

# **KIRKWOOD–BUFF INTEGRALS FROM MOLECULAR SIMULATION**



# **KIRKWOOD–BUFF INTEGRALS FROM MOLECULAR SIMULATION**

## **Proefschrift**

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op donderdag 24 juni om 10:00 uur

door

**Noura DAWASS**

Master of Science in Chemical Engineering,  
Texas A&M University, Qatar  
Geboren te Doha, Qatar.

Dit proefschrift is goedgekeurd door de  
promotor: Prof.dr.ir. T.J.H. Vlugt  
copromotor: Dr. O.A. Moulton

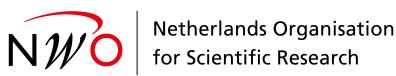
Samenstelling promotiecommissie:

Rector Magnificus,  
Prof.dr.ir. T.J.H. Vlugt  
Dr. O.A. Moulton

voorzitter  
Technische Universiteit Delft, promotor  
Technische Universiteit Delft, copromotor

Onafhankelijke leden:  
Dr. B.P. Tighe  
Prof.dr. D.N. Theodorou  
Prof.dr. F.C. Grozema  
Prof.dr.-ing. J. Gross  
Prof.dr.ir. J. Sietsma

Technische Universiteit Delft  
National Technical University of Athens, Greece  
Technische Universiteit Delft  
University of Stuttgart, Germany  
Technische Universiteit Delft



This work was sponsored by NWO Exacte Wetenschappen (Physical Sciences) for the use of supercomputer facilities with financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organization for Scientific Research, NWO). The promotor acknowledges NWO-CW for a VICI grant.

*Keywords:* Molecular Thermodynamics, Molecular Simulation, Finite-size Effects, Solution Theory  
*Printed by:* Ridderprint | [www.ridderprint.nl](http://www.ridderprint.nl)

Copyright © 2020 by N. Dawass

ISBN 978-94-6384-224-2

An electronic version of this dissertation is available at  
<http://repository.tudelft.nl/>.

*To my mother and my superhero, Momena.*

*To my late grandfather, Jeddo Khader, I know you would have appreciated  
this endeavor the most.*



# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The Kirkwood–Buff theory . . . . .	3
1.2	Inversion of the Kirkwood–Buff theory . . . . .	6
1.3	Kirkwood–Buff Integrals from molecular simulations . . . . .	8
1.3.1	Thermodynamic properties from molecular simulations . . . . .	8
1.3.2	Methods for computing Kirkwood–Buff Integrals . . . . .	9
1.4	Applications of Kirkwood–Buff Integrals computed from molecular simulation . . . . .	16
1.4.1	Partial molar enthalpies . . . . .	16
1.4.2	Properties of single-ions in salt solutions . . . . .	16
1.4.3	Mass transfer in multicomponent liquids . . . . .	17
1.4.4	Other applications . . . . .	19
1.5	Scope of this thesis . . . . .	20
<b>2</b>	<b>Shape Effects of Kirkwood–Buff Integrals of Finite Systems</b>	<b>23</b>
2.1	Introduction . . . . .	24
2.2	Numerical computation of $w(x)$ . . . . .	25
2.2.1	Importance sampling algorithm for computing $p(i)$ . . . . .	27
2.2.2	Computing $w(x)$ . . . . .	28
2.3	Shape effects of Kirkwood–Buff Integrals . . . . .	28
2.3.1	The functions $w(x)$ for a cube, cuboids, and spheroids . . . . .	28
2.3.2	Kirkwood–Buff Integrals for various shapes . . . . .	32
2.4	Conclusions . . . . .	37
<b>3</b>	<b>Finite-Size Effects of Kirkwood–Buff Integrals from Molecular Simulations</b>	<b>39</b>
3.1	Introduction . . . . .	40
3.2	Finite-size effects of subvolumes . . . . .	41
3.3	Finite-size effects of radial distribution functions . . . . .	42
3.3.1	Van der Vegt correction . . . . .	43
3.3.2	Inverse- $N$ finite-size correction . . . . .	44
3.3.3	Cortes-Huerto et al. correction . . . . .	44
3.3.4	Simulation details . . . . .	46

3.4	<b>Results and discussion . . . . .</b>	46
3.4.1	<b>Finite-size effects of subvolumes . . . . .</b>	46
3.4.2	<b>Finite-size effects of radial distribution functions . . . . .</b>	52
3.5	<b>Conclusions . . . . .</b>	63
<b>4</b>	<b>Surface Effects of Kirkwood–Buff Integrals from Molecular Simulations</b>	<b>65</b>
4.1	<b>Introduction . . . . .</b>	66
4.2	<b>Methods . . . . .</b>	68
4.2.1	<b>Simulation details . . . . .</b>	69
4.3	<b>Results . . . . .</b>	70
4.3.1	<b>Estimation of Kirkwood–Buff Integrals . . . . .</b>	70
4.3.2	<b>Estimation of surface effects . . . . .</b>	84
4.4	<b>Conclusions . . . . .</b>	90
<b>5</b>	<b>Thermodynamic and Transport Properties of Urea–Choline Chlorine Mixtures</b>	<b>91</b>
5.1	<b>Introduction . . . . .</b>	92
5.2	<b>Methods . . . . .</b>	93
5.2.1	<b>Radial distribution functions of pseudo–binary mixtures . . . . .</b>	93
5.2.2	<b>Maxwell–Stefan diffusion coefficients of pseudo–binary mixtures . . . . .</b>	98
5.2.3	<b>Force field . . . . .</b>	99
5.2.4	<b>Simulation details . . . . .</b>	99
5.3	<b>Results and discussion . . . . .</b>	101
5.3.1	<b>Kirkwood–Buff Integrals of urea–choline chloride mixtures . . . . .</b>	101
5.3.2	<b>Thermodynamic properties of urea–choline chloride mixtures . . . . .</b>	105
5.3.3	<b>Transport properties of urea–choline chloride mixtures . . . . .</b>	108
5.4	<b>Conclusions . . . . .</b>	112
<b>6</b>	<b>Solubility of Gases in Monoethylene Glycol</b>	<b>113</b>
6.1	<b>Introduction . . . . .</b>	114
6.2	<b>Methods . . . . .</b>	116
6.2.1	<b>Force fields . . . . .</b>	116
6.2.2	<b>The CFCMC method in the osmotic ensemble . . . . .</b>	117
6.2.3	<b>Simulation details . . . . .</b>	119

6.3 Results and discussion . . . . .	122
6.3.1 Solubility of CO <sub>2</sub> in monoethylene glycol . . . . .	122
6.3.2 Solubility of CH <sub>4</sub> , H <sub>2</sub> S and N <sub>2</sub> in monoethylene glycol . . . . .	127
6.4 Conclusions . . . . .	130
<b>Conclusions</b>	<b>133</b>
<b>Bibliography</b>	<b>137</b>
<b>Summary</b>	<b>161</b>
<b>Samenvatting</b>	<b>165</b>
<b>Curriculum Vitae</b>	<b>169</b>
<b>Publications</b>	<b>171</b>
<b>Acknowledgments</b>	<b>173</b>



# 1

## INTRODUCTION

A homogenous solution consists of two or more components that are uniformly mixed. Homogenous liquid solutions are present in biological, medical, geological, and industrial applications. To design and develop industrial processes, knowledge of the thermodynamic and transport properties of solutions is essential [1–6]. This task is not simple, especially for systems where strong intermolecular forces are present such as aqueous solutions, ionic liquids (ILs), and deep eutectic solvents (DESSs) [7–11]. The advantage of using a molecular theory of solutions, as opposed to experiments and classical thermodynamic models, is that bulk properties are provided by directly considering molecular interactions and structure. In this regard, the Kirkwood–Buff (KB) theory [12] provides an important connection between the microscopic structure of fluid mixtures and the corresponding macroscopic properties. Rooted in statistical mechanics, the KB theory applies to any type of intermolecular interactions, making it one of the most general and important theories for homogenous solutions [12–15]. Kirkwood and Buff [12] expressed thermodynamic quantities such as partial derivatives of chemical potentials with respect to composition, partial molar volumes, and the isothermal compressibility in terms of integrals of radial distribution functions (RDFs) over infinite and open volumes. These integrals, which are considered the key quantity in the KB theory, are referred to as KB Integrals (KBIs). Alternatively, KBIs can be obtained from density fluctuations in the grand-canonical ensemble [13, 16].

The KB theory was derived in 1951, however, it has not gained much interest until the late 70s of the previous century after Ben–Naim [14] proposed the inversion of the KB theory. The inversion of the theory allows the calculation of KBIs from experimental data [17–20]. Thirty years following the inversion of the KB theory, molecular simulation emerged as a powerful tool for studying pure liquids and mixtures [21]. There are two main types of molecular simulation techniques [22, 23]: Molecular Dynamics (MD), where trajectories of molecules are determined by solving Newton's equation of motion numerically; and Monte Carlo (MC) simulations, where relevant states of the system are sampled according to their statistical weight [21–25]. In both simulation techniques, RDFs and local density fluctuations are easily computed, thus in principle enabling the calculation of KBIs. Molecular simulations can be used to study closed systems with a fixed number of molecules, or open systems in which the number of molecules fluctuates [22]. It is important to note that molecular simulations can only be performed for finite systems, while the KB theory requires KBIs for infinite and open systems [12]. This disparity between the KB theory and molecular simulations has to be considered when computing KBIs from simulations. The focus of this thesis is to provide a framework to accurately compute KBIs using molecular simulation. In this chapter, we briefly introduce the KB theory and provide

the most important relations that link microscopic structure with macroscopic properties (section 1.1). This is followed by a discussion of the inversion of the KB theory and its applications (section 1.2). Also, we present the available methods for computing KBIs from molecular simulations (section 1.3) and review the applications of KBIs computed from molecular simulations (section 1.4). Finally, the scope of the thesis is presented (section 1.5).

## 1.1. THE KIRKWOOD-BUFF THEORY

One of the fundamental quantities for describing the microscopic structure of fluids, are RDFs [13, 26]. Essentially, RDFs provide the probability of finding a molecule at a distance  $r$  from a central molecule. For homogenous and isotropic fluids, RDFs  $g_{\alpha\beta}(r)$  are defined as [13]:

$$g_{\alpha\beta}(r) = \frac{\rho_{\alpha\beta}(r)}{\rho_\beta} \quad (1.1)$$

where  $\rho_{\alpha\beta}(r)$  is the local density of component  $\beta$  at a distance  $r$  from a central molecule of type  $\alpha$ , and  $\rho_\beta$  is the bulk density of component  $\beta$ . RDFs can be determined from scattering experiments as well as from molecular simulation. Using molecular simulations, RDFs are frequently computed using particles counting [22]. Alternatively, force-based computations of RDFs can be implemented [27–29]. Commonly, particle counting is adopted in molecular simulation packages with RDFs computed from [26]:

$$g_{\alpha\beta}(r) = \frac{V}{N_\alpha N_\beta} \left\langle \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta} \delta(\mathbf{r} - \mathbf{r}_j + \mathbf{r}_i) \right\rangle \quad (1.2)$$

where  $N_\alpha$  and  $N_\beta$  are the number of molecules of components  $\alpha$  and  $\beta$ , respectively.  $\delta$  is the Dirac delta function,  $r_i$  is the position of atom  $i$ , and the brackets  $\langle \dots \rangle$  indicate an ensemble average. When  $\alpha$  equals  $\beta$ , terms where  $i = j$  should be excluded in the double summation of Eq. (1.2). RDFs are central in the KB theory, where the local structure of fluids is related to macroscopic properties. In this section, we review the most important relations derived by Kirkwood and Buff [12]. For the original formulation of the theory, the reader is referred to the paper by Kirkwood and Buff [12]. A very detailed derivation was presented by Newman [30], and an alternative derivation was provided by Hall [15].

In the grand-canonical ( $\mu TV$ ) ensemble, thermodynamic quantities are related to KBIs  $G_{\alpha\beta}^\infty$  for an open and infinite system as [12]:

$$G_{\alpha\beta}^\infty = \int_0^\infty dr 4\pi r^2 [g_{\alpha\beta}^\infty(r) - 1] \quad (1.3)$$

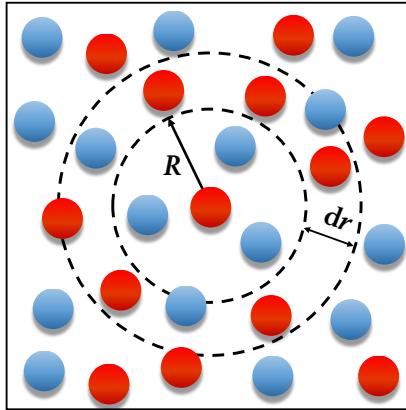


Figure 1.1: A schematic representation of a radial shell in a binary mixture (composed of molecules of types  $\alpha$  and  $\beta$ ) formed by considering a distance  $R$  from a central molecule of type  $\alpha$ . The number of molecules of type  $\beta$  inside radial shell elements with the width  $dr$  is used to compute the RDF  $g_{\alpha\beta}(r)$ .

where  $r$  is the particle distance and  $g_{\alpha\beta}^\infty(r)$  is the radial distribution function (RDF) of species  $\alpha$  and  $\beta$  for an infinitely large system. In Eq. (1.3), species  $\alpha$  and  $\beta$  can be the same. For a shell with thickness  $dr$  centred around a molecule of type  $\alpha$  in an infinite system (see Figure 1.1), the number of molecules of type  $\beta$  is  $4\pi r^2 dr \rho_\beta$  and  $4\pi r^2 dr \rho_\beta g_{\alpha\beta}^\infty(r)$  for an ideal gas and real fluid, respectively. Here,  $\rho_\beta = \langle N_\beta \rangle / V$  is the average number density of species  $\beta$ . Integrating from zero to infinity over the excess number of molecules of type  $\beta$ ,  $(4\pi r^2 dr \rho_\beta [g_{\alpha\beta}^\infty(r) - 1])$ , yields  $\rho_\beta G_{\alpha\beta}^\infty$ . Hence, KBIs  $G_{\alpha\beta}^\infty$  provide the average excess (or depletion) per unit density of molecules of type  $\beta$  around a central molecule of type  $\alpha$ , and the affinity between components  $\alpha$  and  $\beta$  is reflected. It is important to note that this interpretation of KBIs only holds for infinite systems, as indicated by the upper bound of the integral in Eq. (1.3). Truncating the integral of Eq. (1.3) to a distance  $R$  yields the average excess of type  $\beta$  within a sphere of radius  $R$ . We will demonstrate later in this thesis (section 2.3) that the resulting truncated integral does not represent the KBIs in the thermodynamic limit.

