$

In the case where two of the particles belong to the same molecule, the Kirkwood approximation is not satisfactory [26]. Hence, we employ the so-called normalized site-site superposition approximations (NSSA) of Taylor and Lipson [25]

$$g^{(3)}_{\alpha\gamma\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma, \mathbf{x}_2^\eta) \approx g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_2^\eta) \tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) \qquad (44)$$

with

$$\begin{aligned} \tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) &= \frac{g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma)}{n^\eta_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma)}, \\ \tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) &= \frac{g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta)}{n^\gamma_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta)} \qquad (45) \end{aligned}$$

and the normalization functions

$$\begin{aligned} n^\eta_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) &= \int_\Omega g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_2^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta, \\ n^\gamma_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) &= \int_\Omega g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_2^\eta) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) d\mathbf{x}_2^\gamma. \qquad (46) \end{aligned}$$

Applying the Kirkwood approximation and the NSSA approximation for the intermolecular and intramolecular terms of equation (35) finally gives

$$\begin{aligned} \nabla_{\mathbf{x}_1^\alpha} g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) &= \beta \mathbf{F}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \\ &\quad + \beta g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1}^{s} \bar{\rho}_S \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_3^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_3^\eta)\, d\mathbf{x}_3^\eta \\ &\quad + \beta \tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\alpha}^{s} \int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_1^\eta) \tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma, \mathbf{x}_1^\eta)\, d\mathbf{x}_1^\eta \\ &\quad + \beta \tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\gamma) \sum_{\eta=1,\eta\neq\gamma}^{s} \int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) \tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}_1^\alpha, \mathbf{x}_2^\eta) g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma, \mathbf{x}_2^\eta)\, d\mathbf{x}_2^\eta \qquad (47) \end{aligned}$$

In the last line of (47) we can now substitute the intramolecular pair distribution function according to (42). The term which includes the intramolecular force $\mathbf{F}^i_{\alpha\eta}$, however, requires a more detailed examination of the limit $\kappa \to \infty$.

If we insert the definition of the intramolecular potential (37) and the intramolecular pair distribution from (39), we have

$$\begin{aligned}
&\lim_{\kappa\to\infty}\sum_{\eta=1,\eta\neq\alpha}^{s}\beta\tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)\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)\tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma,\mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta\\
=&\lim_{\kappa\to\infty}\sum_{\eta=1,\eta\neq\alpha}^{s}\beta\tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)\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}\tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma,\mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta\\
=&\lim_{\kappa\to\infty}\sum_{\eta=1,\eta\neq\alpha}^{s}\tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)\sum_{\eta=1,\eta\neq\alpha}^{s}\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)\tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma,\mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta\\
=&\lim_{\kappa\to\infty}\sum_{\eta=1,\eta\neq\alpha}^{s}\tilde{g}^{(2)}_{\alpha\gamma;\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)\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}\tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma,\mathbf{x}_1^\eta)\right)\ d\mathbf{x}_1^\eta\\
=&\sum_{\eta=1,\eta\neq\alpha}^{s}\frac{g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)}{n^\eta_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)}\nabla_{\mathbf{x}_1^\alpha}\int_\Omega\frac{\delta(r_1^{\alpha\eta}-r_0^{\alpha\eta})}{4\pi(r_0^{\alpha\eta})^2}\tilde{g}^{(2)}_{\gamma\eta;\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}\nabla_{\mathbf{x}_1^\alpha}\ln\left(\int_\Omega\frac{\delta(r_1^{\alpha\eta}-r_0^{\alpha\eta})}{4\pi(r_0^{\alpha\eta})^2}\tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma,\mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta\right).
\end{aligned}\tag{48}$$

In the last line of equation (48) we approximated $g^{(2)}_{\gamma\eta}$ in the definition of $n^\eta_{\alpha\gamma}$ by its normalized form $\tilde{g}^{(2)}_{\gamma\eta;\alpha}$, such that we can write the fraction as the gradient of a logarithm.

A further simplification of equation (47) is possible by introducing a product ansatz for the solution

$$g^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)=g^0_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)e^{-u^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)}\tag{49}$$

with

$$g^0_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)=e^{-\beta v_{\alpha\gamma}(\mathbf{x}_1^\alpha-\mathbf{x}_2^\gamma)},\tag{50}$$

where $v_{\alpha\gamma}(\mathbf{x}_1^\alpha-\mathbf{x}_2^\gamma)$ denotes the potential between sites α and γ of molecules one and two, respectively. Here, $u^{(2)}_{\alpha\gamma}$ is the new unknown function. This approach together with the application of the divergence to both sides of the equation leads after some calculation to the following equation for the intermolecular site-site pair distribution function:

$$\begin{aligned}
\Delta_{\mathbf{x}_1^\alpha}u^{(2)}_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)\ =\ &\beta\sum_{\eta=1}^{s}\bar{\rho}_S\nabla_{\mathbf{x}_1^\alpha}\cdot\int_\Omega\mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_3^\eta)g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_3^\eta)g^{(2)}_{\gamma\eta}(\mathbf{x}_2^\gamma,\mathbf{x}_3^\eta)\ d\mathbf{x}_3^\eta\\
&-\sum_{\eta=1,\eta\neq\alpha}^{s}\Delta_{\mathbf{x}_1^\alpha}\ln\left(\int_\Omega\omega_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta)\tilde{g}^{(2)}_{\gamma\eta;\alpha}(\mathbf{x}_2^\gamma,\mathbf{x}_1^\eta)\ d\mathbf{x}_1^\eta\right)\\
&-\beta\sum_{\eta=1,\eta\neq\gamma}^{s}\nabla_{\mathbf{x}_1^\alpha}\cdot\frac{\int_\Omega\mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\eta)\tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\eta)\omega_{\gamma\eta}(\mathbf{x}_2^\gamma,\mathbf{x}_2^\eta)\ d\mathbf{x}_2^\eta}{n^\eta_{\alpha\gamma}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\gamma)}.
\end{aligned}\tag{51}$$

Here, we introduced the notation

$$\omega_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta):=g^{(2)}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta)=\frac{\delta(r_1^{\alpha\eta}-r_0^{\alpha\eta})}{4\pi(r_0^{\alpha\eta})^2}\tag{52}$$

for the intramolecular pair distribution functions. Note that we switch here to the ω-notation to agree with the literature. The dot $\cdot$ in the right hand side of (51) indicates the scalar product of two vectors in $\mathbb{R}^3$. We call equation (51) for any pair $\alpha,\gamma=1,\dots,s$ with $\alpha\neq\gamma$ the site-site BGY3dM (SS-BGY3d) equations.

B. The BGY3dM Equations

Next, we consider a single solute molecule immersed into the solvent. We wish to compute the mean density of the solvent around a fixed configuration of the solute. If we assume that the solute is described by N_M particles with fixed configuration $\mathbf{x}^M:=(\mathbf{x}_1^M,\dots,\mathbf{x}_{N_M}^M)$ the mean solvent density ρ^S_α of particle type α can be expressed by

a conditional N_M+1-particle distribution

$$\begin{aligned}\rho_\alpha^S(\mathbf{x}_1^\alpha) &:= \bar{\rho}_S g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M)\\ &= \bar{\rho}_S \frac{g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha,\mathbf{x}^M)}{g^{(N_M)}(\mathbf{x}^M)}.\end{aligned} \tag{53}$$

In analogy to the BGY equation for the site-site pair distribution functions of a molecular fluid we can derive the BGY equation for the N_M+1-particle distribution function where N_M particles belong to the solute:

$$\begin{aligned}&\nabla_{\mathbf{x}_1^\alpha} g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha,\mathbf{x}^M) \;=\\ &\quad \beta \mathbf{F}_\alpha(\mathbf{x}_1^\alpha,\mathbf{x}^M) g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha,\mathbf{x}^M)\\ &\quad +\beta\sum_{\eta=1}^{s}\bar{\rho}_S\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\eta) g_{\alpha\eta}^{(N_M+2)}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\eta,\mathbf{x}^M)\; d\mathbf{x}_2^\eta\\ &\quad +\beta\sum_{\eta=1,\eta\neq\alpha}^{s}\int_\Omega \mathbf{F}^i_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta) g_{\alpha\eta}^{(N_M+2)}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta,\mathbf{x}^M)\; d\mathbf{x}_1^\eta.\end{aligned} \tag{54}$$