Kirkwood and Buff [12] formulated a relation between integrals over RDFs and fluctuations in the number of molecules in the grand-canonical ensemble,

$$G_{\alpha\beta}^{\infty} = \int_0^{\infty} dr 4\pi r^2 [g_{\alpha\beta}^{\infty}(r) - 1] = \lim_{V \rightarrow \infty} \left[ V \frac{\langle N_{\alpha} N_{\beta} \rangle - \langle N_{\alpha} \rangle \langle N_{\beta} \rangle}{\langle N_{\alpha} \rangle \langle N_{\beta} \rangle} - \frac{V \delta_{\alpha\beta}}{\langle N_{\alpha} \rangle} \right] \quad (1.4)$$

where  $N_{\alpha}$  and  $N_{\beta}$  are the number of molecules of type  $\alpha$  and  $\beta$ , inside the volume  $V$ .  $\langle N_{\alpha} \rangle$  is the average number of molecules  $\alpha$  and  $\langle N_{\alpha} N_{\beta} \rangle$  is the average product of the number of molecules of components  $\alpha$  and  $\beta$ . It is important to note that Eq. (1.4) holds for any isotropic fluid. Fluctuations in the number of molecules relate to several thermodynamic properties [31, 32]. For a binary system, the following relations can be derived that relate KBIs to [12]:

1. partial derivatives of chemical potential with respect to the number of molecules,

$$\begin{aligned} \left( \frac{\partial \mu_{\alpha}}{\partial N_{\alpha}} \right)_{T,P,N_{\beta}} &= \frac{\rho_{\beta} k_{\text{B}} T}{\rho_{\alpha} V \eta} \\ \left( \frac{\partial \mu_{\alpha}}{\partial N_{\beta}} \right)_{T,P,N_{\alpha}} &= \left( \frac{\partial \mu_{\beta}}{\partial N_{\alpha}} \right)_{T,P,N_{\beta}} = -\frac{k_{\text{B}} T}{V \eta} \end{aligned} \quad (1.5)$$

2. partial molar volumes,

$$\begin{aligned} v_{\alpha} &= \left( \frac{\partial V}{\partial N_{\alpha}} \right)_{T,P,N_{\beta}} = \frac{1 + \rho_{\beta} (G_{\beta\beta} - G_{\alpha\beta})}{\eta} \\ v_{\beta} &= \left( \frac{\partial V}{\partial N_{\beta}} \right)_{T,P,N_{\alpha}} = \frac{1 + \rho_{\alpha} (G_{\alpha\alpha} - G_{\alpha\beta})}{\eta} \end{aligned} \quad (1.6)$$

3. the isothermal compressibility,

$$\kappa_T = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_{T,N_{\alpha},N_{\beta}} = \frac{\zeta}{k_{\text{B}} T \eta} \quad (1.7)$$

where  $\left( \frac{\partial \mu_{\alpha}}{\partial N_{\alpha}} \right)_{T,P,N_{\beta}}$  is the partial derivative of the chemical potential of component  $\alpha$  with respect to  $N_{\alpha}$  at a constant temperature  $T$ , pressure  $P$  and  $N_{\beta}$ . Similarly,  $\left( \frac{\partial \mu_{\beta}}{\partial N_{\beta}} \right)_{T,P,N_{\alpha}}$  is the partial derivative of the chemical potential of component  $\beta$  with respect to  $N_{\beta}$  at a constant  $T$ ,  $P$  and  $N_{\alpha}$ . In Eqs. (1.5) and (1.7),  $k_{\text{B}}$  is the Boltzmann constant.  $v_{\alpha}$  is the partial molar volume of component  $\alpha$  at a constant  $T$ ,

$P$ , and  $N_\beta$ .  $v_\beta$  is the partial molar volume of component  $\beta$  at a constant  $T$ ,  $P$ , and  $N_\alpha$ .  $\kappa_T$  is the compressibility at a constant  $T$ .  $\eta$  and  $\zeta$  are auxiliary quantities that were defined for convenience [13],

$$\eta = \rho_\alpha + \rho_\beta + \rho_\alpha \rho_\beta G_f \quad (1.8)$$

$$\zeta = 1 + \rho_\alpha G_{\alpha\alpha} + \rho_\beta G_{\beta\beta} + \rho_\alpha \rho_\beta (G_{\alpha\alpha} G_{\beta\beta} - G_{\alpha\beta}^2) \quad (1.9)$$

In Eq. (1.8), the term  $G_f = G_{\alpha\alpha} + G_{\beta\beta} - 2G_{\alpha\beta}$  can be used to indicate the thermodynamic ideality of a binary mixture (i.e. it has the value of zero for ideal solutions). Expressions for ternary and multi-component mixtures of these thermodynamic quantities in terms of KBIs are available in literature [13, 17, 33].

## 1.2. INVERSION OF THE KIRKWOOD–BUFF THEORY

Prior to the use of molecular simulation to compute KBIs, the inversion of the KB theory [13, 14] was used to obtain KBIs from experimental data. In this section we will briefly discuss the inversion procedure, and some of its applications.

For a binary mixture with components  $\alpha$  and  $\beta$ , partial molar volumes, the isothermal compressibility, and partial derivatives of chemical potential with respect to number of molecules are related to KBIs  $G_{\alpha\alpha}^\infty$ ,  $G_{\beta\beta}^\infty$  and  $G_{\alpha\beta}^\infty$  (Eqs. (1.6), (1.7), and (1.5)). Moreover, the Gibbs-Duhem relations apply to these thermodynamic quantities,

$$\begin{aligned} \rho_\alpha \left( \frac{\partial \mu_\alpha}{\partial N_\alpha} \right)_{T,P,N_\beta} + \rho_\beta \left( \frac{\partial \mu_\beta}{\partial N_\alpha} \right)_{T,P,N_\beta} &= 0 \\ \rho_\beta \left( \frac{\partial \mu_\beta}{\partial N_\beta} \right)_{T,P,N_\alpha} + \rho_\alpha \left( \frac{\partial \mu_\alpha}{\partial N_\beta} \right)_{T,P,N_\alpha} &= 0 \\ \rho_\alpha v_\alpha + \rho_\beta v_\beta &= 1 \end{aligned} \quad (1.10)$$

where  $v_\alpha$  and  $v_\beta$  are the partial molar volumes of components  $\alpha$  and  $\beta$ , respectively. Using Eqs. (1.5), (1.6), (1.7), and (1.10), Ben-Naim [14] derived the following expression for KBIs of binary mixtures,

$$G_{\alpha\beta}^\infty = k_B T \kappa_T - \frac{\delta_{\alpha\beta}}{\rho_\alpha} + \rho k_B T \frac{(1 - \rho_\alpha v_\alpha)(1 - \rho_\beta v_\beta)}{\rho_\alpha \rho_\beta \left( \frac{\partial \mu_\alpha}{\partial N_\beta} \right)_{T,P,N_\alpha}} \quad (1.11)$$

where the isothermal compressibility  $\kappa_T$  and partial molar volumes  $v_\alpha$  and  $v_\beta$  are obtained from experiments. The term  $\left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\alpha}$  can be obtained using second derivatives of the Gibbs excess energy, or experimental vapor pressure data [13]. In Refs. [33] and [34], equations for KBIs in terms of thermodynamic properties were derived for ternary mixtures.

Ben-Naim [14] introduced the inversion procedure in 1977 and applied it to a mixture of water and ethanol. For water (W) and solute (S) systems, it was shown that KBIs obtained from experimental data are useful for studying several local phenomena: (1) the quantity  $G_{WS}^\infty = G_{SW}^\infty$  indicates the affinity between the solvent and the solute; (2) KBIs of water,  $G_{WW}^\infty$ , reflect the water-water affinity, which can be used to study the changes in the molecular structure of water when adding solutes; and (3) KBIs of solutes,  $G_{SS}^\infty$ , are of particular interest for studying hydrophobic interactions.

Following the work of Ben-Naim [14], the inversion of the KB theory was applied to study various types of binary and ternary mixtures at the molecular level [19, 35–43]. For instance, Patil [39] computed KBIs of water-butanol mixtures from experimental data of partial molar volumes, isothermal compressibilities, and vapor pressures. KBIs of the considered system were used to study local structure at various concentrations. Similarly, Matteoli et al. [38] used partial molar volumes and isothermal compressibilities of mixtures of water and different organic co-solvents to find KBIs. The KBIs obtained from the inversion procedure were taken as a measure of the net attraction or repulsion, indicating the level of hydrophobicity of these mixtures. More recently, Kobayashi et al. [44] used KBIs to study properties of residual water in ionic liquids. The authors found that the values of KBIs computed using molecular simulation agree with KBIs obtained from experimental data. However, the inversion of the KB theory requires the partial derivatives,  $\left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\alpha}$ , which are difficult to accurately obtain from experimental data [45]. Matteoli et al. [38] demonstrated how the accuracy of KBIs obtained from experimental data is very sensitive to uncertainties in partial derivatives of the chemical potential. Alternatively, KBIs can be obtained from local fluctuations in number of molecules measured by small angle scattering experiments [46], such as SANS and SAXS [46–51]. Perera et al. [52] examined a number of water-alcohol mixtures using KBIs and demonstrated that both methods are reliable and should provide similar values of KBIs. Perera et al. [52] pointed out possible sources of errors leading to inaccurate KBIs when using experimental data. For instance, the largest differences between the two methods were observed at the range where the values of the term  $\left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\alpha}$  in

Eq. (1.11) is close to zero. Almásy et al. [53] obtained KBIs from SANS as well as from vapor pressure data for an ionic liquid. The authors found that scattering experiments and thermodynamic data provided similar KBIs.

### 1.3. KIRKWOOD–BUFF INTEGRALS FROM MOLECULAR SIMULATIONS

#### 1.3.1. THERMODYNAMIC PROPERTIES FROM MOLECULAR SIMULATIONS

Knowledge of chemical potentials and other thermodynamic properties is of great importance for studying the phase equilibria of solutions [54]. Computing excess properties of multicomponent systems using molecular simulation is not trivial. To compute chemical potentials, a number of methods have been developed and evaluated, such as thermodynamic integration [55], and perturbation theory [56]. One of the most widely used methods is the Widom's Test Particle Insertion (WTPI) method [57], where a test particle is randomly inserted in the simulation box and the average Boltzmann factor of the resulting energy change is calculated. In general, molecular insertions are found to be challenging when simulating dense fluids or when strong interactions are present [58]. Recently, the Continuous Fractional Component Monte Carlo (CFCMC) method [59–62] have been developed to improve the efficiency of molecular insertions. By varying the interactions of the fractional molecule with the surrounding molecules, molecules are added/removed gradually during MC simulations. Even with applying these advanced methods, simulating complex fluids such as salt solutions in open ensembles is still challenging [58]. Alternatively, excess properties of solutions can be computed using the KB theory [54]. In the previous section, we showed that KBIs relate directly to partial derivatives of the chemical potential with respect to composition, partial molar volumes, and the isothermal compressibility. KBIs also relate to other thermodynamic properties such as the excess Gibbs energy of mixing. Other than predicting thermodynamic properties, KBIs can be used to investigate local behaviour of solutions, and to connect information obtained from molecular simulations to experimental measurements [13, 14, 63].

Knowledge of solution thermodynamics is also required when studying diffusion. To connect Fick diffusion coefficients, which are measured experimentally, to so-called Maxwell–Stefan (MS) diffusivities computed from MD simulations, the so-called thermodynamic factor  $\Gamma$  is used [16, 64, 65]. The non-ideality of solutions is also quantified by  $\Gamma$  [16, 66, 67]. For an  $n$ -component system,

$$\Gamma_{\alpha\beta} = \delta_{\alpha\beta} + x_\alpha \left( \frac{\partial \ln \gamma_\alpha}{\partial x_\beta} \right)_{T,P,\Sigma} \quad (1.12)$$

where  $\gamma_\alpha$  is the activity coefficient of component  $\alpha$ . In Eq. (1.12), the symbol  $\Sigma$  indicates that  $\gamma_\alpha$  is differentiated with respect to the mole fraction of component  $\beta$ ,  $x_\beta$ , while keeping the mole fraction of the other components constant, except the  $n^{\text{th}}$  one. For a binary mixture at a constant temperature and pressure we have,

$$\Gamma = 1 + x_1 \frac{d \ln \gamma_1}{d x_1} = 1 + x_2 \frac{d \ln \gamma_2}{d x_2} \quad (1.13)$$

where the sum of the mole fractions ( $x_1 + x_2$ ) equals unity when the differentiation is carried out [65, 68]. For a specific solution, the thermodynamic factor provides an indication of the phase stability, since  $\Gamma$  relates to the second derivative of the Gibbs energy with respect to composition [65].  $\Gamma$  is positive for a thermodynamically stable mixture and negative for an unstable one [65]. As discussed earlier, computing properties such as activity coefficients and their derivatives is challenging for fluids with strong interactions. To avoid simulations that require molecular insertions, thermodynamic factors can be computed from KBIs. For a binary system,  $\Gamma$  can be computed using [13, 64, 69]:

$$\Gamma_{\alpha\beta} = 1 - \frac{x_\alpha \rho_\beta (G_{\alpha\alpha} + G_{\beta\beta} - 2G_{\alpha\beta})}{1 + \rho_\beta x_\alpha (G_{\alpha\alpha} + G_{\beta\beta} - 2G_{\alpha\beta})} \quad (1.14)$$

The term  $G_{\alpha\alpha} + G_{\beta\beta} - 2G_{\alpha\beta}$  describes the strength of  $\alpha - \beta$  interactions compared to  $\alpha - \alpha$  and  $\beta - \beta$  interactions. If this term is negative, then  $\alpha - \beta$  attractive interactions are stronger than  $\alpha - \alpha$  and  $\beta - \beta$  interactions and as a result  $\Gamma > 1$ . For an ideal gas, the term  $G_{\alpha\alpha} + G_{\beta\beta} - 2G_{\alpha\beta}$  will be zero and hence  $\Gamma = 1$ . We will use this in chapter 5 to analyse the interactions of DESs. Expressions relating  $\Gamma_{ij}$  to KBIs for ternary [13, 70] and quaternary [71] mixtures are available in literature. The thermodynamic factor also plays an important role for correcting finite-size effects of diffusion coefficients [72, 73]. This is discussed further in section 1.4.3.

### 1.3.2. METHODS FOR COMPUTING KIRKWOOD–BUFF INTEGRALS

KBIs can be computed from fluctuations in the number of particles or RDFs, which are both accessed by molecular simulation. KBIs are defined for infinitely large systems while a finite number of molecules are studied by molecular simulation. To estimate KBIs in the thermodynamic limit ( $G_{\alpha\beta}^\infty$ ) using microscopic information of finite systems, three main approaches can be adopted. The most common approach is to simply truncate KBIs to the size of the simulation box, which results in integrals that converge poorly to the thermodynamic limit [13, 74]. In the second approach, RDFs from finite systems were extended to the thermodynamic limit, using mathematically involved methods [75, 76] that

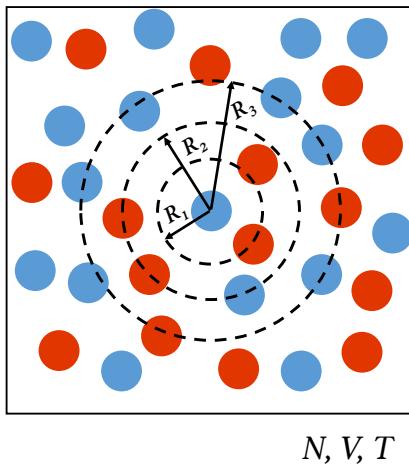


Figure 1.2: A schematic representation of the Small System Method (SSM) [16], where thermodynamic properties are computed from the scaling of properties of small subvolumes with the inverse size of the subvolume.

are difficult to extend to complex molecular systems. Recently, Krüger and co-workers [74] applied the Small System Method (SSM) [16] to develop a practical approach for computing KBIs from molecular simulation. This approach was derived for estimating KBIs of fluids, which is the focus of this thesis, and was extended to solids as shown in the recent work of Miyaji et al. [77]. In the following subsection, the method of Krüger and co-workers [74] will be presented. After that, a summary of the other methods available in literature for computing KBIs from molecular simulation is provided.

### THE METHOD OF KRÜGER AND CO-WORKERS

According to the SSM, properties of small subvolumes, that can be of the order of a few molecular diameters, are treated in terms of thermodynamics of small systems rather than classical thermodynamics [78]. According to Hill's thermodynamics of small systems, properties of open embedded subvolumes scale with the inverse size of the subvolumes [78, 79]. This also applies to KBIs of finite subvolumes,  $G_{\alpha\beta}^V$  [63, 74]. For a specific system,  $G_{\alpha\beta}^V$  computed from a number of subvolumes with different sizes, scale linearly with the inverse size of the subvolume [63, 74, 80, 81]. See Figure 1.2 for an example of multiple subvolumes embedded in a simulation box.