To this end, $\mathbf{F}_\alpha(\mathbf{x}_1^\alpha,\mathbf{x}^M)$ is the total force exerted on the solvent particle $\mathbf{x}_1^\alpha$ due to the solute. Next, we employ the so-called n-level Kirkwood closure relations [37] and approximate the intermolecular and intramolecular N_M+2-particle distribution functions by

$$\begin{aligned}&g_{\alpha\eta}^{(N_M+2)}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\eta,\mathbf{x}^M) \approx\\ &\qquad \frac{g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha,\mathbf{x}^M) g_{\alpha\eta}^{(2)}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\eta) g_\eta^{(N_M+1)}(\mathbf{x}_2^\eta,\mathbf{x}^M)}{g^{(N_M)}(\mathbf{x}^M)},\\ &g_{\alpha\eta}^{(N_M+2)}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta,\mathbf{x}^M) \approx\\ &\qquad \frac{\tilde{g}_{\alpha;\eta}^{(N_M+1)}(\mathbf{x}_1^\alpha,\mathbf{x}^M)\omega_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta)\tilde{g}_{\eta;\alpha}^{(N_M+1)}(\mathbf{x}_1^\eta,\mathbf{x}^M)}{g^{(N_M)}(\mathbf{x}^M)},\end{aligned} \tag{55}$$

with

$$\begin{aligned}\tilde{g}_{\alpha;\eta}^{(N_M+1)}(\mathbf{x}_1^\alpha,\mathbf{x}^M) &:= \frac{g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha,\mathbf{x}^M)}{n_\alpha^\eta(\mathbf{x}_1^\alpha,\mathbf{x}^M)}\\ \tilde{g}_{\eta;\alpha}^{(N_M+1)}(\mathbf{x}_1^\eta,\mathbf{x}^M) &:= \frac{g_\eta^{(N_M+1)}(\mathbf{x}_1^\eta,\mathbf{x}^M)}{n_\eta^\alpha(\mathbf{x}_1^\eta,\mathbf{x}^M)}\end{aligned} \tag{56}$$

and

$$\begin{aligned}n_\alpha^\eta(\mathbf{x}_1^\alpha) &:= \int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta)\tilde{g}_{\eta;\alpha}^{(N_M+1)}(\mathbf{x}_1^\eta|\mathbf{x}^M)\; d\mathbf{x}_1^\eta,\\ n_\eta^\alpha(\mathbf{x}_1^\eta) &:= \int_\Omega \omega_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta) g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) d\mathbf{x}_1^\alpha.\end{aligned} \tag{57}$$

For the approximation of the intramolecular distribution function we additionally simplified the NSSA approximation (55) of Taylor and Lipson [25] by replacing $g_\eta^{(N_M+1)}$ by its normalized form $\tilde{g}_{\eta;\alpha}^{(N_M+1)}$ in (57). Now, we can insert the above approximations into equation (55) and consider the modeling of the rigid bonds of the molecules in the same way as for the SS-BGY3dM equations. Then, we divide the whole equation by $g^{(N_M)}(\mathbf{x}^M)$ and employ the product approach

$$g_\alpha(\mathbf{x}_1^\alpha) := g_\alpha^{(N_M+1)}(\mathbf{x}_1^\alpha|\mathbf{x}^M) = g_\alpha^0(\mathbf{x}_1^\alpha;\mathbf{x}^M) e^{-u_\alpha(\mathbf{x}_1^\alpha)} \tag{58}$$

with

$$g_\alpha^0(\mathbf{x}_1^\alpha;\mathbf{x}^M) = e^{-\beta v_\alpha(\mathbf{x}_1^\alpha-\mathbf{x}^M)}, \tag{59}$$

where $v_\alpha(\mathbf{x}_1^\alpha-\mathbf{x}^M)$ describes the total potential between solvent particle α and the solute. Finally, we apply the divergence and obtain the equation

$$\begin{aligned}&\Delta_{\mathbf{x}_1^\alpha} u_\alpha(\mathbf{x}_1^\alpha) \;=\\ &-\beta\sum_{\eta=1}^{s}\bar{\rho}_S\nabla_{\mathbf{x}_1^\alpha}\cdot\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\eta) g_{\alpha\eta}^{(2)}(\mathbf{x}_1^\alpha,\mathbf{x}_2^\eta) g_\eta(\mathbf{x}_2^\eta)\; d\mathbf{x}_2^\eta\\ &-\sum_{\eta=1,\eta\neq\alpha}^{s}\Delta_{\mathbf{x}_1^\alpha}\ln\left(\int_\Omega \tilde{g}_{\eta;\alpha}(\mathbf{x}_1^\eta)\omega_{\alpha\eta}(\mathbf{x}_1^\alpha,\mathbf{x}_1^\eta)\; d\mathbf{x}_1^\eta\right),\end{aligned} \tag{60}$$

for the mean solvent density

$$\rho_\alpha^S(\mathbf{x}_1^\alpha) = \bar{\rho}_S g_\alpha(\mathbf{x}_1^\alpha) = \bar{\rho}_S g_\alpha^0(\mathbf{x}_1^\alpha) e^{-u_\alpha(\mathbf{x}_1^\alpha)} \tag{61}$$

of solvent site $\alpha = 1,\ldots,s$. We coin this equations the BGY3dM equations. They require the site-site pair distribution functions $g_{\alpha\eta}^{(2)}$ as input, which can be computed by the SS-BGY3dM equations beforehand. The intramolecular distribution functions $\omega_{\alpha\eta}$ are given by relation (52).

IV. COMPUTATIONAL DETAILS

To solve the SS-BGY3dM and BGY3dM equations (51) and (61) all functions are approximated on a regular grid with mesh size $h = L/m$, where L is the length of the domain $\Omega = [0,L]^3$ in one direction and m is the number of grid points in one direction. Although the site-site pair distribution functions are radial symmetric, the SS-BGY3dM equations are solved with full three-dimensional resolution and the computed pair distribution functions are used as input for the BGY3dM equations. The nonlinear integro-differential equations (51) and (61) are solved by a standard Picard iteration with simple mixing, i.e. with an additional damping. In every step of the iteration a Poisson problem with Dirichlet boundary conditions $u_{\alpha\gamma}(\partial\Omega) = 0$ and $u_\alpha(\partial\Omega) = 0$, $\alpha,\gamma = 1,\ldots,s$ has to be solved. Since we compute the right hand sides of equations (51) and (61) by means of Fourier transformations, the solution of the Poisson problem is computed by a diagonal scaling in Fourier space. The convolution integrals are solved by the convolution theorem, i.e.

$$f * g = \mathcal{F}_3^{-1}(\mathcal{F}_3(f)\mathcal{F}_3(g)), \tag{62}$$

where $*$ denotes the convolution and the three-dimensional Fourier transformation is defined as

$$\hat{g}(\mathbf{k}) := \mathcal{F}_3(g)(\mathbf{k}) = \int_{\mathbb{R}^3} g(\mathbf{x}) e^{-2\pi\imath\mathbf{k}\cdot\mathbf{x}}\; d\mathbf{x}, \tag{63}$$

$$\mathcal{F}_3^{-1}(\hat{g})(\mathbf{x}) = \int_{\mathbb{R}^3} \hat{g}(\mathbf{k}) e^{2\pi\imath\mathbf{k}\cdot\mathbf{x}}\; d\mathbf{k}. \tag{64}$$

We employ the parallel FFT algorithm of the FFTW [38]. For simplicity, all appearing differential operators are applied in Fourier space. The Fourier transform of the intramolecular distribution functions $\omega_{\alpha\gamma}$ is analytically known

$$\mathcal{F}_3\left(\delta(r_{12}-r_0^{\alpha\gamma})\right)(\mathbf{k}) = \frac{2}{|\mathbf{k}|}\sin(2\pi|\mathbf{k}|r_0^{\alpha\gamma})r_0^{\alpha\gamma}, \qquad (65)$$

and has not to be computed. For numerical stability, the divisions by the normalization functions have to be regularized. Hence, we replace the exact divisions by

$$\frac{\int_\Omega \mathbf{F}_{\alpha\eta}(\mathbf{r}-\mathbf{r}')\tilde{g}^{(2)}_{\alpha\eta;\gamma}(\mathbf{r}-\mathbf{r}')\omega_{\gamma\eta}(\mathbf{r}')\ d\mathbf{r}'}{\max(n^\eta_{\alpha\gamma}(\mathbf{r}),\epsilon_\omega)}, \qquad \forall\ \mathbf{r}\in\Omega. \qquad (66)$$

and

$$\frac{g_\alpha(\mathbf{x})}{\max(n^\gamma_\alpha(\mathbf{x}),\epsilon_g)}, \qquad \forall\ \mathbf{x}\in\Omega. \qquad (67)$$

with the regularization parameters $\epsilon_\omega = 10^{-1}$ and $\epsilon_g = 10^{-2}$, respectively.

A. The Coulomb Potential

The application of the discrete Fourier transform requires periodicity of the involved functions with respect to the computational domain Ω. Assuming that the functions are of short range, i.e., they decay faster to zero than $1/r^3$ with the distance r in three-dimensional space, we can simply choose the domain large enough, so that all functions are sufficiently small at the boundaries. The effect of the periodicity can then be eliminated by zero padding. Hence, the convolution integrals can be computed by discrete Fourier transforms for all short-range potentials as e.g. the Lennard-Jones potential. However, if the Coulomb potential $1/|\mathbf{r}|$ is involved, which is not of short range, we cannot directly apply the discrete Fourier transform, since the force does not vanish at the boundaries and is not periodic with respect to the domain. Nevertheless, the convolution integrals can be computed by means of discrete Fourier transforms as follows: We consider the (notationally simplified) convolution integral

$$\int_\Omega \mathbf{F}(\mathbf{r}'-\mathbf{r})g_\alpha(\mathbf{r}'-\mathbf{r})g_\gamma(\mathbf{r}')\ d\mathbf{r}' \qquad (68)$$

with the total force $\mathbf{F} = \mathbf{F}^{LJ} + \mathbf{F}^C$ consisting of a Lennard-Jones and a Coulomb part. We split the Coulomb force into a short-range and a long-range part, such that the long-range part is smooth and therefore has fast decaying Fourier components. This is achieved by adding and subtracting a smooth charge distribution around the point charge at $\mathbf{r} = 0$. As in the particle-mesh-Ewald method [39] the charge distribution is chosen to be a Gaussian

$$\varrho(\mathbf{r}) = \left(\frac{G}{\sqrt{\pi}}\right)^3 e^{-G^2|\mathbf{r}|^2} \qquad (69)$$

with a parameter G that determines the width of the function. The Coulomb potential (and its force) between two particles is divided by means of this shielding function into the following parts

$$\begin{aligned} v^C(\mathbf{r}) &= v^{Cs}(\mathbf{r}) + v^{Cl}(\mathbf{r}) = q_\gamma\Phi^s(\mathbf{r}) + q_\gamma\Phi^l(\mathbf{r}),\\ \mathbf{F}^C(\mathbf{r}) &= \mathbf{F}^{Cs}(\mathbf{r}) + \mathbf{F}^{Cl}(\mathbf{r}) = -\nabla v^{Cs}(\mathbf{r}) - \nabla v^{Cl}(\mathbf{r}), \end{aligned} \qquad (70)$$

with Φ^s and Φ^l the solutions of the Poisson equations

$$\begin{aligned} -\Delta\Phi^s &= \frac{1}{\epsilon_0}q_\alpha(\delta_3-\varrho) \qquad \text{in } \mathbb{R}^3,\\ -\Delta\Phi^l &= \frac{1}{\epsilon_0}q_\alpha\varrho \qquad \text{in } \mathbb{R}^3, \end{aligned} \qquad (71)$$

where q_α and q_γ denote the charges of the two particles and ϵ_0 is the dielectric constant. For the special choice of the function ϱ as in (69), the solution can be given analytically. We have

$$\begin{aligned} v^{Cs}(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0}q_\alpha q_\gamma\frac{\operatorname{erfc}(G|\mathbf{r}|)}{|\mathbf{r}|},\\ v^{Cl}(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0}q_\alpha q_\gamma\frac{\operatorname{erf}(G|\mathbf{r}|)}{|\mathbf{r}|}, \end{aligned} \qquad (72)$$

with erf the error function and erfc $= 1 -$ erf the complementary error function. The complementary error function decays rapidly, whereas the error function decays as slowly as $\frac{1}{|\mathbf{r}|}$ and is smooth even at $|\mathbf{r}| = 0$.

We now want to use these properties and transform the integral (68) 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 split as discussed above. Hence, we transform the convolution integral (68) according to

$$\begin{aligned} &\int_\Omega \mathbf{F}(\mathbf{r}'-\mathbf{r})g_\alpha(\mathbf{r}'-\mathbf{r})g_\gamma(\mathbf{r}')\ d\mathbf{r}' = \qquad (73)\\ &\qquad \int_\Omega \left(\mathbf{F}(\mathbf{r}'-\mathbf{r})g_\alpha(\mathbf{r}'-\mathbf{r}) - \mathbf{F}^{Cl}(\mathbf{r}'-\mathbf{r})\right)g_\gamma(\mathbf{r}')\ d\mathbf{r}'\\ &\quad + \int_\Omega \mathbf{F}^{Cl}(\mathbf{r}'-\mathbf{r})g_\gamma(\mathbf{r}')\ d\mathbf{r}'. \end{aligned}$$

The first term can be treated as before, since the part in outer brackets is of short range. The second integral can be written as

$$\int_\Omega \mathbf{F}^{Cl}(\mathbf{r}'-\mathbf{r})g_\gamma(\mathbf{r}')\ d\mathbf{r}' = -\nabla_{\mathbf{r}}\int_\Omega v^{Cl}(\mathbf{r}'-\mathbf{r})g_\gamma(\mathbf{r}')\ d\mathbf{r}', \qquad (74)$$

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

$$\mathcal{F}_3(v^{Cl})(\mathbf{k}) = \frac{q_\alpha q_\gamma}{\epsilon_0}e^{-\frac{\pi^2}{G^2}|\mathbf{k}|^2}. \qquad (75)$$

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 eliminated in Fourier space by the inverse operator of the left hand side of the SS-BGY3dM or BGY3dM equations.

Now, the convolution integrals with the long-range Coulomb force can efficiently be computed. But still, the assumption of periodicity of the involved functions introduces an error at the boundaries. Since we know that the distribution functions have to be of short range [40], it follows that we have $g_\alpha^S(\partial\Omega) \approx 0$ for a domain Ω, which is large enough. Hence, we employ the boundary conditions $g_\alpha^S(\partial\Omega) = 0$ and thereby $u_\alpha(\partial\Omega) = v_\alpha^{Cl}(\partial\Omega)$. Here, we restrict our discussion to the BGY3dM equations since the exact same considerations also hold for the SS-BGY3dM equations. To enforce these boundary conditions and to remove the errors due to the periodic boundary conditions, we can compute this error as the solution of

$$\begin{aligned} \Delta u_\alpha^* &= 0 \quad \text{in } \Omega \\ \text{with} \quad u_\alpha^*(\partial\Omega) &= u_\alpha(\partial\Omega) - v_\alpha^{Cl}(\partial\Omega), \end{aligned} \tag{76}$$

where $u_\alpha(\partial\Omega)$ are the boundary values of the previously computed solution of the BGY3dM equations for particle type α. Subtraction of the solution u_α^* from u_α leads exactly to the solution of the BGY3dM equation with the desired Dirichlet boundary condition $\bar{u}_\alpha(\partial\Omega) = v_\alpha^{Cl}(\partial\Omega)$ where $\bar{u}_\alpha = u_\alpha - u_\alpha^*$.

The solution of equation (76) is computed by a simple finite-difference scheme with a seven-point stencil. The resulting systems of equations can be solved by any iterative method. For convenience we choose the GMRES method with block Jacobi preconditioning as it is implemented in PETSc [41–43]. The computational effort necessary to solve this problem is very small compared to the costs for solving the BGY3dM equations, since the repeated solution of (76) becomes very efficient at later steps of the Picard iteration. Then, subsequent iterates u_α^l, with l denoting the iteration number, differ only little at the boundaries and we have $u_\alpha^{*l} \approx u_\alpha^{*l-1}$. Therefore, only few GMRES iterations are sufficient to compute the solution u_α^{*l} up to a prescribed accuracy.

V. RESULTS

To validate the BGY3dM method for the computation of solvent densities around arbitrary solutes, we compare results computed by the BGY3dM method with results obtained from MD simulations. For this, we employ the HCl-like models of Hirata et al. [44] as solvent. They are two site models where the intermolecular interaction is described by a combination of Lennard-Jones and Coulomb forces. Hence, the total potential between two particles of species α and γ with $\alpha, \gamma = \mathrm{H}, \mathrm{Cl}$ can be written as

$$\begin{aligned} v_{\alpha\gamma}^I(r) &= v_{\alpha\gamma}^{LJ}(r) + v_{\alpha\gamma}^C(r) \\ &= 4\epsilon_{\alpha\gamma}\left(\left(\frac{\sigma_{\alpha\gamma}}{r}\right)^{12} - \left(\frac{\sigma_{\alpha\gamma}}{r}\right)^{6}\right) + \epsilon_C \frac{q_\alpha q_\gamma}{r} \end{aligned} \tag{77}$$

with $r = |\mathbf{x}^\alpha - \mathbf{x}^\gamma|$. The specific parameters for the HCl-like models can be found in Table I. The two models only differ in the value of σ_{H}, i.e., for the first model (HCl1) it is $\sigma_{\mathrm{H}}^1 = 2.735$ Å and both particles have large Lennard-Jones spheres, whereas in the second model (HCl2) we have $\sigma_{\mathrm{H}}^2 = 0.4$ Å and the Lennard-Jones sphere of the hydrogen is completely embedded inside that of the chloride atom.