In 2013, Krüger et al. [74] derived an expression for KBIs of finite and open subvolumes embedded in a reservoir. Similar to the original KB theory [12] for infinitely large and open systems, Krüger et al. [74] derived an expression that relates local density fluctuations inside the subvolume with the integral of the RDF of the system. This was achieved by first considering the average densities and the fluctuations in density of the open subvolume,  $V$ , embedded in a large reservoir. The open subvolume,  $V$  is grand-canonical. The system is characterized by the following variables: temperature ( $T$ ), volume of the subvolume ( $V$ ), and chemical potentials ( $\mu_\alpha$  and  $\mu_\beta$  for a binary system). In this ensemble, we consider the average number of molecules,  $\langle N_\alpha \rangle$ , and the average number of  $\alpha$  and  $\beta$  pairs,  $\langle N_\alpha N_\beta \rangle$ , expressed as integrals of the one molecule density ( $\rho_\alpha^{(1)}(\mathbf{r}_1)$ ) and the two molecule density ( $\rho_{\alpha\beta}^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ ),

$$\int_V \rho_\alpha^{(1)}(\mathbf{r}_1) d\mathbf{r}_1 = \langle N_\alpha \rangle \quad (1.15)$$

$$\int_V \int_V \rho_{\alpha\beta}^{(2)}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 = \langle N_\alpha N_\beta \rangle - \delta_{\alpha\beta} \langle N_\alpha \rangle \quad (1.16)$$

Integration of the local densities over the subvolume  $V$  yields the average number of molecules in the grand-canonical ensemble [13]. Subsequently, the density fluctuations in the subvolume  $V$  are expressed as:

$$\int_V d\mathbf{r}_1 \int_V d\mathbf{r}_2 [\rho_{\alpha\beta}^{(2)}(\mathbf{r}_1, \mathbf{r}_2) - \rho_\alpha^{(1)}(\mathbf{r}_1) \rho_\beta^{(1)}(\mathbf{r}_2)] = \langle N_\alpha N_\beta \rangle - \langle N_\alpha \rangle \langle N_\beta \rangle - \delta_{\alpha\beta} \langle N_\alpha \rangle \quad (1.17)$$

For fluid systems,  $\rho_\alpha^{(1)}(\mathbf{r}_1)$  and  $\rho_{\alpha\beta}^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ , can be replaced by  $c_\alpha$ , and  $c_\alpha c_\beta g_{\alpha\beta}(r_{12})$  due to translational and rotational invariance, respectively. Here,  $c_\alpha$  is the macroscopic number density given by  $c_\alpha = \langle N_\alpha \rangle / V$ . The function  $g_{\alpha\beta}(r_{12})$  is the RDF and  $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$ . For a finite multicomponent fluid, the integral,  $G_{\alpha\beta}^V$ , is defined by simply dividing Eq. (1.17) by  $c_\alpha c_\beta V$ :

$$G_{\alpha\beta}^V \equiv \frac{1}{V} \int_V \int_V [g_{\alpha\beta}(r_{12}) - 1] d\mathbf{r}_1 d\mathbf{r}_2 \equiv V \frac{\langle N_\alpha N_\beta \rangle - \langle N_\alpha \rangle \langle N_\beta \rangle}{\langle N_\alpha \rangle \langle N_\beta \rangle} - \frac{V \delta_{\alpha\beta}}{\langle N_\beta \rangle} \quad (1.18)$$

In the limit  $V \rightarrow \infty$  and for homogeneous conditions, the double integrals of Eq. (1.18) can be reduced to a single integral by applying the transformation:  $\mathbf{r}_2 \rightarrow \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ , which yields the original expression for the KB integral for infinitely large systems (Eqs. (1.3) and (1.4)). However, for a finite subvolume,  $V$ , applying this transformation is not possible since the domain of integration over

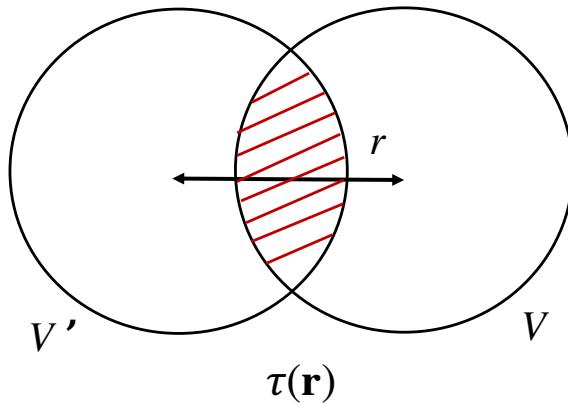


Figure 1.3: Schematic representation of the definition of the function  $\tau(\mathbf{r})$  (Eq. (1.22)), the overlap volume between two spheres separated by a distance  $r$ .

$\mathbf{r}$  depends on  $\mathbf{r}_1$ . In this case, the double volume integrals in Eq. (1.18) are reduced to a single radial integral by rewriting the left hand side (L.H.S) of Eq. (1.18) as

$$G_{\alpha\beta}^V = \int_0^\infty drc(r) w(r) (g_{\alpha\beta}(r) - 1) \quad (1.19)$$

where  $drc(r)$  is a hyperspherical volume element and  $w(r)$  is a purely geometric function characteristic of the volume  $V$  defined as:

$$w(r) \equiv \frac{1}{V} \int_V d\mathbf{r}_1 \int_V d\mathbf{r}_2 \delta(r - |\mathbf{r}_1 - \mathbf{r}_2|) \quad (1.20)$$

Once the function  $w(r)$  is known, the 2D dimensional integral of Eq. (1.18) reduces to the one-dimensional integral of Eq. (1.19), and the expression for KBIs for finite subvolumes is obtained.

For the calculation of  $w(r)$ , we first rewrite the L.H.S of Eq. (1.18) as

$$G_{\alpha\beta}^V = \frac{1}{V} \int_V d\mathbf{r} \tau(\mathbf{r}) (g_{\alpha\beta}(\mathbf{r}) - 1) \quad (1.21)$$

where the integral is over all of space and

$$\tau(\mathbf{r}) \equiv \int_V d\mathbf{r}_1 \int_V d\mathbf{r}_2 \delta(\mathbf{r} - \mathbf{r}_1 + \mathbf{r}_2) \quad (1.22)$$

The function  $\tau(\mathbf{r})$  has a simple geometrical interpretation: it is the overlap between the subvolume  $V$  and the same subvolume  $V$  shifted by  $\mathbf{r}$ . This may be seen by making the variable substitution  $\mathbf{r}'_2 = \mathbf{r}_2 + \mathbf{r}$  which yields  $\tau(\mathbf{r}) \equiv$

Table 1.1 Exact expressions of the geometrical function  $w(x)$  for hyperspheres in 1-3 dimensions (i.e. line, circle, and sphere) [74]. Here,  $x = r/L_{\max}$  ( $0 < x < 1$ ),  $L_{\max}$  is the maximum distance between two points, and  $c(r)dr$  is the hyperspherical volume element with thickness  $dr$ . For  $x \geq 1$ ,  $w(x) = 0$ .

Dimension	$c(r)$	$w(x)$
1D	2	$1 - x$
2D	$2\pi r$	$2/\pi(\arccos(x) - x\sqrt{1-x^2})$
3D	$4\pi r^2$	$1 - 3x/2 + x^3/2$

$\int_V d\mathbf{r}_1 \int_{V'} d\mathbf{r}'_2 \delta(\mathbf{r}'_2 - \mathbf{r}_1)$ , where  $V'$  is the subvolume  $V$  shifted by  $\mathbf{r}$  (see Figure 1.3). The function  $w(r)$  is obtained from  $\tau(\mathbf{r})$  by integrating over  $4\pi$  solid angle ( $\Omega$ ) and dividing by  $V$ . We have

$$w(r) = \frac{1}{V} \int d\mathbf{r}' \tau(\mathbf{r}') \delta(r - |\mathbf{r}'|) = \frac{r^{D-1}}{V} \int d\Omega \tau(\mathbf{r}) \quad (1.23)$$

where  $D$  is the dimensionality of space. In the following, we consider for  $V$  hyperspheres of radius  $R$ , where by symmetry, the overlap volume does not depend on  $\Omega$ , so  $\tau(\mathbf{r}) = \tau(r)$ . The volume of a hypersphere is  $V = R^D \int d\Omega/D$  which, together with Eq. (1.23), yields

$$w(r) = \frac{\tau(r) D r^{D-1}}{R^D} \quad (1.24)$$

The overlap volumes  $\tau(r)$  of hyperspheres in  $D=1-3$  dimensions (i.e. segment, circle and sphere) are well known [82]. From these, the corresponding functions  $w(r)$  are obtained using Eq. (1.24). The functions  $w(r)$  are computed up to the maximum distance between two points in a subvolume  $L_{\max}$ . It is convenient to define the dimensionless distance  $x = r/L_{\max}$ . The corresponding functions  $w(x)$  are listed in Table 1.1. Using  $w(x)$  and the L.H.S of Eq. (1.18), we arrive at the final expression for KBIs for finite subvolumes,

$$G_{\alpha\beta}^V = \int_0^{L_{\max}} [g_{\alpha\beta}(r) - 1] c(r) w(x) dr \quad (1.25)$$

where we have used the fact that  $w(x) = 0$  for  $x \geq 1$ .

KBIs computed from small subvolumes scale with the inverse size of the subvolumes. This scaling law can be explained by the concept of thermodynamics of small systems as mentioned earlier. Alternatively, Krüger et al. [74] showed that finite-size effects of the subvolume emerge from pairs of molecules  $\alpha - \beta$ , where particle  $\alpha$  is inside subvolume  $V$ , and particle  $\beta$  is outside  $V$  (the simulation box which contains  $V$  is denoted by  $L^3$ ). To account for the contribution

of these molecular pairs to the KBIs, the function  $Q_{\alpha\beta}$  is defined by splitting the integral domain in Eq. (1.18) over the surrounding  $\int_V \int_{L^3}$  and  $\int_V \int_{L^3 - V}$  (this is possible as RDFs have a finite range),

$$\begin{aligned} Q_{\alpha\beta} &\equiv \int_V d\mathbf{r}_1 \int_{L^3 - V} d\mathbf{r}_2 (g_{\alpha\beta}(r_{12}) - 1) \\ &= \int_V d\mathbf{r}_1 \int_{L^3} d\mathbf{r}_2 (g_{\alpha\beta}(r_{12}) - 1) - \int_V d\mathbf{r}_1 \int_V d\mathbf{r}_2 (g_{\alpha\beta}(r_{12}) - 1) \end{aligned} \quad (1.26)$$

where non-zero contributions to  $Q_{\alpha\beta}$  originate from molecule pairs where one molecule is inside  $V$  and the other one outside  $V$ . Assuming a finite correlation length  $\zeta$  for a layer surrounding the subvolume, we have  $[g_{\alpha\beta}(r_{12}) - 1] \approx 0$  for  $r_{12} > \zeta$ . The volume of this layer, and thus  $Q_{\alpha\beta}$ , increases linearly with the surface area  $A$  of the spherical subvolume (for a radius of the sphere much larger than  $\zeta$ ). In the case of an infinitely large system, the homogeneous conditions allow for the substitution  $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$  in the integral over  $L^3$ , resulting in:

$$\begin{aligned} Q_{\alpha\beta} &= \int_V d\mathbf{r}_1 \int_{L^3} d\mathbf{r} (g_{\alpha\beta}(r_{12}) - 1) - \int_V d\mathbf{r}_1 \int_V d\mathbf{r}_2 (g_{\alpha\beta}(r_{12}) - 1) \\ &\approx VG_{\alpha\beta}^\infty - VG_{\alpha\beta}^V \end{aligned} \quad (1.27)$$

where  $G_{\alpha\beta}^\infty$  is the KB integral for an infinite volume. As  $Q_{\alpha\beta}$  scales linearly with the surface area  $A$ , the difference  $(G_{\alpha\beta}^\infty - G_{\alpha\beta}^V)$  scales as  $A/V$ , i.e. inversely with the linear dimension of the subvolume ( $A/V \sim 1/L$ ),

$$G_{\alpha\beta}^V(L) = G_{\alpha\beta}^\infty + \frac{F_{\alpha\beta}^\infty}{L} \quad (1.28)$$

where  $F_{\alpha\beta}^\infty$  is a scaling constant proportional to the function  $Q_{\alpha\beta}$  defined earlier. From extrapolating  $G_{\alpha\beta}^V$  to  $1/L \rightarrow 0$ , KBIs in the thermodynamic limit are obtained.

### OTHER METHODS FOR COMPUTING KIRKWOOD–BUFF INTEGRALS

Similar to the method of Krüger and co-workers [74], the approach of Cortes-Huerto et al. [83] uses small subvolumes to estimate KBIs in the thermodynamic limit. In their approach, Cortes-Huerto et al. [83] apply a correction for RDFs that is independent of the interparticle distance. The methods of Krüger and co-workers [74, 80, 81, 84], and Cortes-Huerto et al. [83] provide practical approaches to computing KBIs for any isotropic fluid, while addressing system size effects and RDF-related finite-size effects. Other available methods for computing KBIs are more complicated, and found to be difficult to extend to systems

with internal degrees of freedom. Wedberg et al. [75, 85] presented a method for extending KBIs to the thermodynamic limit using Verlet's extension of RDFs [86]. The Verlet extension method [86] can be applied to estimate RDFs beyond the size of the finite simulation box, which are then used to extrapolate to KBIs to the thermodynamic limit, by truncating Eq. (1.3) to a value much larger than half the size of the simulation box. The approach of Wedberg et al. [75] was verified for pure Lennard Jones (LJ) and Stockmayer fluids. A drawback of this approach is the complexity of the numerical procedure as it required the effective interaction potential for two molecules at distance  $r$ . Moreover, it is not trivial to extend the method to systems of molecules with intramolecular degrees of freedom.

KBIs can be computed from molecular simulations of finite number of molecules using static structure factors [76, 87]. The structure factor of a liquid,  $S(q)$ , is related to the Fourier transform of pair distribution functions, and  $q$  is the magnitude of change of a reciprocal lattice vector [23]. Structure factors can be measured from scattering experiments, where  $q$  is a function of the wave length and the scattering angle. At the zero wavelength limit,  $q = 0$ , structure factors are directly related to KBIs [51, 88]. However, the values of  $S(q = 0)$  cannot be measured directly. Similarly, with molecular simulation, structure factors can be computed for a set of values of  $q$ , and then extrapolated to the limit  $q = 0$  to find KBIs. In the work of Nichols et al. [76], structure factors are computed from fluctuations in the number of molecules of finite systems. Each lattice vector  $q$  corresponds to a set of different sampling volumes, or sub-cells inside the simulation box, from which density fluctuations are computed. Rather than considering subvolumes formed by a central molecule, Nichols et al. [76] considered fluctuations in slab-like regions that resulted from dividing the simulation box. As a result, the whole volume is considered and all the information is used. From fluctuations, written as a 3D Fourier series, structure factors are computed and this was used to obtain the thermodynamic properties that relate to KBIs (partial deviates of chemical potential with respect to composition, partial molar volumes, and isothermal compressibilities). For a LJ fluid, Nichols et al. [76] found that it was difficult to extrapolate structure factors to  $q = 0$ . Instead, thermodynamic properties computed from subcells (i.e specific range of  $q$ ) were extrapolated to the limit  $q = 0$ . Extrapolation of thermodynamic properties is needed to remedy finite-size effects. While the method of Nichols et al. [76] provides accurate thermodynamic properties, compared to truncated KBIs, it is computationally involved even for systems with no intramolecular interactions. Structure factors were also used in the work of Rogers [87] to compute KBIs from simulations of closed and finite systems. As in the work of Nichols et al. [76], information from

the entire volume of the simulation box was used. However, both methods were applied to compute KBIs of systems of molecules with no intramolecular degrees of freedom such as LJ fluids.

## 1.4. APPLICATIONS OF KIRKWOOD–BUFF INTEGRALS COMPUTED FROM MOLECULAR SIMULATION

### 1.4.1. PARTIAL MOLAR ENTHALPIES

In Ref. [89], Schnell et al. proposed a method to compute partial molar enthalpies from molecular simulation in the canonical ensemble. Following the SSM [16], enthalpies  $\hat{H}$  of small subvolumes embedded in a larger reservoir are used. From nanothermodynamics, an expression for the change of  $\hat{H}$  with respect to the average number of molecules  $\langle N_\alpha \rangle$  was derived in terms of fluctuations in density and energy,

$$\left( \frac{\partial \hat{H}}{\partial \langle N_\alpha \rangle} \right)_{T, V, \mu_{\beta \neq \alpha}} = \frac{\langle EN_\alpha \rangle - \langle N_\alpha \rangle \langle E \rangle + \langle N_\alpha \rangle k_B T}{\langle N_\alpha^2 \rangle - \langle N_\alpha \rangle^2} \quad (1.29)$$

in which  $E$  is the partial energy of the subvolume. As shown in the previous section, properties of small subvolumes scale with the inverse size of the subvolume ( $1/L$ ). Extrapolating the derivatives of Eq. (1.29) to the thermodynamic limit yields partial enthalpies at constant volume  $\left( \frac{\partial H}{\partial N_\alpha} \right)_{T, V, N_{\beta \neq \alpha}}$ . To find partial molar enthalpies,  $\left( \frac{\partial H}{\partial N_\alpha} \right)_{T, P, N_{\beta \neq \alpha}}$ , a Legendre transform was performed. To convert from enthalpies in the canonical ensemble to partial molar enthalpies, KBIs of the studied system are needed. The method of Krüger and co-workers [74, 80, 81, 84] to obtain KBIs for finite subvolumes was used. This approach was also applied by Skorpa et al. [90] to compute the heat of reaction of  $H_2$  dissociation using a reactive force field.

### 1.4.2. PROPERTIES OF SINGLE-IONS IN SALT SOLUTIONS

Simulating closed and finite systems to compute KBIs has the advantage of accessing single-ion properties [91]. Essentially, to apply the KB theory to a salt solution, the system has to be treated as a binary mixture where ions are indistinguishable [13], as shown in Ref [20]. In this case, relations between KBIs and thermodynamic properties of binary mixtures can be applied (e.g. see section 1.1). For a ternary mixture of a dissociating monovalent substance ( $AB \rightarrow A + B$ ) and a solvent (e.g. water,  $W$ ), KBIs are subject to the following electroneutrality conditions,

$$\rho G_{WA} = \rho G_{WB} \quad (1.30)$$

$$1 + \rho G_{AA} = \rho G_{AB} \quad (1.31)$$

$$1 + \rho G_{BB} = \rho G_{AB} \quad (1.32)$$

where  $\rho$  is the number density of the salt ( $\rho_A = \rho_B = \rho$ ). Eqs. (1.30), (1.31) and (1.32) imply that the number of molecules of species  $A$  and  $B$  cannot be varied independently. Ben-Naim [13] showed that the above constraints introduce a singularity to the equations relating KBIs  $G_{\alpha\beta}^\infty$  to thermodynamic quantities. It is important to note that the KB theory is general for any type of interactions and the issue of singularity is not due to the strong electrostatic interactions present in salt solutions. Rather, it is a result of the closure constraints imposed by Eqs. (1.30), (1.31) and (1.32), and it does not apply to KBIs defined in open systems [12]. Eqs. (1.30), (1.31) and (1.32) hold for any dissociating molecule  $AB$  where the number of molecules has to be conserved simultaneously in the system, i.e.  $N_A = N_B$ . The approach of using KBIs of finite subvolumes of Krüger and co-workers [74, 80, 81, 84] allows KBIs of single ions to be computed from simulations in the canonical ensemble with open subvolumes embedded in the simulation box. As a result, the charge neutrality of the reservoir is maintained ( $N_A = N_B$ ), while the electroneutrality condition is not applied inside the subvolume, and therefore the grand-canonical ensemble is accessed. In the work of Schnell et al. [91], KBIs of a sodium chloride (NaCl) solution were computed as well as partial molar volumes of water,  $\text{Na}^+$ , and  $\text{Cl}^-$ . The partial molar volume of  $\text{Na}^+$  was reported to have a negative value [91]. In Ref. [7], a similar observation was reported when computing the partial molar volumes of  $\text{Na}^+$  and  $\text{Cl}^-$ . The authors of Ref. [7] investigated the possibility of computing single-ion properties using molecular simulations by considering two methods. The first method is based on the changes in average potential energy and box volume when inserting an ion into a pure liquid. The second method depends on computing the reversible work associated with inserting an ion into a liquid.

#### 1.4.3. MASS TRANSFER IN MULTICOMPONENT LIQUIDS

KBIs computed from molecular simulation can be applied to connect Fick diffusion coefficients to MS diffusivities [70, 92]. The generalized Fick's law relates the molar flux,  $J_i$ , to the Fick diffusivity,  $D_{ij}$  [66, 68],

$$J_i = -c_t \sum_{j=1}^{n-1} D_{ij} \nabla x_j \quad (1.33)$$

where  $c_t$  is the total molar concentration and  $\nabla x_j$  is the gradient of the mole fraction  $x_j$ , which is the driving force in Fick's law. In a molar reference frame, we have  $\sum_{i=1}^n J_i = 0$ . MS diffusivities can be computed from MD simulations and Fick diffusivities can be measured by experiments [64, 66, 68, 70, 92]. The MS diffusivity  $D_{ij}$  can be considered as an inverse friction term in an equation where the gradient in chemical potential  $\nabla \mu_i$  of component  $i$  is related to differences in the average velocities between species:

$$-\frac{1}{RT} \nabla \mu_i = \sum_{j=1(j \neq i)}^n \frac{x_j(u_i - u_j)}{D_{ij}} \quad (1.34)$$

where  $(u_i - u_j)$  is the difference between the average velocities of components  $i$  and  $j$ . As chemical potentials cannot be measured directly, it is not possible to directly compare MS diffusivities to experiments. It is more convenient to compute MS diffusivities using molecular simulation. Details on this are provided in Refs. [70, 73, 92]. Often, Fick diffusivities depend more strongly on concentration than MS diffusivities [64, 68]. Moreover, it is possible to predict diffusion of multicomponent mixtures ( $n > 2$ ) from the knowledge of MS diffusivities of the corresponding binary mixtures [64, 69, 70]. For a mixture with more than two components, Fick diffusivities depend on the type of reference frame, unlike MS diffusivities [64, 68, 93]. In a molar reference frame, Fick diffusivities and the thermodynamic factor can be used to compare MS diffusivities with experimental data [66],

$$[D] = [B]^{-1} [\Gamma] \quad (1.35)$$

where  $[D]$  is the Fick diffusion coefficients matrix. The elements of the matrix  $[B]$  can be computed using

$$B_{ii} = \frac{x_i}{D_{in}} + \sum_{j=1(j \neq i)}^n \frac{x_j}{D_{ij}} \quad \text{with } i = 1, 2, \dots, (n-1) \quad (1.36)$$

$$B_{ij} = -x_i \left( \frac{1}{D_{ij}} - \frac{1}{D_{in}} \right) \quad \text{with } i = 1, 2, \dots, (n-1) \quad \text{and } i \neq j \quad (1.37)$$

and the elements of matrix  $[\Gamma]$  can be expressed as a function of KBIs. For binary systems, the relation between  $\Gamma_{ij}$  and KBIs is provided by Eq. (1.14). In Refs. [70] and [64], KBIs from simulations of finite systems were computed using the approach of Krüger and co-workers [74, 80, 81, 84]. KBIs of binary and ternary mixtures were used to compute the thermodynamic factors and convert MS diffusivities computed from simulations to Fick diffusivities measured by experiments. The proposed method was applied to binary and ternary alcohol mixtures [64, 70].

In Refs. [72, 73, 94], it was shown that KBIs can be used to correct finite-size effects of computed MS diffusion coefficients. MS diffusion coefficients depend on the size of the simulated system. These finite-size effects originate from hydrodynamic interactions [72, 95, 96]. In the studies by Jamali et al. [72, 94], a correction based on viscosity and the thermodynamic factor was used to compensate for this effect. For binary and ternary systems, KBIs were obtained from molecular simulation and used to compute thermodynamic factors. The finite-size correction was validated for various molecular systems such as organic fluids. Jamali et al. [72, 94] found the finite-size effects of MS diffusivities to be significant, especially when mixtures are close to demixing, or when molecules are large compared to the size of the simulation box [97, 98].

#### 1.4.4. OTHER APPLICATIONS

In section 1.1, we presented the relations that link KBIs to partial derivatives of the chemical potential with respect to the number of molecules (Eq. (1.5)), partial molar volumes (Eq. (1.6)), and isothermal compressibilities (Eq. (1.7)) for binary systems. Based on these relations, other properties can be estimated from KBIs. Galata et al. [54] used the KB theory to compute thermodynamic mixing properties and excess properties of liquid mixtures. In their work, the authors focus on computing partial derivatives of the chemical potential with composition and the Gibbs energy of mixing,  $\Delta_{\text{mix}}G$ , which are important quantities for the prediction of phase equilibria of liquid mixtures. The prediction of  $\Delta_{\text{mix}}G$  and other mixing properties from KBIs was validated using binary ideal and real LJ mixtures [54]. The KBIs were found using simulations of finite volumes, and finite-size effects were corrected using the approach of Cortes-Huerto et al. [83].

KBIs can be used to interpret findings from simulations of biological molecules. In Ref. [99], Pierce et al. presented a review of the applications of the KB theory to biological systems. One of the valuable applications of the KB theory is to study the effects of co-solvents on biomolecules. Molecular simulation provide local information on the co-solvents surrounding biomolecules and how such an environment affects the structure of biomolecules [99–103]. In 2004, Smith [102] demonstrated how KBIs can be used to relate simulation results, which provide preferential interaction, to macroscopic thermodynamic data [104]. Other than studying solvents surrounding biomolecules, the KB theory can be applied directly to systems with interacting biomolecules. However, this application can be hindered by difficulties associated with sampling the phase space of such systems.

KBIs can be used for the development and parameterization of force fields [105–107]. Weerasinghe and Smith provide KB derived force fields (KBFFs) for a number of mixtures, such as sodium chloride in water [106], urea and water [108], acetone and water [109], and methanol and water [110]. The force fields were parameterized so that KBIs obtained from experimental data are reproduced. More on the use of experimentally obtained KBIs are provided in section 1.2. The authors of Refs. [105–107] found that macroscopic properties can be accurately computed using KBFFs, while addressing solute-solute and solvent-solute molecular structure of the systems considered. For instance, in Ref. [105] the derived KBFFs was able to reproduce microstructure of alkaline earth halide salts in water. Ion-ion and ion-water distances provided by the force field were found to agree with those measured by neutron scattering experiments. The same KB force field yielded satisfactory predictions of several macroscopic quantities such as partial molar volumes, and partial derivatives of chemical potential with respect to density. Mijaković et al. [111] compared several force fields, including KBFFs, for ethanol-water mixtures. The authors reported that the KB derived force field performed better than other force fields when computing KBIs and several thermodynamic properties such as excess volumes, excess enthalpy, and self-diffusion coefficients.

## 1.5. SCOPE OF THIS THESIS

The focus of this thesis is to provide methods to accurately compute KBIs from molecular simulation. In this way, one does not have to rely on experimental data or insertion and deletion of molecules. To successfully estimate KBIs from molecular simulation using Eq. (1.25), the following effects have to be studied:

1. Shape effects related to the subvolume embedded in the system.
2. Finite-size effects related to the size of the system used to compute KBIs. To obtain of KBIs that converge to the correct value in the thermodynamic limit, a sufficient number of molecules has to be simulated.
3. Finite-size effects related to the sizes of the subvolumes embedded in the system.
4. Finite-size effects related to the computed RDFs. KBIs should be computed from RDFs of open systems, while nearly always, RDFs are computed from closed systems.

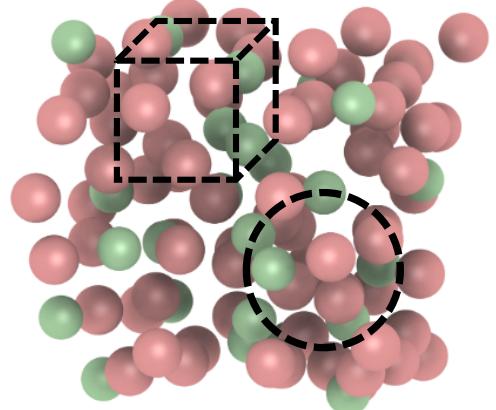
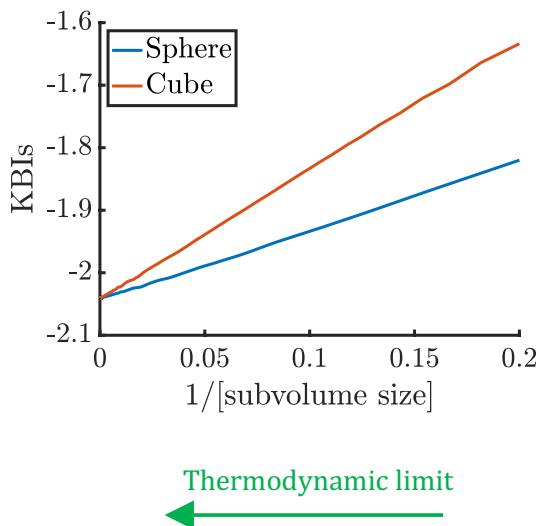
In this thesis, we develop a framework to accurately compute KBIs based on the method of Krüger and co-workers [74]. In chapters 2 and 3, the shape and size effects related to computing KBIs using small subvolumes are studied. In

chapter 2, a numerical method is developed to compute KBIs for subvolumes with arbitrary shapes. We show that truncating KBIs (i.e. truncating the integral of Eq. (1.3)) is nonphysical. In chapter 3, an analytic RDF was used to investigate finite-size effects of the system and the subvolumes. RDF-related effects were studied by simulating Weeks-Chandler-Andersen (WCA) particles [112]. RDFs of systems with different sizes were corrected using three RDF correction methods. When using finite systems, the method of Gaunly and van der Vegt [113] resulted in the best estimation of KBIs. Additionally, in chapter 3 it is shown that the accuracy may be affected by how KBIs of small subvolumes are extrapolated to the thermodynamic limit. If the size of the simulated system is sufficient, KBIs of small subvolumes should scale linearly with the inverse size of the subvolume. However, identifying a linear regime is not straightforward. In chapter 4, other extrapolation techniques are considered such as an expression to directly compute KBIs in the thermodynamic limit from RDFs of finite volumes and the scaling of  $LG_{\alpha\beta}^V$  with  $L$ . The scaling of KBIs depends on the size of the subvolume and a term related to surface effects. In chapter 4, different methods are used to extrapolate to KBIs in the thermodynamic limit and to compute the surface term for LJ and WCA fluids. From chapters 3 and 4, a practical method to compute KBIs using molecular simulations is established. In chapter 5, these findings are used to compute KBIs of the Deep Eutectic Solvent (DES) consisting of choline chloride and urea at various compositions. The computed KBIs are then used to obtain thermodynamic factors and partial molar volumes. The thermodynamic factors were used to examine interactions of choline chloride-urea mixtures and connect MS diffusivities with Fick diffusion coefficients. This clearly shows that the KB theory is useful for computing thermodynamic properties while avoiding insertion/deletion of molecules. However, in some cases, it is more convenient to perform molecular simulations in open ensemble. For instance, when computing the solubility of small gas molecules in liquid solvents. An example of this is presented in chapter 6, where the CFCMC method was used to predict the solubility of a number of gases in monoethylene glycol (MEG).



# 2

## SHAPE EFFECTS OF KIRKWOOD-BUFF INTEGRALS OF FINITE SYSTEMS



---

This chapter is based on the paper: N. Dawass, P. Krüger, J. M. Simon, and T. J. H. Vlugt, *Kirkwood-Buff integrals of finite systems: Shape effects*, Molecular Physics, **116**, 1573-1580 (2018) [84]

## 2.1. INTRODUCTION

In chapter 1, the method of Krüger and co-workers [74] was presented, where KBIs of finite subvolumes are used to estimate KBIs in the thermodynamic limit. Considering the shape of the subvolume, the sphere is the natural choice in simple liquids, but other shapes may be more convenient for specific applications. For example, the KB theory was previously applied to study the interactions between large biomolecules and the surrounding solvent molecules [99, 114, 115]. Giambasu et al. [116] used KBIs to study the ionic atmosphere surrounding nucleic acids. In their work, selecting the shape of the subvolume depended on the inhomogeneous region surrounding the nucleic acids [116]. For instance, hexagonal prisms were used to study the fluctuations of solvent molecules around DNA. In principle, it is possible to compute KBIs using the right hand side (R.H.S) of Eq. (1.18) for any shape of the subvolume  $V$ . The size of the subvolume can be gradually increased as shown in Figure 1.2, and the number of particles in each subvolume is then used to compute  $G_{\alpha\beta}^V$  using the R.H.S of Eq. (1.18). Cubic subvolumes have been used in the works of Schnell et al. [16], Cortes-Huerto et al. [83] and others [70, 113, 117] in combination with the R.H.S of Eq. (1.18). The advantage of using cubic subvolumes is that one does not need to compute distances between molecules and the center of the subvolume. The alternative formulation of finite-size KBIs (Eq. (1.25)), i.e. direct integration over the RDF, has only been applied to spherical subvolumes [74]. It is important to note that Eq. (1.25) is valid for subvolumes of any shape, provided the geometrical function  $w(x)$  is known for that shape.

The objective of this chapter is to present a unified framework for computing KBIs for subvolumes of arbitrary convex shape. We provide a numerical method to compute the function  $w(x)$  based on umbrella sampling MC. Once the function  $w(x)$  is computed for a specific shape, it can be used for any size of the subvolume. We compute the function  $w(x)$  for the following shapes: square, cube, and spheroids and cuboids with different aspect ratios. Numerical tables of these functions are provided in a data repository (see Ref. [118]). We also investigate the effect of the shape of the subvolume on the computation of KBIs. We will show that using a cubic or spherical subvolume leads to the same KB integral in the thermodynamic limit, and that for large subvolumes KBIs scale as area over volume, independent of the shape of the subvolume. This scaling will also explain why truncation of KBIs (i.e. truncating the integral of Eq. (1.3)) leads to nonphysical results.