$m_{\mathrm{Cl}} = 35.453$ u	$m_{\mathrm{H}} = 1.008$ u
$q_{\mathrm{Cl}} = -0.2$ 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	$\sigma_{\mathrm{H}}^1 = 2.735$ Å, $\sigma_{\mathrm{H}}^2 = 0.4$ Å

TABLE I: Parameter values for the HCl-like model solvents.

The parameters for the Lennard-Jones potential are computed according to the Lorentz-Berthelot mixing rules

$$\epsilon_{\alpha\gamma} = \sqrt{\epsilon_\alpha \epsilon_\gamma}, \quad \sigma_{\alpha\gamma} = 0.5(\sigma_\alpha + \sigma_\gamma), \quad \alpha, \gamma = \mathrm{H}, \mathrm{Cl}. \tag{78}$$

The constant intramolecular distance between a hydrogen and chloride atom of the same molecule is set to $r_0^{\mathrm{HCl}} = 1.257$ Å. The molecular number density is chosen as $\bar{\rho}_S = 0.0018$ Å^{-3} with $\Omega = [-10\text{ Å}, 10\text{ Å}]^3$ and the temperature as $T = 420$ K ($\beta = 1.1989$). With this choice of unit system we additionally have $\epsilon_C \approx 331.84$ (kcal Å)/(mol e^2). We use $m = 256$ and $m = 128$ grid points in one direction for the solution of the SS-BGY3dM and the BGY3dM equations, respectively. The iteration of the (SS-)BGY3dM equations is stopped if the L_∞-norm of the difference of subsequent iterates is less than 10^{-2}. Note that this value for the stopping criterion is sufficient, because the model error of the (SS-)BGY3dM equations is larger anyway as we will see in the following section.

A. Site-Site Pair Distribution Functions

Now, we first compute the site-site pair distribution functions by means of the SS-BGY3dM equations (51) and compare them to the pair distribution functions computed by a MD simulation of the pure solvent, see e.g. [45] for details concerning the computation of pair distribution functions by MD. The results are shown in Figures 2 and 3.

Note that the SS-BGY3dM equations are solved in three-dimensions, but Figures 2 and 3 only show the radial component of the computed site-site pair distribution functions. We compute three error values to quan-

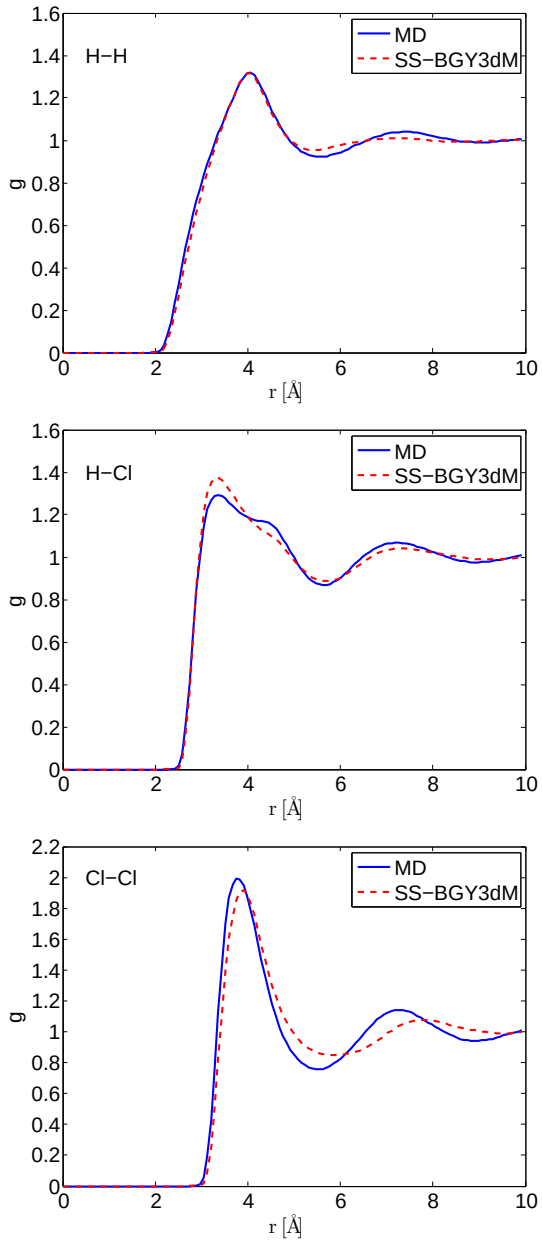

FIG. 2: Radial component of the site-site pair distribution functions for the HCl-like model (HCl1). Comparison between SS-BGY3dM and MD results.

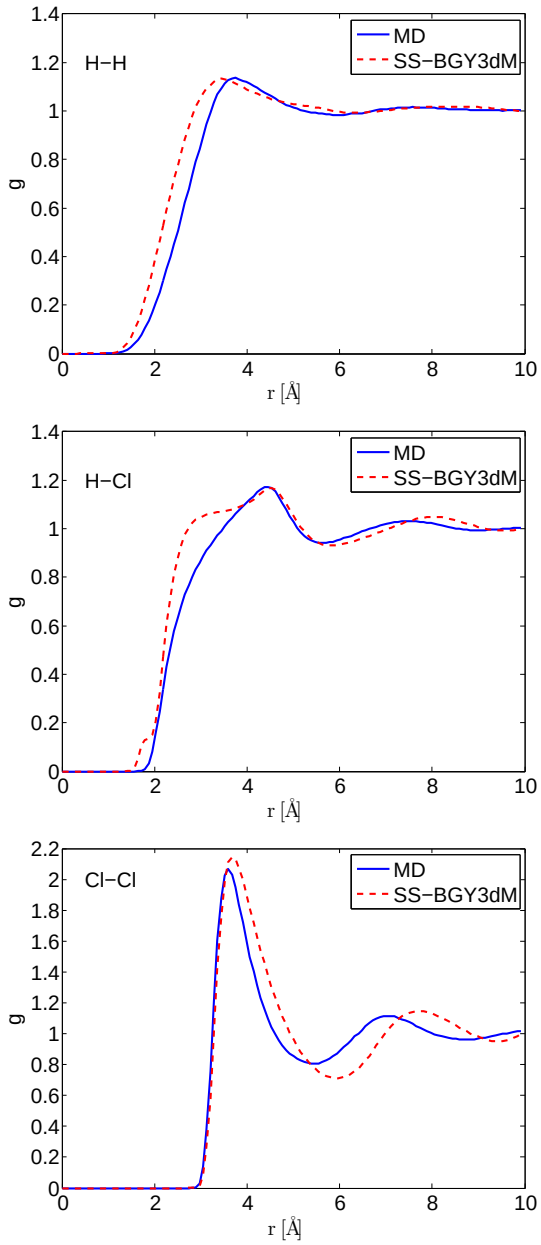

FIG. 3: Radial component of the site-site pair distribution functions for the HCl-like model (HCl2). Comparison between SS-BGY3dM and MD results.

titatively compare the results:

$$
\begin{aligned}
e_{L_2^h} &:= \frac{1}{N}\left(\sum_{\mathbf{i}} |(g_h)_{\mathbf{i}} - (g_h^{MD})_{\mathbf{i}}|^2\right)^{\frac{1}{2}}, \\
e_{L_\infty^h} &:= \max_{\mathbf{i}} |(g_h)_{\mathbf{i}} - (g_h^{MD})_{\mathbf{i}}|, \\
e_{max} &:= |\max_{\mathbf{i}}(g_h)_{\mathbf{i}} - \max_{\mathbf{i}}(g_h^{MD})_{\mathbf{i}}|,
\end{aligned} \tag{79}
$$

where $(g_h)_{\mathbf{i}}$ denotes the respective solution of the SS-BGY3dM equations at grid point $\mathbf{i}$, $\mathbf{i} \in [0, N-1]^3$, and $(g_h^{MD})_{\mathbf{i}}$ is the distribution function at grid point $\mathbf{i}$ computed with MD. To this end, the MD results of the pair distribution functions are interpolated on the three-dimensional grid to compute the above errors. The values for the HCl-like model solvents can be found in Table II.

The apparent differences between the SS-BGY3dM and MD results in Figures 2 and 3 are typical for any method employing the Kirkwood approximation. That is, the exact position and the height of the first peak do not exactly match those of the MD results except for the H-H distribution of HCl1. The frequency of the subse-

	MD	SS-BGY3dM			
	max g	max g	$e_{L_2^h}$	$e_{L_\infty^h}$	e_{max}
H-H (HCl1)	1.32	1.32	3.549_{-6}	6.223_{-2}	0.00
H-Cl (HCl1)	1.29	1.37	4.385_{-6}	8.698_{-2}	0.08
Cl-Cl (HCl1)	1.99	1.93	1.491_{-5}	3.194_{-1}	0.06
H-H (HCl2)	1.14	1.13	6.893_{-6}	2.692_{-1}	0.00
H-Cl (HCl2)	1.17	1.18	7.516_{-6}	3.000_{-1}	0.01
Cl-Cl (HCl2)	2.07	2.48	2.896_{-5}	5.756_{-1}	0.41

TABLE II: Comparison of SS-BGY3dM with MD results for the site-site pair distribution functions of the HCl-like models.

	MD	BGY3dM			
	max g	max g	$e_{L_2^h}$	$e_{L_\infty^h}$	e_{max}
H (solute: HCl)	1.96	1.56	3.594_{-5}	4.134_{-1}	0.39
Cl (solute: HCl)	3.39	2.38	9.671_{-5}	1.215_{+0}	1.00
H (solute: hexane)	2.42	1.92	4.802_{-5}	5.525_{-1}	0.50
Cl (solute: hexane)	4.49	3.96	9.880_{-5}	1.299_{+0}	0.52

TABLE III: Comparison of BGY3dM with MD results for the site density distributions of the HCl-like solvent around a single HCl and a hexane molecule as solute.

quent oscillation is too low. These errors are known to be a consequence of the two-particle superposition approximation, see e.g. [46]. A comparison of the error values for the different site-site distribution functions of HCl1 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 with the results for HCl2 uncovers another difficulty. All errors are increased for this model, which is due to the decreased value of σ_{H} considered for HCl2. This leads to a small Lennard-Jones sphere of the hydrogen atom which is completely embedded inside that of the chloride atom and results in a worse 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. Similar problems have been observed for the solution of the extended RISM equations of Hirata and Rossky in [44].

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 equations perform better for more similar particle species. However, note that the overall approximation of the presented method is still reasonable. All important features of the distribution functions of the HCl-like HCl1 are reproduced. We conclude that the general form and especially the modeling of the intramolecular bonds within the SS-BGY3dM equations is verified by the results.

B. Site Density Distributions

Next, we test the BGY3dM model with respect to the computation of solvent densities around a solute molecule. For this, we again employ the HCl-like model (HCl1) of Hirata et al. [44] already described above. First, 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 for the BGY3dM equations, are computed by the SS-BGY3dM model. All simulation parameters are chosen as described above. Details on how the site density distribution is computed by MD can be found in [47]. In this case, a total of $3.2\ 10^8$ MD time steps were necessary to reach a satisfactory level of convergence.

The computed site densities and their deviation are depicted in Figures 4 and 5. The computed error quantities can be found in Table III. The MD results still show distinct fluctuations, but all features of the distribution functions have clearly developed. A comparison of the results of the BGY3dM model and MD shows a satisfying agreement. The low L_2-errors indicate a good overall approximation. The first peak and the subsequent oscillation pattern are reproduced with a sufficient accuracy considering the approximations involved in the model. The main difference can be observed at the location of the first peak of the distributions. All other errors are not resolved in the plots due to the fluctuations of the MD results. The L_∞-error is about 0.4 and 1.2 for the hydrogen and the chloride distribution, respectively, and is also 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 the discussion above. Recall that the site-site pair distribution functions are required as input of the BGY3dM model. They are computed with the approximate SS-BGY3dM model. Hence, the approximation error enters twice: directly via the approximation involved in the BGY3dM model and by the use of the approximated site-site pair distribution functions computed with the SS-BGY3dM model.

Finally, we compare the site distribution functions of the HCl-like model solvent around a single hexane molecule ($CH_3(CH_2)_4CH_3$) as solute. The potential parameters for hexane are taken from the general-purpose force field OPLS [48]. The computed site densities and the difference between the BGY3dM and MD results are shown in Figures 6 and 7. The computed error quantities can be found in Table III.

As for the HCl molecule as solute the overall agreement between the BGY3dM and MD results is good. The computed error values are similar to those for HCl as solute.

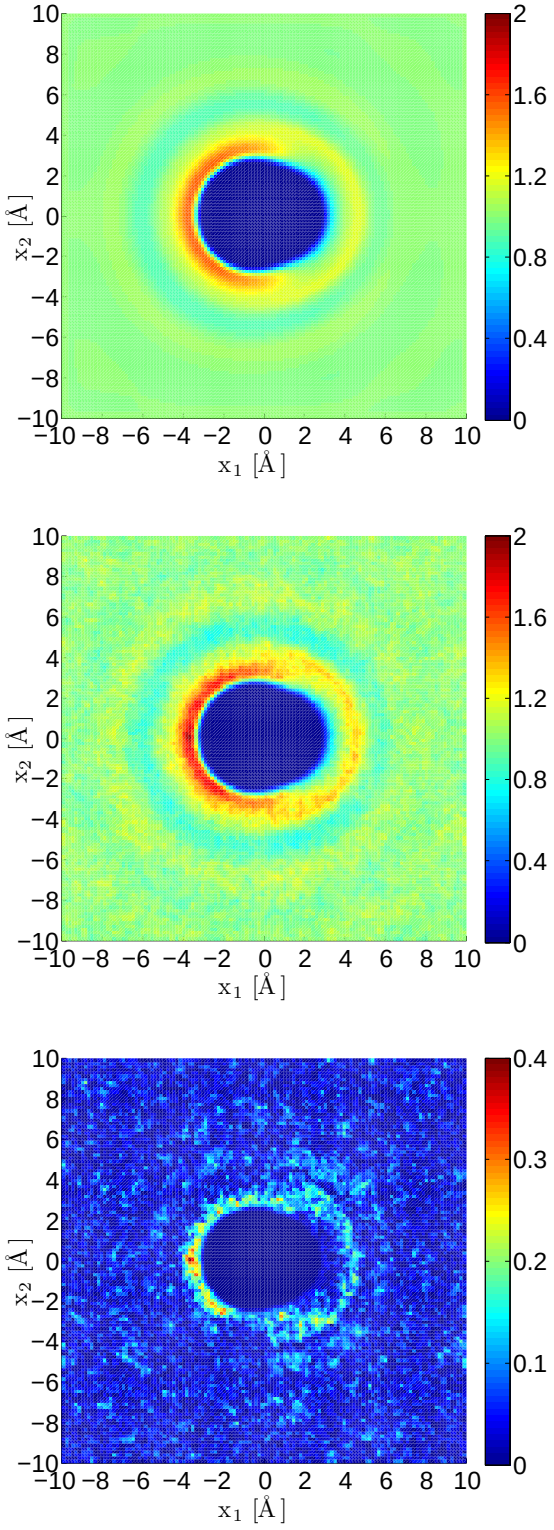

FIG. 4: Hydrogen distribution around a single HCl molecule at the $x_3 = 0$ plane. Top: BGY3dM. Middle: MD. Bottom: Difference between BGY3dM and MD.