The chapter is organized as follows. In section 2.2, the numerical method used to compute  $w(x)$  is introduced. The method is verified by comparing our numerical results to the analytic expressions for a sphere (3D), circle (2D), and line (1D). In section 2.3.2, the function  $w(x)$  is computed numerically for a cube

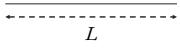
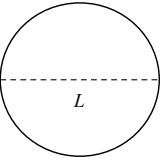
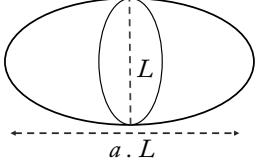
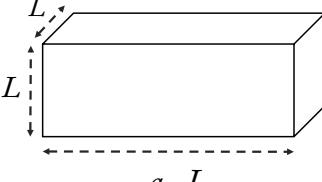
and for spheroids and cuboids with different aspect ratios. From this, finite volume KBIs for a liquid with a model RDF are computed for various shapes, both in 2D and 3D. Our findings are summarized in section 2.4.

## 2.2. NUMERICAL COMPUTATION OF $w(x)$

In this section, we present a numerical method to compute the function  $w(x)$  for convex subvolumes. Table 2.1 shows a schematic representation of the shapes studied here. For cuboids and spheroids,  $w(x)$  depends on the aspect ratio,  $a$ , and the function  $w(x)$  is computed for each  $a$ .

To find  $w(x)$ , we first compute  $w(r)$  and then normalize the distance  $r$  using the maximum distance between two points in the subvolume,  $L_{\max}$  (see Table 2.1), so  $x = r/L_{\max}$ . The function  $w(r)$  is proportional to the probability distribution function  $p(r)$  for finding two points inside the subvolume,  $V$ , separated by distance  $r$  [74]. Therefore, by construction we obtain  $w(r=0) = 1$  and  $w(r=L_{\max}) = 0$ , so consequently  $w(x=0) = 1$  and  $w(x=1) = 0$ . To compute the probability distribution function  $p(r)$  numerically, distances between two points inside the subvolume are divided into  $N$  bins ( $i_1, i_2, \dots, i_N$ ) of equal sizes, separated by  $\Delta r$ . Each bin contains all distances between  $i\Delta r$  and  $(i-1)\Delta r$ . As a result of this discretization, we sample the probability distribution  $p(i)$ , which is then used to compute  $w(i)$ . The value of  $\Delta r$  has to be chosen such that the function  $p(i)$  is properly sampled. We find that a small value of  $\Delta r$  results in poor statistics, especially in the first few bins. We recommend setting  $\Delta r$  to  $L/100$  ( $L$  is the characteristic length of the subvolume, see Table 2.1). To further improve the statistics, umbrella sampling [22, 119] is implemented for computing  $w(x)$ . This introduces a weightfunction  $W(i)$  which modifies the sampled distribution of distances.

Table 2.1 Subvolume shapes considered in this work.  $L$  is the characteristic linear dimension of the shape and  $L_{\max}$  is the largest possible distance between two points inside the subvolume. A cube and sphere are included as special cases of cuboid and spheroid with aspect ratio  $a = 1$ .

Shape of the subvolume		$L_{\max}$
Line (1D)		$L$
Circle (2D)		$L = 2 * R$
Square (2D)		$\sqrt{2}L$
Spheroid (3D)		$\begin{cases} L * a & a \geq 1 \\ L & a < 1 \end{cases}$
Cuboid (3D)		$L\sqrt{2 + a^2}$

### 2.2.1. IMPORTANCE SAMPLING ALGORITHM FOR COMPUTING $p(i)$

In the algorithm below, we show how the probability distribution function  $p(i)$  and the weight function  $W(i)$  are computed. Note that the algorithm presented in this work is for a 3D subvolume, however, it is trivial to adjust it to other dimensions. The algorithm follows the following steps:

1. Set  $\Delta r$  and the maximum allowed displacement for random displacements.
2. Set a weight function  $W(i)$  to zero for all bins.
3. Choose two random points ( $P_1$  and  $P_2$ ) inside the subvolume  $V$ .
4. For each sampling cycle (we typically performed  $10^{11}$  cycles):
  - (a) Select a single point,  $P_1$  or  $P_2$ , randomly. Assume that  $P_i$  is selected (the other point is denoted by  $P_j$ ).
  - (b) Give a random displacement to point  $P_i$  leading to  $P_{\text{new}}$ .
  - (c) Check if this new position falls inside the subvolume. If it does not, skip to step (f), otherwise carry on with the next step (d).
  - (d) Determine the normalized distance,  $r/L_{\max}$ , between  $P_{\text{new}}$  and  $P_j$  and determine the bin number corresponding to this distance,  $i_{\text{new}}$ . The bin number corresponding to the old distance is denoted by  $i_{\text{old}}$ .
  - (e) Accept the displacement if a uniformly distributed random number between 0 and 1 is less than  $\exp[W(i_{\text{new}}) - W(i_{\text{old}})]$ , otherwise the displacement is rejected. If the displacement move is accepted, update  $P_i$  and  $i_{\text{old}}$  such that  $P_i = P_{\text{new}}$  and  $i_{\text{old}} = i_{\text{new}}$ .
  - (f) Compute the normalized distance between  $P_i$  and  $P_j$  and the bin number,  $i$  corresponding to that distance. Update the sampling of the observed probability distribution function  $p_{\text{biased}}(i)$ .
5. After a large number of cycles, remove the bias caused by the weightfunction:

$$p(i) = p_{\text{biased}}(i) \exp[-W(i_{\text{old}})] \quad (2.1)$$

6. Update and save  $W(i)$  for the consecutive computations of  $p(i)$  using an iterative updating scheme:

$$W(i) \rightarrow W(i) - (1/2) \ln p_{\text{biased}}(i) \quad (2.2)$$

and shift  $W(i)$  so that its minimum equals zero.

7. Repeat steps 1 to 4 while updating  $W(i)$  until a satisfactory sampling of  $p(i)$  is reached. For  $10^{11}$  cycles, running the algorithm takes approximately 150 minutes on a modern computer.

**2**

### 2.2.2. COMPUTING $w(x)$

The function  $w(i)$  is proportional to the distribution function  $p(i)$  divided by the volume of the bin in a hypersphere with dimension  $D$ :

$$w(i) = \frac{p(i)}{i^D - (i-1)^D} \quad (2.3)$$

In Eq. (2.3), the prefactors for the bin volumes are not included yet since in the next step  $w(i)$  is normalized using the known value  $w(0) = 1$ . Since we do not obtain statistics exactly at  $r = 0$ , we interpolate to  $w(0)$  using the first two bins  $w(1)$  and  $w(2)$ ,

$$w(i) \rightarrow \frac{w(i)}{w(1) - \frac{(w(2)-w(1))}{2}} \quad (2.4)$$

Similarly, the distances are normalized to find the relative distance  $x$ ,

$$x(i) = \frac{(i-1/2)\Delta r}{L_{\max}} \quad (2.5)$$

Interpolation can be used to find  $w(x)$  for any value of  $x$ .

## 2.3. SHAPE EFFECTS OF KIRKWOOD–BUFF INTEGRALS

### 2.3.1. THE FUNCTIONS $w(x)$ FOR A CUBE, CUBOIDS, AND SPHEROIDS

In this section, we present the function  $w(x)$ , computed numerically for different shapes. To validate our numerical method (section 2.2), we compute the function  $w(x)$  for subvolumes where the analytic expressions are known (line, circle, and sphere, see Table 1.1) and make a comparison. In Figure 2.1 (a), the comparison between analytic and numerical functions  $w(x)$  is shown for a line, circle, and sphere. For these shapes, the numerical results reproduce the theoretical solution very well. The average absolute difference between analytic and numerical values are  $9 \times 10^{-3}$ ,  $5 \times 10^{-3}$ , and  $2 \times 10^{-4}$  for a sphere, circle, and line, respectively. Therefore, the algorithm of section 2.2 can be used to numerically compute the function  $w(x)$  for any convex subvolume in 1D, 2D, or 3D.

We compute the function  $w(x)$  numerically for subvolumes where analytic expressions for  $w(x)$  are not available. In Figure 2.1 (b), we show the function  $w(x)$  computed numerically for a cube and sphere, which are the most commonly used shapes for the subvolume when computing KBIs . Figure 2.2 shows

the function  $w(x)$  for spheroids and cuboids with the following aspect ratios,  $a = 0.1, 1, 2, 5$ , and  $10$ . Clearly, the function  $w(x)$  varies a lot with the aspect ratio  $a$ , and this function is very different for a sphere and a cube. We found that it is difficult to accurately fit  $w(x, a)$  with polynomial functions. In a data repository (see Ref. [118]), we provide tabulated data of the function  $w(x)$  for various shapes of the subvolumes.

2

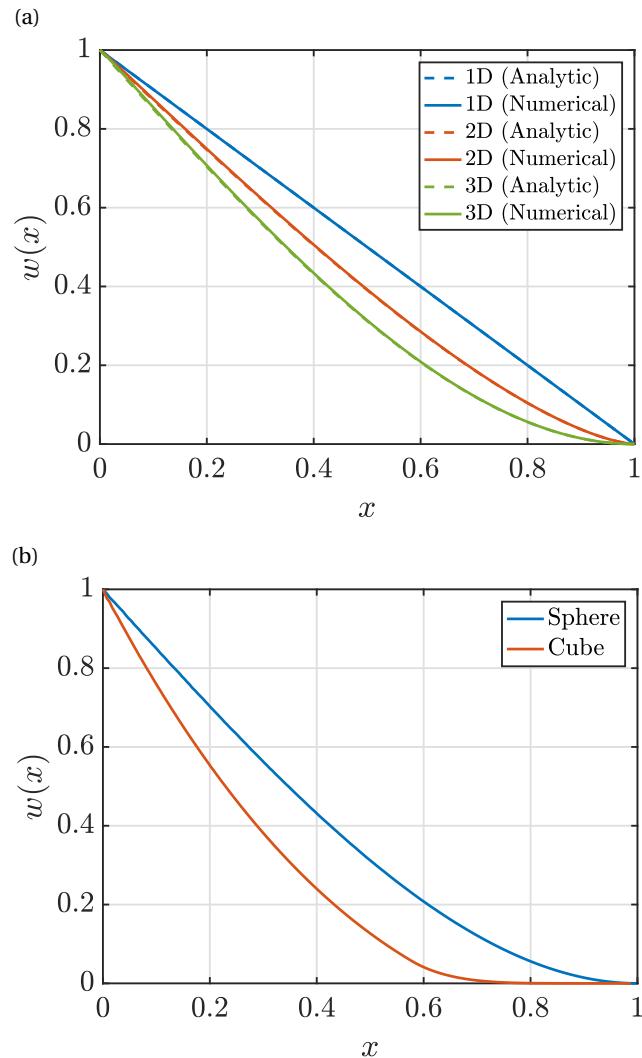


Figure 2.1: (a) The function  $w(x)$  for a line (1D), circle (2D), and a sphere (3D). The function  $w(x)$  is computed numerically using the MC algorithm provided in section 2.2 and the analytic functions are listed in Table 1.1. In all cases, the numerical solution matches the theoretical expressions of  $w(x)$ . (b) The function  $w(x)$  for a sphere and a cube.

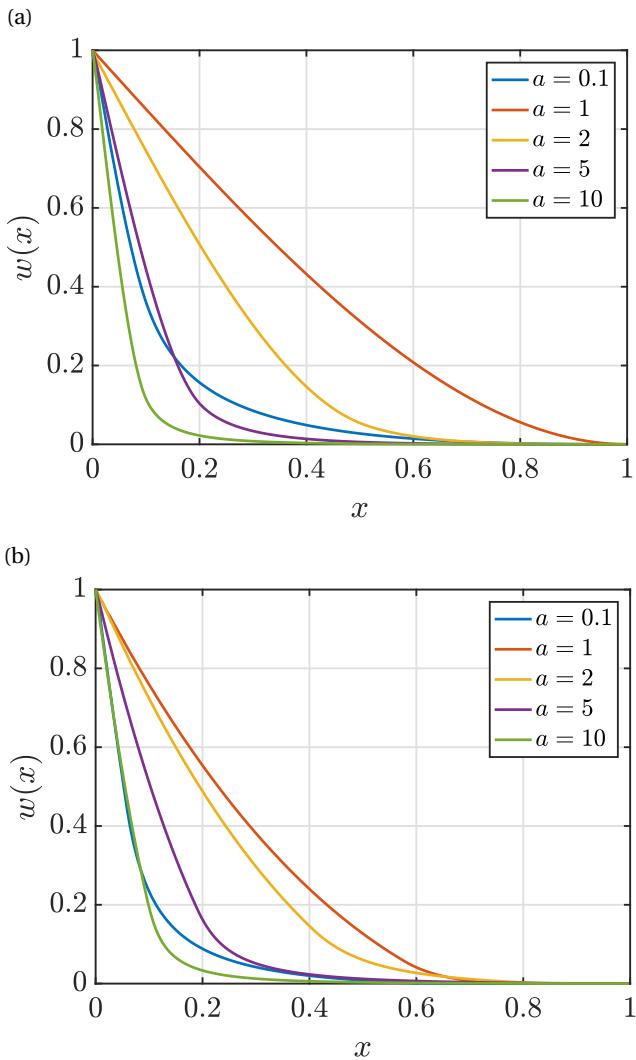


Figure 2.2: The function  $w(x)$  computed numerically using the MC algorithm provided in section 2.2, for (a) spheroids, and (b) cuboids with different aspect ratios  $a$ .

### 2.3.2. KIRKWOOD–BUFF INTEGRALS FOR VARIOUS SHAPES

We compute KBIs for subvolumes with different geometries. The expression for KBIs of finite subvolumes,  $G_{\alpha\beta}^V$ , is provided by Eq. (1.25). To investigate shape effects, a liquid with the following analytic RDF model [86, 120] is used,

$$g(r) - 1 = \begin{cases} \frac{3/2}{r/\sigma} \exp\left[\frac{1-r/\sigma}{\chi}\right] \cos\left[2\pi\left(\frac{r}{\sigma} - \frac{21}{20}\right)\right] & \frac{r}{\sigma} \geq \frac{19}{20}, \\ -1, & \frac{r}{\sigma} < \frac{19}{20} \end{cases} \quad (2.6)$$

where  $\sigma$  is the diameter of the particles, and  $\chi$  is the length scale at which the fluctuations of the RDF decay. This RDF mimics density fluctuations around a central particle for a typical isotropic liquid. The RDF parameters are fixed at  $\sigma = 1$  and  $\chi = 2$ . Here, we work with a single-component fluid therefore the indices  $\alpha$  and  $\beta$  are dropped. The use of an analytic  $g(r)$  eliminates errors due to uncertainties in numerically obtained RDFs [74, 113]. The functions  $w(x)$  are obtained numerically in tabulated form, and the value of  $w(x)$  at any  $x$  is obtained by interpolation. The integral of Eq. (1.25) is obtained by numerical integration using the trapezoidal rule [121].

In Figure 2.3, we show the KBIs for finite subvolumes,  $G^V/\sigma^3$ , plotted as a function of the inverse of the length of the subvolumes,  $\sigma/L$ . Figure 2.3 (a) shows the KBIs computed for spheroids with different aspect ratios ( $a = 1, 2, 5, 10$ ), and Figure 2.3 (b) shows the same for cuboids. In Figure 2.3 (b), we use analytic and numerical functions  $w(x)$  for spherical subvolumes ( $a = 1$ ). Integrating using the analytic or numerical functions  $w(x)$  yields practically identical values of the KBIs, and differences are of the same order as the error introduced by the numerical integration of Eq. (1.25). Changing the aspect ratio affects the slope of the lines of  $G^V/\sigma^3$  versus  $\sigma/L$ . All lines approach the same value of the KB integral in the limit  $\sigma/L \rightarrow 0$ , which is expected as in the thermodynamic limit the KB integral should be independent of the shape of the subvolume. Figure 2.3 shows that the shape of the subvolume affects the slope of the plots of  $G^V/\sigma^3$  versus  $\sigma/L$ . The dependence of the slope on the shape of the finite subvolume was previously reported in the work of Strøm et al. [122] using arguments based on small-scale thermodynamics. These authors found that plotting KBIs as a function of the surface to volume ratio of the subvolume should eliminate shape effects.

In Figure 2.4 (a), we show the KBIs plotted as a function of the surface area to volume ratio ( $A\sigma/V$ ) of the subvolume for the following shapes: sphere, cube, spheroid with  $a = 2$ , and cuboid with  $a = 2$ . As expected from the work by Strøm et al. [122], all KBIs approach the same value of  $G^\infty/\sigma^3$  with the same slope. In the thermodynamic limit,  $G^V/\sigma^3$  seems to only be a function of  $A/V$ . For large subvolumes, values of  $w(r)$  at small distances are more significant when inte-

grating  $G^\infty/\sigma^3$  (Eq. (1.25)). Therefore,  $w(r)$  has a universal behavior at small distances which leads to the universality of KBIs in Figure 2.4 (a). Using a Taylor expansion around  $r = 0$ , one can show that  $w(r)$  can be written as function of  $A/V$  [81],

$$w(r) \approx \left(1 - \frac{r}{4} \frac{A}{V} + \mathcal{O}(r^2)\right) \quad (2.7)$$

In Figure 2.4 (b), we plot the ratio of  $w(r)$  using Eq. (2.7) to numerical  $w(r)$  for a sphere, cube, spheroid with  $a = 2$ , and cuboid with  $a = 2$ . A subvolume with  $L = 1$  is used. At small distances ( $r < 0.1$ ), the ratio is practically 1 for all shapes considered. Eq. (2.7) shows that the function  $w(r)$  depends on the size  $r$  and the ratio  $A/V$ . The shape contribution originates from the term  $\mathcal{O}(r^2)$ . Therefore, properties of large subvolumes are independent of shape. This is referred to as the so-called shape thermodynamics limit, where properties of the subvolume are dependent on the size but not the shape of the subvolume [123]. In the conventional thermodynamic limit, properties are independent of both size and shape of the subvolume.

It is important to note that Eq. (2.7) provides a physical reason for the poor convergence of truncated KBIs (i.e. truncating the integral of Eq. (1.3)). If we consider a subvolume  $V$  with zero surface area  $A = 0$ , this will yield the weight function  $c(r)w(r) = 4\pi r^2$ . Substituting this in the expression of KBIs of finite subvolumes (Eq. (1.25)), one arrives at an expression of KBIs truncated to finite distances,

$$\hat{G}_{\alpha\beta}(R) = \int_0^R dr 4\pi r^2 \left[ g_{\alpha\beta}^\infty(r) - 1 \right] \quad (2.8)$$

Therefore, truncated KBIs correspond to the nonphysical case of finite-size KBIs with subvolumes  $V$  and zero surface area.

We verified numerically that findings from computing KBIs using different 3D shapes apply to KBIs computed using 2D shapes as well. In Figure 2.5, we show  $G^V/\sigma^2$  vs.  $\sigma/L$  for a circle and a square. For a circle, we compute the KBIs using the analytic function for  $w(x)$ , and also using the numerically function  $w(x)$ . Both functions result in identical values of KBIs for all sizes of the subvolume. Using circle or square for the shape of the subvolume results in KBIs that converge to the same value of  $G^\infty/\sigma^2$ .

2

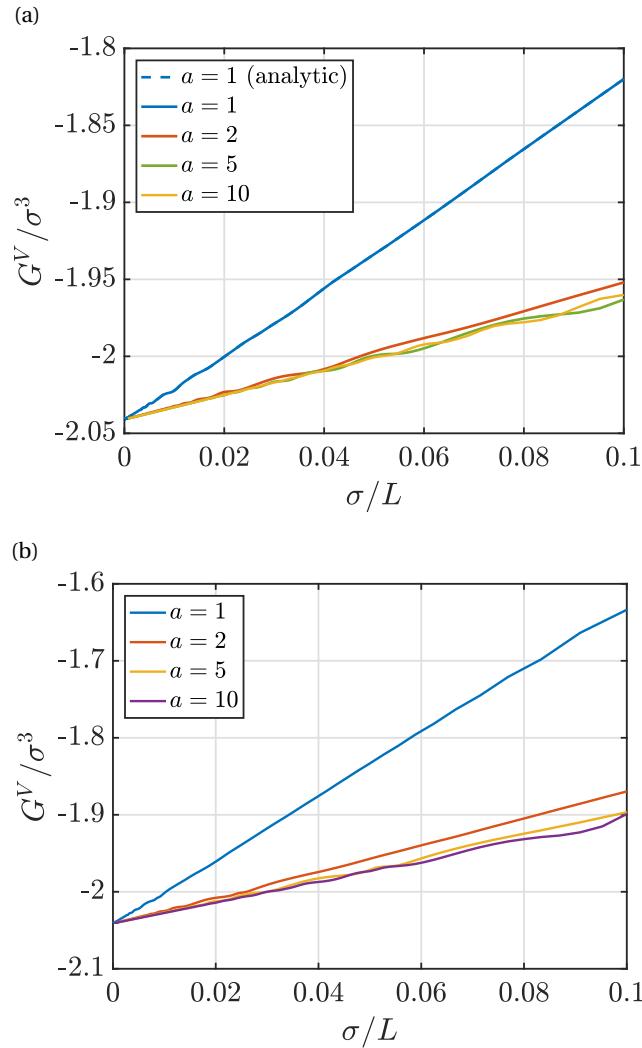


Figure 2.3: KBIs of finite subvolumes  $G^V / \sigma^3$  vs. the inverse of the size of the subvolume,  $\sigma / L$  using the numerically computed  $w(r)$ . The RDF of Eq. (2.6) has parameters  $\sigma = 1$  and  $\chi = 2$ . The subvolumes used have the following shapes: (a) spheroids and (b) cuboids with different aspect ratios,  $a$ . In (a) we also compare the KB integral for a sphere ( $a = 1$ ) using both the numerical  $w(x)$ , and the analytic  $w(x)$  from Table 1.1.

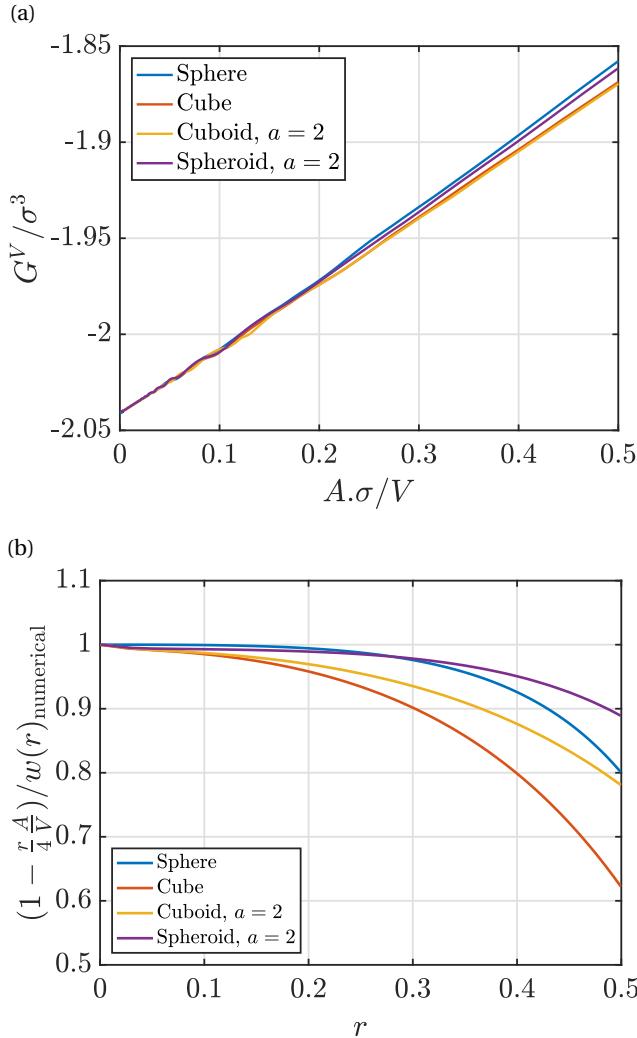


Figure 2.4: (a) KBIs of finite subvolumes,  $G^V / \sigma^3$  vs. the inverse of the surface area to volume ratio of the subvolume,  $A.\sigma/V$ . The KBIs are found by numerically integrating  $G_{\alpha\beta}^V$  for 3D subvolumes (Table 1.1). The RDF of Eq. (2.6) has parameters  $\sigma = 1$  and  $\chi = 2$ . (b) The ratio of  $w(r)$  computed using Eq. (2.7) to numerical  $w(r)$  of subvolumes with  $L = 1$  and different shapes.

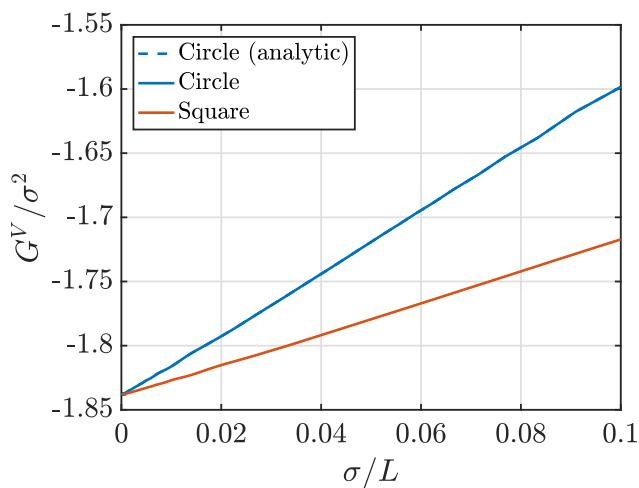


Figure 2.5 KBIs of finite subvolumes,  $G^V/\sigma^2$ , vs. the inverse of the size of the subvolume,  $\sigma/L$  using the numerically obtained  $w(r)$ . KBIs are obtained by numerically integrating Eq. (1.25) for 2D subvolumes (Table 1.1). The RDF of Eq. (2.6) has parameters  $\sigma = 1$  and  $\chi = 2$ . For the circle, the analytic expression of the function  $w(x)$  can be found in Table 1.1.

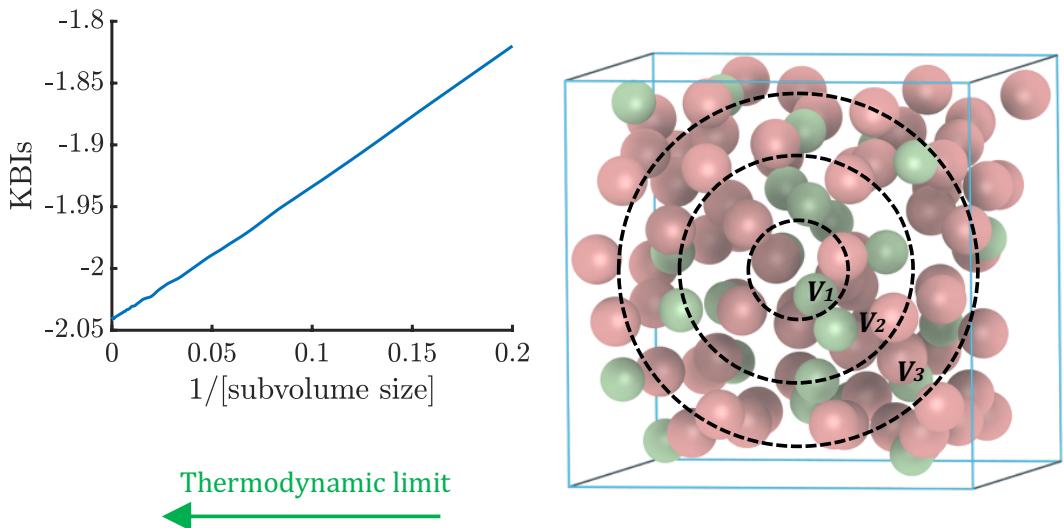
## 2.4. CONCLUSIONS

We have introduced a method for computing KBIs for finite subvolumes of arbitrary convex shapes. This requires a numerical method to obtain the geometrical function  $w(x)$ , which is needed when computing KBIs from RDFs. We showed that  $w(x)$  is related to the probability of finding two particles inside a subvolume  $V$  at a certain distance, and we presented a numerical scheme based on umbrella sampling MC for this. The numerical method was verified by comparing the results with analytic expressions for hyperspherical subvolumes in 1D (line), 2D (circle), and 3D (sphere). The method was used to compute the function  $w(x)$  for the following subvolumes: square, cube, and spheroids and cuboids with different aspect ratios. These functions are tabulated in a data repository (see Ref. [118]). We computed KBIs for subvolumes with different shapes, using an analytic RDF model representing an isotropic liquid. In the thermodynamic limit, KBIs are independent of the shape of the subvolume, and the approach to the thermodynamic limit only depends on the area over volume ratio, and not the shape of the subvolume. This is due to the observation that for small  $r$ ,  $w(r)$  is only a function of  $r$  and the surface to volume ratio of the subvolume, and independent of the shape of the subvolume. At small  $r$ , a universal expression for  $w(r)$  was derived and used to show how truncated KBIs (Eq. (2.8)) correspond to the nonphysical case of a finite volume with zero surface area. It would be interesting to investigate whether or not these findings are applicable to non-isotropic liquids.



# 3

## FINITE-SIZE EFFECTS OF KIRKWOOD-BUFF INTEGRALS FROM MOLECULAR SIMULATIONS



This chapter is based on the paper: N. Dawass, P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, J. M. Simon, and T. J. H. Vlugt, *Finite-size effects of Kirkwood–Buff integrals from molecular simulations*, Molecular Simulation, **44**, 599–612 (2017) [80]

### 3.1. INTRODUCTION

To compute KBIs using molecular simulation, Krüger and co-workers [74] derived an expression for KBIs of finite subvolumes, embedded in larger reservoirs. From the scaling of KBIs of finite subvolumes with the inverse size of these subvolumes, estimates for KBIs in the thermodynamic limit are obtained. When computing KBIs from molecular simulations of closed, and finite systems one must be aware of two system-size effects that originate from: (1) computing finite-volume KBIs from subvolumes embedded in closed and finite simulation boxes [122]; and, (2) using RDFs obtained from finite and closed simulation boxes, in contrast to open, infinite systems as required by the KB theory. For the second effect, RDFs from open systems,  $g^\infty(r)$ , can be estimated from  $g(r)$  obtained from molecular simulations of closed and finite systems. Previously, Krüger et al. [74] used a correlation based on expanding the difference between  $g(r)$  and  $g^\infty(r)$  in a Taylor series in the inverse of the number of molecules. Ganguly and van der Vegt [113] suggested a correction based on the excess or depletion of molecules within a distance,  $r$ , around a central molecule. Cortes-Huerto et al. introduced a correlation to compute the KBIs that includes a correction to the RDF finite-size effects [83].

The objective of this chapter is to investigate finite-size effects associated with the computation of KBIs from molecular simulations (e.g. MD) of finite and closed systems. For the two types of finite-size effects studied (Figure 3.1), we present the inaccuracies as a function of the simulation box size. Also, we investigate subvolume sizes and provide the distances up to which RDFs need to be calculated. This chapter is organized as follows. The finite-size effects are studied using two sets of simulations. The first set of simulations consider MC sampling of KBIs from subvolumes embedded in a larger simulation box. The KBIs in the thermodynamic limit, and the inaccuracies associated with these computations are investigated for various sizes for the simulation box. Second, we demonstrate the finite-size effects related to the RDF by performing MD simulation of a WCA [112] mixture. Other than varying the simulation box size used to obtain KBIs, we also test various RDF correction methods to estimate RDFs of infinite systems,  $g^\infty(r)$ . In section 3.2 details regarding the simulation methods and specifications of the systems studied are provided. In section 3.4, we present two main sets of results: (1) the appropriate system sizes from which KBIs could be extrapolated and the magnitude of the inaccuracies made when simulating finite subvolumes; and (2) the numerical inaccuracies resulting from computing RDFs using closed and finite simulation boxes, and the effect of applying the discussed correction methods of the RDFs. Our findings are summarized in section 3.5.

### 3.2. FINITE-SIZE EFFECTS OF SUBVOLUMES

Computing KBIs using molecular simulations of finite simulation boxes results in finite-size effects. These effects impact the accuracy of the computed KBIs from finite subvolumes, and hence the KBIs at the thermodynamic limit. In this section, we describe the system and simulation method used to study the finite-size effects. Also, we relate the surface area of the subvolume embedded in the simulation box to the finite-size effects of the subvolume.

To deal solely with effects originating from the finite size of the system, and not RDF related effects, we study a fluid that is described by the analytic RDF of Eq. (2.6). As we will consider a pure component fluid, for simplicity the indices  $\alpha$  and  $\beta$  are dropped in this section. Also, the parameter  $\sigma$  is set to unity throughout this work, so we use  $L_{\text{box}}$  for the size of the box instead of  $L_{\text{box}}/\sigma$ . We show how finite KBIs (Eq. (1.25)) scale with the inverse of the size of the spherical subvolume,  $V$  (the subvolume can have any other shape but in this study, we choose to consider spherical subvolumes).