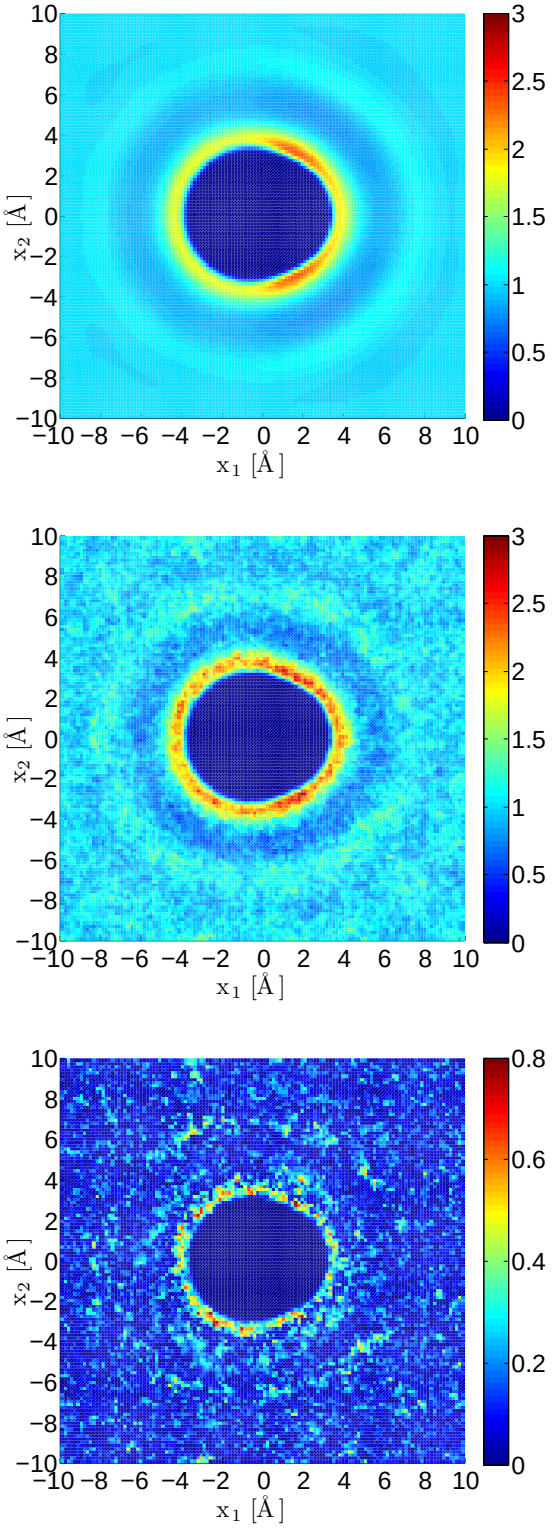

FIG. 5: Chloride distribution around a single HCl molecule at the $x_3 = 0$ plane. Top: BGY3dM. Middle: MD. Bottom: Difference between BGY3dM and MD.

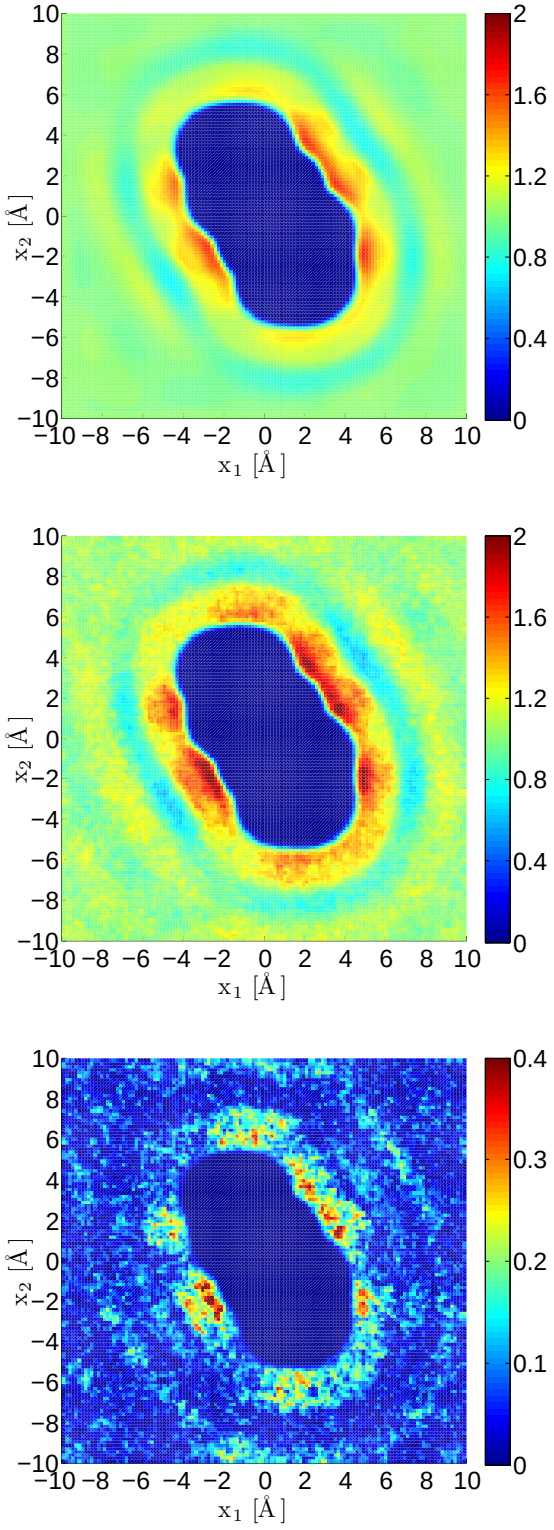

FIG. 6: Hydrogen distribution around a hexane molecule at the $x_3 = 0$ plane. Top: BGY3dM. Middle: MD. Bottom: Difference between BGY3dM and MD.

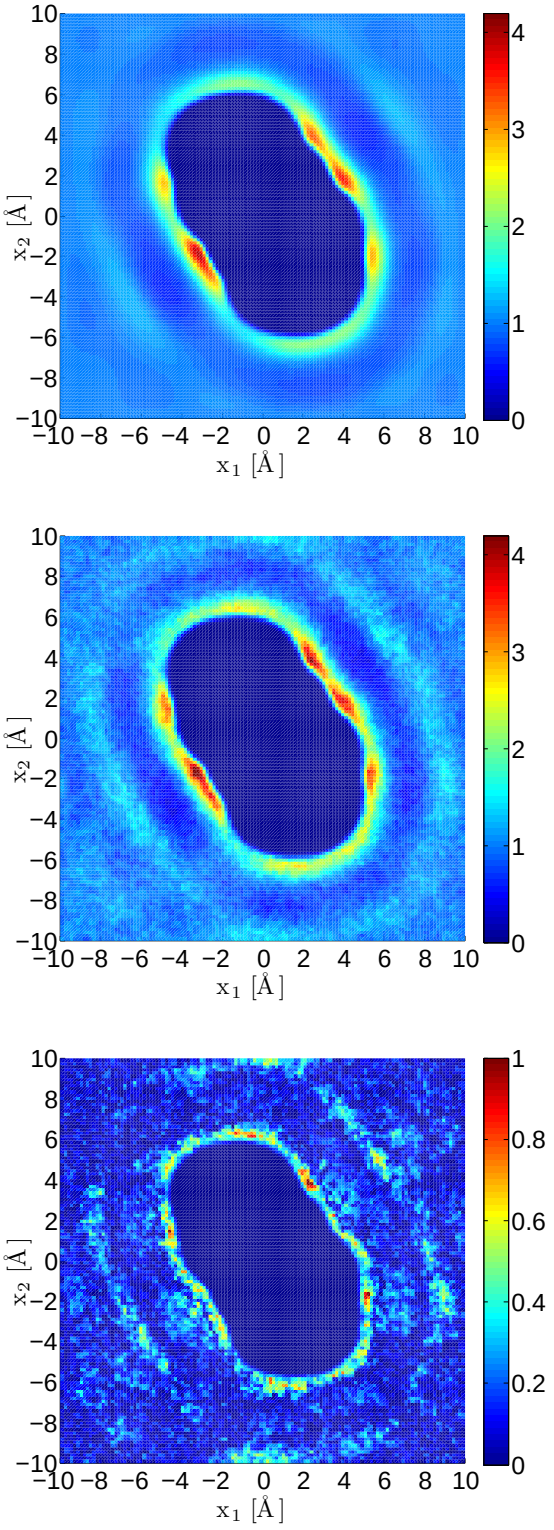

FIG. 7: Chloride distribution around a hexane molecule at the $x_3 = 0$ plane. Top: BGY3dM. Middle: MD. Bottom: Difference between BGY3dM and MD.

But again, the magnitude of the maxima in the first shell of the approximated hydrogen and chloride distributions around the solute are considerably small. Moreover, the predicted magnitude of the oscillation pattern which follows the main peak is also small compared to the MD results. Nevertheless, the positions of the maxima of the first shell and the form of the subsequent oscillation pattern is reproduced well. We can conclude that the accuracy of the computed site distribution functions is nearly independent of the size and form of the considered solute as long as the interactions between solute and solvent are similar in strength. The atoms of the considered HCl and hexane models do not carry large partial charges directly exposed to the solvent. Hence, the approximation of the BGY3dM model yields similar results in both cases. This property is important to predict the error of a computed site density without actually comparing it to MD results.

C. Examples

We have seen that the BGY3dM model leads to a satisfying agreement between the computed site densities and the results of a MD simulation. Now, we present results obtained by the BGY3dM model for a more realistic fluid. For this, 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. [49]. 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. [50]. As before, the functional form of the interaction potential is given as a sum of Lennard-Jones and Coulomb terms, see equation (77) with $\alpha, \gamma = \mathrm{C}, \mathrm{S}$ in this case.

Contrary to the HCl-like model solvent, the carbon disulfide model is a three-site model. Since the two sulfur atoms of CS_2 are identical, we again have to compute three different site-site pair distribution functions and two different site distributions. The pair distribution functions are computed by a MD simulation of 80 CS_2 molecules at a number density of $\rho = 0.01/\text{Å}^3$ and a temperature of $T = 360\text{K}$. The resulting C-C, C-S and S-S pair distributions are depicted in Figure 8. They are used as input for the BGY3dM equations.

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 9 shows the site distributions at the $x_3 = 0$ plane. The carbon distribution exhibits a broad maximum around the entire solute molecule. This maximum results from a superposition of the van der Waals attraction modeled by the Lennard-Jones potential between the carbon atoms and the solute, and from the Coulomb attraction between the solvent carbon and the solute sulfur atoms. The solvent sulfur distribution shows a sharp peak around the solute carbon particle due to the strong Coulomb interaction between them. This

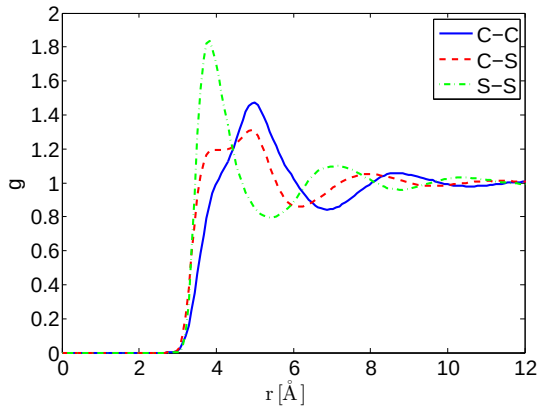

FIG. 8: The site-site pair distribution functions of carbon disulfide computed by MD.

can also be observed in Figure 10, where the charge distribution is plotted. The charge distribution g_{charge} of carbon disulfide can be computed from the site distribution functions g_C and g_S by

$$g_{charge} = q_\mathrm{C} g_\mathrm{C} + 2 q_\mathrm{S} g_\mathrm{S} \tag{80}$$

with q_α the charge of site $\alpha = \mathrm{C}, \mathrm{S}$. As can be expected, a closed band of high sulfur density evolves around the solute carbon whereas the solvent carbon is more likely to be found next to the solute sulfur atoms.

Next, we consider methanol as solute in carbon disulfide as solvent. 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 [51]. Due to the alcohol specific OH-group, methanol is a polar molecule. The oxygen and hydrogen atoms carry strong opposed charges. We again employ the OPLS force field [48] for the parameter set of methanol. To this end, the hydrogen particle of the OH-group is modeled as a charge carrying site without Lennard-Jones interaction. However, for the numerical stability of the BGY3dM equations a hard core potential is required at the position of any atom. Hence, we introduce Lennard-Jones parameters for the oxygen bonded hydrogen and choose $\sigma_\mathrm{H} = 3.4\text{Å}$ and $\epsilon_\mathrm{H} = 0.03\text{kcal/mol}$. The high value of σ_H includes an empirical correction and ensures a stable convergence of the BGY3dM equations. 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 an intramolecular coupling in the BGY3dM model where the two sulfur sites are incorporated without considering their relative position to the carbon site. 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. In order to compensate for this missing three-body effect,

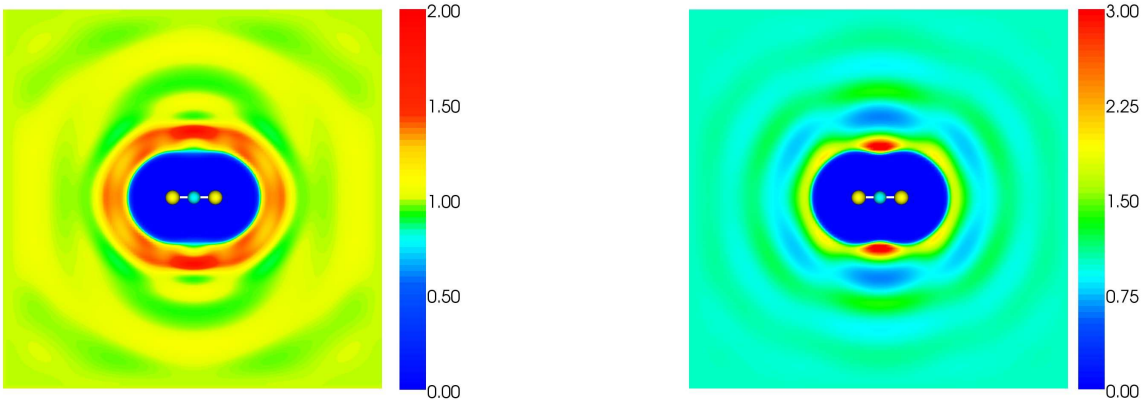

FIG. 9: Distribution functions of carbon disulfide around a CS_2 molecule at the $x_3 = 0$ plane. Carbon distribution (left) and sulfur distribution (right).

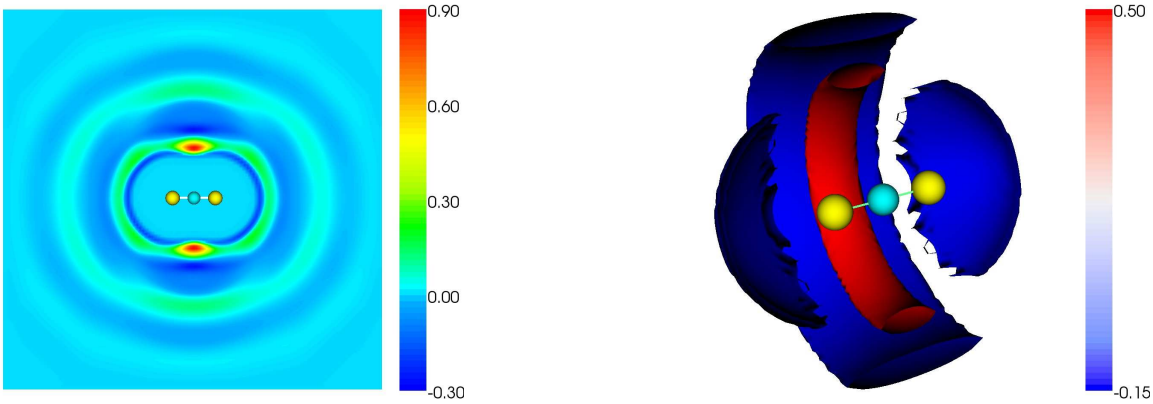

FIG. 10: Charge distribution of carbon disulfide around a CS_2 molecule. Cut at the $x_3 = 0$ plane (left) and isosurface plot (right).

we choose the high value of $\sigma_H = 3.4$Å as an empirical correction.

Figure 11 shows the carbon and sulfur distributions around methanol at the $x_3 = 0$ plane. Here, the methanol molecule is depicted 2Å above the plane for visualization purposes. It is obvious that the strong Coulomb interaction strongly 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. The plots of the charge distributions in Figure 12 also uncover the 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 the different solvent sites. 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.

VI. CONCLUSIONS

We have presented the BGY3dM model for the approximation of solvent densities around solutes of arbitrary shape. The model is directly derived from the YBG-hierarchy and comprises the Kirkwood superposition approximation as closure relation for the intermolecular 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. Since the Kirkwood approximation is not appropriate for terms including intramolecular interactions, we employ a slightly simplified version of the normalized site-site superposition approximation (NSSA) of Taylor et al. [25] for these terms.

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

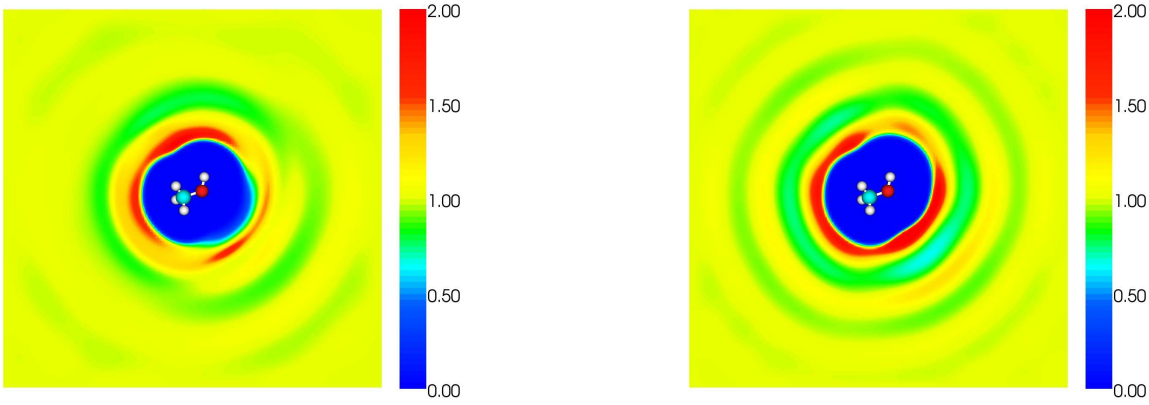

FIG. 11: Distribution functions of carbon disulfide around a methanol molecule at the $x_3 = 0$ plane. Carbon distribution (left) and sulfur distribution (right).