To quantify the inaccuracies resulting from finite size effects of these subvolume, the values of  $G_{\alpha\beta}^V$  are computed from simulations of different sizes of the subvolume,  $R$  (the radius of the spherical subvolume), for a specific simulation box size  $L_{\text{box}}$  (side length of the cubic simulation box). KBIs in the thermodynamic limit  $G_{\alpha\beta}^\infty$  are obtained from extrapolating KBIs of finite subvolumes to infinite subvolume size ( $R \rightarrow \infty$ ). These KBIs are then compared to integrals obtained from the direct numerical integration of Eq. (1.25), with the RDF at each distance computed analytically from Eq. (2.6). To quantify the inaccuracies in the KBIs due to finite-size effects, the differences between KBIs from the numerical integration of Eq. (1.25) and  $G^\infty$  from simulation of finite simulation boxes are compared. Furthermore, we examine the distances up to which the computations of the KB integral are performed (i.e the appropriate subvolume sizes). In molecular simulation, RDFs are typically computed up to half the length of the simulation box ( $L_{\text{box}}/2$ ) and as a result the computed KBIs are limited to this range. However, in simulations one can in principle extend  $r$  up to  $\frac{\sqrt{3}}{2}$  of the box length [124]. In this work, the RDF is extended up to  $\frac{\sqrt{2}}{2}$  of the box length (the range  $\frac{\sqrt{2}}{2} < L_{\text{box}} < \frac{\sqrt{3}}{2}$  involves complex calculations that will not be considered further). In the results section we show how this extension affects the computations of KBIs. When the radius of the subvolume,  $R$ , is larger than half the box size, the surface area and volume of the subvolume are computed by disregarding the sphere caps falling outside the simulation box (see Figure 3.1 (b)). Using  $R - L_{\text{box}}/2$  for the height of the cap falling outside the simulation box, we obtain the following equations for the truncated surface area  $A_{\text{trunc.}}$ , and truncated volume  $V_{\text{trunc.}}$  [82],

$$A_{\text{trunc.}} = 4\pi R \left[ R - 3 \left( R - \frac{L_{\text{box}}}{2} \right) \right], \quad \frac{1}{2} < \frac{R}{L_{\text{box}}} < \frac{\sqrt{2}}{2} \quad (3.1)$$

$$V_{\text{trunc.}} = 2\pi \left[ \frac{2}{3} R^3 - \left( R - \frac{L_{\text{box}}}{2} \right)^2 \left( 2R + \frac{L_{\text{box}}}{2} \right) \right], \quad \frac{1}{2} < \frac{R}{L_{\text{box}}} < \frac{\sqrt{2}}{2} \quad (3.2)$$

3

To compute the KBIs of the system using the RDF of Eq. (2.6), MC sampling is performed to numerically evaluate Eq. (1.18). For simplicity, a brute-force random sampling algorithm is used instead of importance sampling MC. The simulations follow the following algorithm:

1. Specify the size of the cubic simulation box  $L_{\text{box}}$ , the size of the spherical subvolume, with radius  $R$  placed in the center of the box, and the RDF parameters  $\chi$  and  $\sigma$ . In this study, we fix  $\sigma = 1$ , and test different values of  $\chi$  ( $\chi = 1, 2$ , and  $4$ ).
2. Select two random points inside the simulation box,  $P_1$  and  $P_2$ , accept the points if they both fall inside the sphere otherwise, generate new points.
3. Select a random point,  $P_3$ , and accept if the point is outside the sphere. Otherwise, choose a new point  $P_3$ .
4. Calculate the distance between points  $P_1$  and  $P_2$  ( $r_{12}$ ), and points  $P_1$  and  $P_3$  ( $r_{13}$ ). Find  $g(r_{12}) - 1$  and  $g(r_{13}) - 1$  using Eq. (2.6).
5. Performing steps 2 and 3 results in one cycle. After a sufficient number of cycles (in this study we perform  $10^{11}$  cycles), compute the average of the integrals  $\int_V \int_V$  and  $\int_V \int_{L_{\text{box}}^3 - V}$  (from Eq. (1.18) and Eq. (1.26)) over the RDF.
6. Finally, calculate  $G_{\alpha\beta}^V$  (Eq. (1.18)) and  $Q$  (Eq. (1.26)) using the sampled averages of the integrals  $\int_V \int_V$  and  $\int_V \int_{L_{\text{box}}^3 - V}$ .

### 3.3. FINITE-SIZE EFFECTS OF RADIAL DISTRIBUTION FUNCTIONS

While our formulation enables the calculation of KBIs from finite subvolumes, the used RDF has to be of that of an open and infinite system,  $g^\infty(r)$ . Using  $g(r)$  from molecular simulation of a closed system results in a systematic error that affects the accuracy of the KBIs, thus, the  $g(r)$  has to be corrected. There are several approaches available to estimate  $g^\infty(r)$  from RDFs computed in closed systems. In this section, we present the RDF correction methods that we consider in this

work. To study the RDF correction methods, MD simulations of WCA molecules are performed. The system conditions and the simulation details are provided in this section.

### 3.3.1. VAN DER VEGT CORRECTION

When computing KBIs using finite systems, Ganguly and van der Vegt [113] observe a drifting asymptote due to the asymptotic behavior of RDFs in finite systems. Specifically, this asymptotic behavior of the RDF does not converge to one. These authors proposed that the RDF could be corrected using a correlation that takes into account the excess, or depletion of the bulk density of molecule type  $\beta$  around a central molecule of type  $\alpha$  at a distance  $r$  due to the finite-size of the system (simulation box). The bulk density of molecules  $\beta$  is compensated by computing the excess or depletion of number of molecules of species  $\beta$  inside the considered subvolume,  $V$ . The subvolume is formed by taking a distance  $r$  from the central molecule of type  $\alpha$ . The correlation takes into account the depletion of molecule type  $\beta$  around a central molecule of type  $\alpha$ ,  $\Delta N_{\alpha\beta}(r)$ , and provides the corrected RDF,  $g_{\alpha\beta}^{\text{vdV}}(r)$ :

$$g_{\alpha\beta}^{\text{vdV}}(r) = g_{\alpha\beta}(r) \frac{N_\beta \left(1 - \frac{V}{L_{\text{box}}^3}\right)}{N_\beta \left(1 - \frac{V}{L_{\text{box}}^3}\right) - \Delta N_{\alpha\beta}(r) - \delta_{\alpha\beta}} \quad (3.3)$$

where  $g_{\alpha\beta}(r)$  is the RDF obtained from a MC/MD simulation of a closed system (Eq. (1.2)). For infinitely large and open systems,  $g_{\alpha\beta}(r)$  and  $g_{\alpha\beta}^{\text{vdV}}(r)$  are equal.  $N_\beta$  is the number of molecules of type  $\beta$ ,  $L^3$  is the volume of the closed simulation box (which is assumed to be cubic), and  $\delta_{\alpha\beta}$  is the Kronecker delta.  $V$  is the subvolume that surrounds a molecule of type  $\alpha$ , with radius  $r$ . This volume is calculated according to whether  $r$  extends to outside the simulation box or not. When  $\frac{r}{L_{\text{box}}} < \frac{1}{2}$ ,  $V$  is simply the volume of the sphere,  $\frac{4}{3}\pi r^3$ . However, when  $\frac{1}{2} < \frac{r}{L_{\text{box}}} < \frac{\sqrt{2}}{2}$ ,  $V$  corresponds to the volume in Eq. (3.2). The excess or depletion of molecule type  $\beta$  around a molecule of type  $\alpha$ ,  $\Delta N_{\alpha\beta}(r)$ , can be calculated from, depending on whether  $r$  extends to outside the simulation box or not [82]:

$$\Delta N_{\alpha\beta}(r) = \begin{cases} \int_0^r dr' 4\pi r'^2 \rho_\beta [g_{\alpha\beta}(r') - 1], & \frac{r}{L_{\text{box}}} < \frac{1}{2} \\ \int_0^{L_{\text{box}}/2} dr' 4\pi r'^2 \rho_\beta [g_{\alpha\beta}(r') - 1] + \\ \int_{L_{\text{box}}/2}^r dr' 2\pi r' (3L_{\text{box}} - 4r') \rho_\beta [g_{\alpha\beta}(r') - 1], & \frac{1}{2} < \frac{r}{L_{\text{box}}} < \frac{\sqrt{2}}{2} \end{cases} \quad (3.4)$$

### 3.3.2. INVERSE- $N$ FINITE-SIZE CORRECTION

A simple method to correct for the finite-size effect observed in radial distribution functions was presented by Krüger *et al.* [74], where the difference between  $g(r)$  and  $g^\infty(r)$  is expanded in a Taylor series in  $1/N$  [125]:

$$g_{\alpha\beta}^{N_1}(r) = g_{\alpha\beta}^\infty(r) + \frac{c(r)}{N_1} + \mathcal{O}\left(\frac{1}{N_1^2}, r\right) \quad (3.5)$$

3

Here,  $g_{\alpha\beta}^{N_1}(r)$  is the radial distribution function for a closed system with  $N_1$  molecules,  $g_{\alpha\beta}^\infty(r)$  is the RDF corrected for the finite-size effect, and  $c(r)$  is a function that describes deviation from an open system. The function  $c(r)$  is usually not known, but can be estimated using two systems with different sizes at the same thermodynamic state (same density/pressure, temperature, and composition). The term  $\mathcal{O}(\frac{1}{N_1^2}, r)$  is the error associated with the truncation in Eq. (3.5), which is a function of the number of molecules used as well as  $r$ . From Eq. (3.5), the corrected  $g_{\alpha\beta}^\infty(r)$  can be expressed as:

$$g_{\alpha\beta}^\infty(r) = \frac{N_1 g_{\alpha\beta}^{N_1}(r) - N_2 g_{\alpha\beta}^{N_2}(r)}{N_1 - N_2} \quad (3.6)$$

where the subscripts 1 and 2 refer to two systems with different number of molecules, but with same density/pressure, temperature, and composition. This method of correcting for the finite-size effect is straightforward, but it requires two different set of simulations, with different box sizes and number of molecules. The box sizes should not be too different and the resulting  $g_{\alpha\beta}^\infty(r)$  can only be extended to the size of the smallest system. Another shortcoming of this method is related to the numerical accuracy arising from subtracting two numbers of the same magnitude, both in numerator and denominator. The resulting numerical instabilities are increased when using two system sizes that are very close to each other [126].

### 3.3.3. CORTES-HUERTO ET AL. CORRECTION

Another RDF correction is proposed by Cortes-Huerto *et al.* [83]. These authors define KBIs from finite systems in terms of fluctuations of the number of molecules as in the work of Krüger *et al* [74]. Their study considers the fluctuations inside a cubic subvolume (as oppose to spherical subvolumes, but the use of a different subvolume geometry should not affect the values of the KBIs at the thermodynamic limit  $G_{\alpha\beta}^\infty$  [122]). The KBIs are also defined in terms of integrals over the RDF of the system (Eq. (1.18)). To compute KBIs using finite volumes, these authors modify the L.H.S of Eq. (1.18) to include finite effects of RDFs and

finite subvolumes effects. For the finite effects of RDFs, a correction based on a relation from the work of Ben-Naim [13] is used, that strictly only applies when  $r \rightarrow \infty$ ,

$$g_{\alpha\beta}(r) = g_{\alpha\beta}^{\infty}(r) - \frac{1}{L_{\text{box}}^3} \left( \frac{\delta_{\alpha\beta}}{\rho_\alpha} + G_{\alpha\beta}^{\infty} \right) \quad (3.7)$$

The application of Eq. (3.7) implies that the difference between  $g_{\alpha\beta}(r)$  and  $g_{\alpha\beta}^{\infty}(r)$  in Eq. (3.7) is independent of  $r$  (for all values of  $r$ ). In the results section, we compare the validation of all RDF corrections over the whole range of  $r$ . When including the RDF correction (Eq. (3.7)) in the L.H.S of Eq. (1.18) the following expression for the finite KBIs is obtained [83],

$$G_{\alpha\beta}(V, L_{\text{box}}^3) = \frac{1}{V} \int_V \int_V \left[ g_{\alpha\beta}^{\infty}(r_{12}) - 1 \right] d\mathbf{r}_1 d\mathbf{r}_2 - \frac{V}{L_{\text{box}}^3} \left( \frac{\delta_{\alpha\beta}}{\rho_\alpha} + G_{\alpha\beta}^{\infty} \right) \quad (3.8)$$

The effect of the finite size of the subvolume,  $V$ , is accounted for by considering the boundary effects considered through the function  $Q$  (Eq. (1.27)). The double integral in Eq. (3.8),  $\int_V \int_V$ , is expanded to account for the other integration domains,  $\int_V \int_{L_{\text{box}}^3}$  and  $\int_V \int_{L_{\text{box}}^3 - V}$ . As explained earlier in sections 1.3.2 and 3.2, particles in a layer outside  $V$  in the volume  $L_{\text{box}}^3 - V$  contribute to the double integral  $\int_V \int_V$ . This contribution scales with the surface area,  $S$ , of the subvolume. Considering the finite subvolume effect and using  $S/V \propto 1/V^{1/3}$ , we have

$$G_{\alpha\beta}(V, L_{\text{box}}^3) = \frac{1}{V} \int_V \int_{L_{\text{box}}^3} \left[ g_{\alpha\beta}^{\infty}(r_{12}) - 1 \right] d\mathbf{r}_1 d\mathbf{r}_2 - \frac{V}{L_{\text{box}}^3} \left( \frac{\delta_{\alpha\beta}}{\rho_\alpha} + G_{\alpha\beta}^{\infty} \right) + \frac{C_{\alpha\beta}}{V^{1/3}} \quad (3.9)$$

where  $C_{\alpha\beta}$  is a constant that is unique for each thermodynamic state of the system. Cortes-Huerto et al. [83] restrict the volume  $V$  between  $V_\zeta$  and  $L_{\text{box}}^3$ , where  $V_\zeta = 4\pi\zeta^3/3$ . As a result of the values of  $r$  being always larger than  $\zeta$ , the value of  $g_{\alpha\beta}^{\infty}(r_{12})$  is set to one. Additionally, it is assumed that the system is translationally invariant and the transformation  $\mathbf{r}_2 \rightarrow \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$  applies which transforms the integrals in Eq. (3.9) to the ones in the original KBIs expression (Eq. (1.25)). Applying these assumptions, the following expression for KBIs for finite subvolumes was derived [83],

$$G_{\alpha\beta}^V = G_{\alpha\beta}^{\infty} \left( 1 - \frac{V}{V_{\text{box}}} \right) - \frac{V}{V_{\text{box}}} \frac{\delta_{\alpha\beta}}{\rho_\alpha} + \frac{C_{\alpha\beta}}{V^{1/3}} \quad (3.10)$$

$C_{\alpha\beta}$  is a constant originating from the scaling of  $G_{\alpha\beta}^V$  with  $A/V$ , and it is specific to each thermodynamic state. By defining  $\lambda = (V/V_{\text{box}})^{1/3}$ , we can write:

$$\lambda G_{\alpha\beta}^V(\lambda) = \lambda G_{\alpha\beta}^\infty(1 - \lambda^3) - \lambda^4 \frac{\delta_{\alpha\beta}}{\rho_\alpha} + \frac{\alpha_{ij}}{V_{\text{box}}^{1/3}} \quad (3.11)$$

by plotting  $\lambda G_{\alpha\beta}^V(\lambda)$  as a function of  $\lambda$ , one can obtain  $G_{\alpha\beta}^\infty$  from the slope of this plot for low  $\lambda$ . The computed KBIs are compared to those computed by other correction methods. Details of the system studied and of the MD simulations are provided in the following section.

### 3

#### 3.3.4. SIMULATION DETAILS

To study effects resulting from computing RDFs from simulation of finite and closed systems, we examine a binary mixture interacting using the WCA potential [112] where the LJ potential is shifted and truncated at  $2^{1/6}\sigma$ . The WCA mixture is simulated in the *NVT* ensemble using the Nose-Hoover thermostat [22]. The MD package LAMMPS [127] is used to perform the simulations, with 1 million initialization timesteps and 5 million integration timesteps for each run. A timestep of 0.001 in LJ reduced units, which are the units used for other variables, are used. All simulations were performed for the same system properties with  $\sigma_{11} = \sigma_{12} = \sigma_{22} = 1.0$ ,  $\epsilon_{11} = \epsilon_{22} = 1.0$ , and  $\epsilon_{12} = 0.1$ . Additionally, the same thermodynamic condition was maintained for all system sizes. A mixture composition of  $x_1 = 0.75$  was used with the following reduced variables  $T^* = 1.8$  and  $\rho^* = 0.7$ . The box length  $L_{\text{box}}$  and number of molecules  $N$  were varied to obtain  $g_{11}(r)$ ,  $g_{12}(r)$ ,  $g_{22}(r)$  as a function of system size. For all box sizes,  $g_{\alpha\beta}(r)$  were computed up to distances of  $(\sqrt{2}/2)L$ . The RDFs are used to calculate finite subvolumes KBIs (Eqs. (1.25)) which are then extrapolated to the thermodynamic limit to obtain  $G_{11}^\infty$ ,  $G_{12}^\infty$ , and  $G_{22}^\infty$ , respectively. To evaluate the KBIs in the thermodynamic limit for the system studied, simulations of a large system are performed. Specifically,  $G_{11}^\infty$ ,  $G_{12}^\infty$ , and  $G_{22}^\infty$  are evaluated from averaging the results of five simulations for a box with  $L_{\text{box}} = 80$ . Each simulation was initialized with a different configuration, thus allowing for computing the standard deviations.