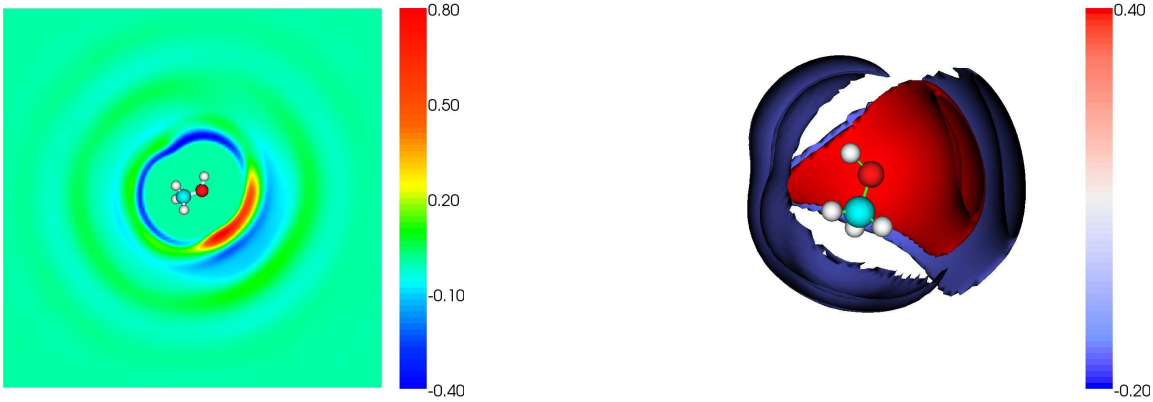

FIG. 12: Charge distribution of carbon disulfide around a methanol molecule. Cut at the $x_3 = 0$ plane (left) and isosurface plot (right).

which are therefore directly dealt with 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 which can efficiently be solved by a finite difference scheme with an iterative GMRES solver. Finally, we also derived the SS-BGY3dM equations 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 MD revealed a good overall performance of our method. 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 consistent with the results. Nevertheless, the involved approximations have to be further modified to reach an improved level of accuracy at realistic temperatures and for stronger interactions. A promising approach would be the introduction of an empirical correction specific for a certain solvent. This would be similar to the empirical bridge functions as they are employed by Du et al. [13] and Kovalenko et al. [6] for water as solvent. A better accuracy could also be gained by employing the optimal superposition approximation of the intramolecular terms as it has been derived by Attard [26]. This, however, requires an iterative solution of an additional system of equations for the functions to be superposed and would therefore increase the computational costs.

In conclusion we note that our results are promising. The BGY3dM model is able to reproduce the important features of the site distribution functions around an arbitrary solute. To consider more realistic solvents, such as water, empirical corrections of the approximations are currently under development.

Acknowledgments

This work was supported by the “Deutsche Forschungsgemeinschaft” through the SFB 611 “Singuläre Phänomene und Skalierung in mathematischen Modellen”.

[1] 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).

[2] B. Roux and T. Simonson, Biophys. Chem. **78**, 1 (1999).

[3] B. Roux, in *Computational Biophysics*, edited by O. Becker, A. D. MacKerrel, B. Roux, and M. Watanabe (Marcel Dekker Inc, New York, 2001).

[4] M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).

[5] D. Beglov and B. Roux, J. Chem. Phys. **103**, 360 (1995).

[6] A. Kovalenko and F. Hirata, J. Chem. Phys. **113**, 2793 (2000).

[7] A. Kovalenko and F. Hirata, Chem. Phys. Lett. **290**, 237 (1998).

[8] A. Kovalenko and F. Hirata, J. Phys. Chem. B **103**, 7942 (1999).

[9] A. Kovalenko and F. Hirata, J. Chem. Phys. **112**, 10391 (2000).

[10] A. Kovalenko and F. Hirata, J. Chem. Phys. **112**, 10403 (2000).

[11] A. Kovalenko and F. Hirata, J. Chem. Phys. **113**, 9830 (2000).

[12] A. Kovalenko and T. N. Truong, J. Chem. Phys. **113**, 7458 (2000).

[13] Q. Du, D. Beglov, and B. Roux, J. Phys. Chem. **104**, 796 (2000).

[14] D. Beglov and B. Roux, J. Phys. Chem. B **101**, 7821 (1997).

[15] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).

[16] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, London, 1986), 2nd ed.

[17] E. Meeron, J. Chem. Phys. **27**, 1238 (1957).

[18] E. E. Salpeter, Ann. Phys. **5**, 183 (1958).

[19] S. A. Rice and J. Lekner, J. Chem. Phys. **42**, 3559 (1965).

[20] S. A. Rice and D. A. Young, Discuss. Faraday Soc. **43**, 16 (1967).

[21] I. Z. Fisher and B. I. Kopeliovich, Dokl. Akad. Nauk. SSSR **133**, 81 (1960).

[22] Y.-T. Lee, F. H. Ree, and T. Ree, J. Chem. Phys. **48**, 3506 (1968).

[23] F. H. Ree, Y.-T. Lee, and T. Ree, J. Chem. Phys. **55**, 234 (1971).

[24] B. C. Eu and H. H. Gan, J. Chem. Phys. **99**, 4084 (1993).

[25] M. P. Taylor and J. E. G. Lipson, J. Chem. Phys. **100**, 519 (1994).

[26] P. Attard, J. Chem. Phys. **102**, 5411 (1995).

[27] H. H. Gan and B. C. Eu, J. Chem. Phys. **99**, 4103 (1993).

[28] H. H. Gan and B. C. Eu, Journal of Polymeric Science: Part B: Polymer Physics **33**, 2319 (1995).

[29] H. H. Gan and B. C. Eu, AIChE Journal: Materials, Interfaces, and Electrochemical Phenomena **42**, 2960 (1996).

[30] H. H. Gan and B. C. Eu, J. Chem. Phys. **110**, 3235 (1999).

[31] M. P. Taylor and J. E. G. Lipson, J. Chem. Phys. **102**, 2118 (1995).

[32] M. P. Taylor and J. E. G. Lipson, J. Chem. Phys. **102**, 6272 (1995).

[33] M. P. Taylor and J. E. G. Lipson, J. Chem. Phys. **104**, 4835 (1996).

[34] M. P. Taylor, J. L. Mar, and J. E. G. Lipson, 1997 **106**, 5181 (J. Chem. Phys.).

[35] M. P. Taylor, J. Luettmer-Strathmann, and J. E. G. Lipson, J. Chem. Phys. **114**, 5654 (2001).

[36] S. Sokolowski and J. Fischer, Mol. Phys. **70**, 1097 (1990).

[37] A. Singer, J. Chem. Phys. **121**, 3657 (2004).

[38] M. Frigo and S. G. Johnson, in *Special Issue on Program Generation, Optimization, and Platform Adaptation* (2005), vol. 93 (2) of *Proceedings of the IEEE*, pp. 216–231.

[39] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. (1993).

[40] A. Alastuey and P. A. Martin, J. Stat. Phys. **39**, 405 (1985).

[41] 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.

[42] S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith, in *Modern Software Tools in Scientific Computing*, edited by E. Arge, A. M. Bruaset, and H. P. Langtangen (Birkhäuser Press, 1997), pp. 163–202.

[43] S. Balay, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang, Tech. Rep. ANL-95/11 - Revision 2.1.5, Argonne National Laboratory (2004).

[44] F. Hirata, B. M. Pettitt, and P. J. Rossky, J. Chem. Phys. **77**, 509 (1982).

[45] M. Griebel, S. Knapek, and G. Zumbusch, *Numerical Simulation in Molecular Dynamics. Numerics, Algorithms, Parallelization, Applications*, Texts in Computational Science and Engineering (Springer, 2007).

[46] P. Attard, Mol. Phys. **74**, 547 (1991).

[47] L. Jager, Ph.D. thesis, Universität Bonn (2007).

[48] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[49] Carbon Disulfide Fact Sheet, *Carbon Disulfide Fact Sheet* (2007), http://www.npi.gov.au/database/substance-info/profiles/18.html.

[50] S.-B. Zhu, J. Lee, and G. W. Robinson, Mol. Phys. **65**, 65 (1988).

[51] Methanol Fact Sheet, *Methanol Fact Sheet* (2007), http://www.npi.gov.au/database/substance-info/profiles/54.html.