### 3.4. RESULTS AND DISCUSSION

#### 3.4.1. FINITE-SIZE EFFECTS OF SUBVOLUMES

KBIs ( $G^\infty$ ) are computed for the liquid with the analytic RDF model in Eq. (2.6), using the MC algorithm discussed in section 3.2. The KBIs for finite subvolumes,  $G^V$  (in this section we drop the indices  $\alpha\beta$  since a pure fluid is studied) are computed using simulation boxes with the following lengths,  $L_{\text{box}} = 7.5, 10, 15, 20, 40$  and, 50. For each box size, spherical subvolumes with  $R$  up to  $(\sqrt{2}/2)L_{\text{box}}$  are

used. The KBIs for finite subvolumes,  $G^V$  scales with the inverse of the sphere size,  $1/R$ . For each box size, the linear part of the scaling of  $G^V$  is extrapolated up to  $1/R \rightarrow 0$ , to find  $G^\infty$ .

Figure 3.2 (a) shows the scaling behavior for the case of simulating the system when  $\chi = 2$ . The regime where the scaling is linear depends on the size of the simulated box. Larger simulation boxes provide larger linear regimes. The accuracy of the computations of the KBIs in the thermodynamic limit  $G^\infty$  depends on the size of the simulation box. The computed KBIs from the MC simulations are compared to KBIs in the thermodynamic limit,  $G^{\infty,\text{num}}$ , obtained by numerically integrating Eq. (1.25) up to very large distances. The absolute differences between the numerically integrated KBIs  $G^{\infty,\text{num}}$ , and KBIs from simulations are computed using

$$\text{Difference\%} = \frac{|G^{\infty,\text{num}} - G^\infty|}{|G^{\infty,\text{num}}|} * 100\% \quad (3.12)$$

In Table 3.1, the differences (%) are listed when using the system sizes shown in Figure 3.2 and for three  $\chi$  parameters,  $\chi = 1, 2$ , and  $4$ . For these parameters, the values of  $G^{\infty,\text{num}}/\sigma^3$  are  $-1.785$ ,  $-2.041$ , and  $-2.172$  respectively. The values of  $G^\infty/\sigma^3$  were obtained by extrapolating the linear part of the lines in Figure 3.2, which extend until  $R = L_{\text{box}}/2$  (indicated by a dot for each line). In general, for all fluctuation length parameters,  $\chi$ , the difference decreases with the system size. For simulation boxes with  $L_{\text{box}} = 7.5$  or  $10$  the difference is equal to or larger than  $1\%$ , but the deviation decreases by approximately  $75\%$  and  $90\%$  when increasing  $L_{\text{box}}$  to  $15\sigma$  and  $20\sigma$ , respectively. Finally, obtaining the KBIs from simulation boxes with  $L_{\text{box}} = 40\sigma$  and  $L_{\text{box}} = 50\sigma$  results in marginal differences.

The finite-size effect of the subvolumes is shown more clearly when plotting  $A/V$  instead of  $1/R$  as shown in Figure 3.2 (b). This is due to the fact that the linear scaling of  $G^V$  with  $1/R$  is correct up to  $R = L_{\text{box}}/2$  ( $A/V \sim 1/R$ ). When  $R$  is larger than  $L_{\text{box}}/2$ , parts of the sphere fall outside the simulation box and for these distances the ratio  $A_{\text{trunc.}}/V_{\text{trunc.}}$  (Eq. (3.1) and (3.2)) is used instead of  $A/V$ . Another important observation made from Figure 3.2 (b) is related to the size of subvolume used to compute KBIs. For each simulation box size, the dot in the function  $G^V$  indicates the point where the radius of the spherical subvolume equals  $L_{\text{box}}/2$ . As shown in Figure 3.2 (b), increasing  $R$  beyond this value will not extend the linear regime. This finding is manifested when looking at the correlation between  $G^V$  and the ratio between the area and volume of the sphere ( $A/V$ ) which is presented in Figure 3.2 (b). As shown by Eq. (1.27),  $G^V$  scales with  $A/V$ , but this scaling does not continue when  $R$  is larger than half the simulation box size. When the size of the subvolume extends beyond the simulation box the number of molecules surrounding the embedded subvolume decreases greatly,

Table 3.1: Differences (%), calculated using Eq. (3.12), between KBIs obtained from direct numerical integration of Eq. (1.25),  $G^{\infty,\text{num}}/\sigma^3$ , and integrals computed using MC simulations of various simulation box sizes,  $G^{\infty}/\sigma^3$ . The system used is the fluid described by the analytic RDF of Eq. (2.6). The values of the KBIs from numerical integration,  $G^{\infty,\text{num}}/\sigma^3$ , are  $-1.785$ ,  $-2.041$ , and  $-2.172$  for  $\chi = 1$ ,  $\chi = 2$ , and  $\chi = 4$ , respectively. The values of  $G^V/\sigma^3$  obtained from MC simulations of finite simulation boxes are extrapolated to the thermodynamic limit to obtain  $G^{\infty}/\sigma^3$ .

Box length ( $L_{\text{box}}$ )	$\chi = 1$	$\chi = 2$	$\chi = 4$
7.5	1.979	2.262	2.408
10	1.052	1.042	1.000
15	0.323	0.228	0.140
20	0.154	0.105	0.069
40	0.005	0.036	0.007
50	0.003	0.017	0.055

and the subvolume cannot be considered as grand-canonical. Thus, extending the computations of KBIs beyond  $L_{\text{box}}/2$  is not necessary and does not improve the accuracy of the computed KBIs. In fact, when extrapolating  $G^{\infty}$  this range should be avoided and only the linear part of the function  $G^V$  should be used. In section 3.4.2, the best range used for extrapolating the scaling of  $1/R$  and  $G^V$ , to properly obtain KBIs, is further discussed. Finally, we examine the scaling of the function  $Q$  (Eq. (1.27)) with the surface area of the spherical subvolume. Figure 3.3 shows how the function  $Q$  scales linearly with the surface area. The integrals over  $L_{\text{box}}^3$  in Eq. (1.27) become larger as the number of molecules around surface area of the subvolume increase. For all simulation box sizes, the values of the function  $Q$  decrease when  $R$  is larger than  $L_{\text{box}}/2$  due to the decrease in the number of molecules at the boundary of the sphere  $V$ . In any case, regardless of the shape of the subvolume, whether a fully embedded sphere or a truncated sphere, the function  $Q$  scales linearly with the surface area.

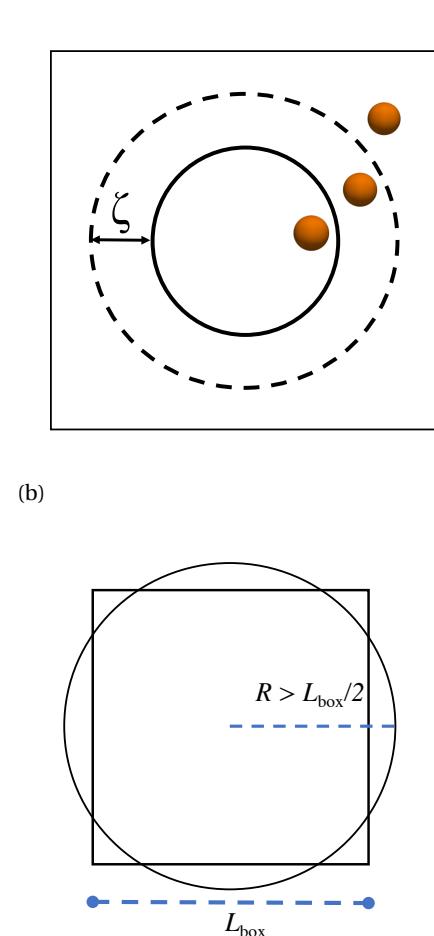


Figure 3.1: Schematic representation of finite-size effects in section 3.2. (a) related to the function  $Q$  in Eq. (1.26) which represents contributions to KBIs emerging from molecular interactions at the surface of the subvolume  $V$ . The only contributions to  $Q$  are from molecule pairs with one molecule (1) inside the subvolume  $V$  and the other molecule (2) in a surrounding layer of thickness  $\zeta$ , where  $\zeta$  is the correlation length of the fluctuations of the RDF (Eq.(2.6)). Molecules outside this layer ( $r_{12} > \zeta$ ) do not contribute to  $Q$ . (b) related to the finite-size of the simulation box. Shown is a schematic representation of the effect of extending  $R$  beyond half the simulation box length,  $L_{\text{box}}/2$ . The volume of the sphere no longer equals  $\frac{4}{3}\pi R^3$ , and the sphere caps falling outside the box has to be subtracted. Eq. (3.1) and (3.2) provide the area and volume of the truncated spherical subvolume.

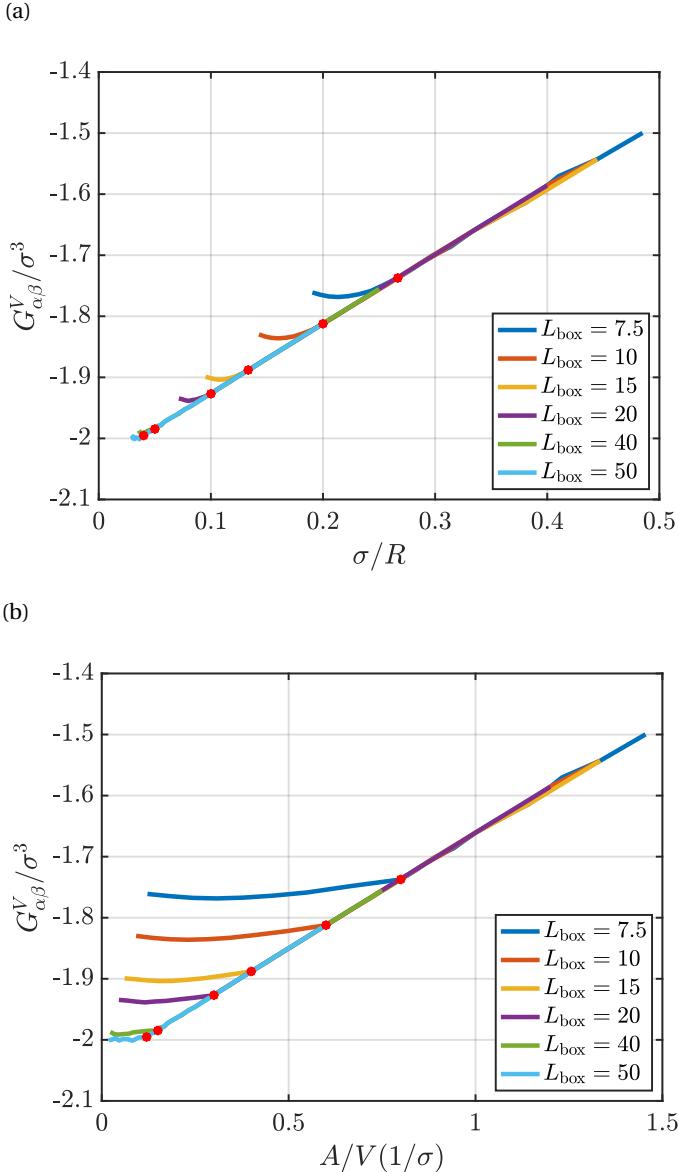


Figure 3.2: KBIs from finite subvolumes of the fluid described with the RDF of Eq. (2.6), with  $\chi = 2$ .  $G^V/\sigma^3$  scales with: (a) the inverse of the sphere size  $1/R$ , and (b) the ratio of the area of the sphere to its volume  $A/V$ . The integrals are computed using MC simulations for different lengths of the simulation box,  $L_{\text{box}}$ , and different radii,  $R$ , of the spherical subvolume. The dots show the points where  $R=L_{\text{box}}/2$  for each box size.

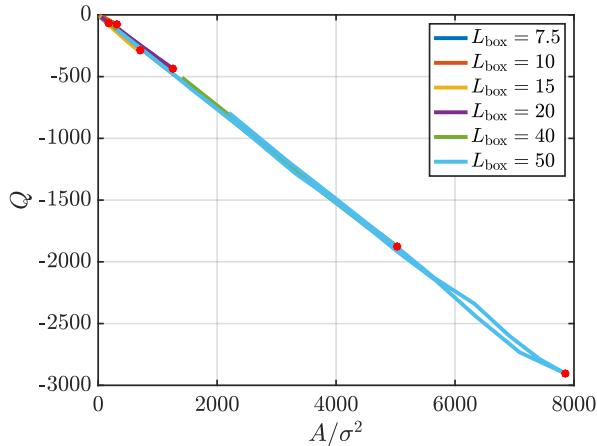


Figure 3.3: The function  $Q$  (Eq. (1.27)) of the fluid described by the RDF of Eq. (2.6), with  $\chi = 2$ . The function scales linearly with the surface area,  $A/\sigma^2$ , for all the box sizes studied. The dots show the points where  $R=L_{\text{box}}/2$ .

### 3.4.2. FINITE-SIZE EFFECTS OF RADIAL DISTRIBUTION FUNCTIONS

Here, we study the RDF-related effects discussed in section 3.3. RDFs obtained from simulations of closed, and finite systems have to be corrected. Figure 3.4 shows the enhanced scaling of  $G_{\alpha\beta}^V$  with  $1/R$  when applying an RDF correction. The correction methods for  $g(r)$  are applied to KBIs computed from closed, finite simulations of the WCA mixture described in section 3.4.2. In the case of obtaining the RDFs from MD simulations, the extrapolation to  $G_{\alpha\beta}^\infty$  is not straightforward. In the following section we show how to identify the linear regime.

#### IDENTIFYING THE LINEAR REGIME

We consider the RDF computed for the binary WCA mixture, while applying the van der Vegt correction, since we find that it provides the most accurate KBIs out of the three methods studied in this work (details are provided in section 3.4.2). The MD simulation details and system conditions are provided in section 3.3.4. The study is performed to identify the extrapolation range from the scaling of  $\sigma/R$  with  $G_{\alpha\beta}^V$  from simulation boxes with  $L_{\text{box}} = 10$ ,  $L_{\text{box}} = 20$ , and  $L_{\text{box}} = 40$ . To find the best linear range, the start of the fitting range,  $1/a$ , and the end,  $1/b$ , are varied (see Figure 3.4 (b)). For each extrapolation range, the square of the correlation coefficient (denoted as  $c^2$  in this work) is computed to assess the linearity of the selected range. Also, the difference between the extrapolated KBIs,  $G_{\alpha\beta}^\infty$ , and the integrals computed from a large system ( $L_{\text{box}} = 80$ ) are computed, using Eq. (3.12). To relate the tested ranges to the dimensions of the system,  $a$  and  $b$  are related to the molecular diameter  $\sigma$ . The variable  $a$  is set based on how many molecular diameters should be discarded at the beginning of the distances at which the RDF is computed. The starting point for the  $1/R$  extrapolation could be varied as follows,

$$\frac{1}{a} = \frac{1}{x\sigma} \quad (3.13)$$

Once  $x$ , and subsequently  $a$ , is set the end point of the extrapolation range could be again related to  $\sigma$ ,

$$\frac{1}{b} = \frac{1}{a + y\sigma} \quad (3.14)$$

In this section, we demonstrate how to find the best extrapolation range for the case of the values of  $G_{22}^V$ . To determine the accuracy of the computed KBIs, differences between values of  $G_{22}^V$  of finite simulation boxes and  $G_{22}^V$  of a very large simulation box ( $L_{\text{box}} = 80$ ) are considered (difference%). The results for the effect of varying  $x$  and  $y$  on the quality of the linear fit ( $c^2$ ) and the accuracy in KBIs computations (difference%) are shown in Figure 3.5, 3.6, and 3.7 for the box sizes  $L_{\text{box}} = 40$ ,  $L_{\text{box}} = 20$  and  $L_{\text{box}} = 10$ , respectively. In these figures, the values of  $-\ln(1 - c^2)$  and  $-\ln(\text{difference})$  are plotted as functions of  $x$  and  $y$ . Also, in

each plot a dashed line is added to indicate values of  $c^2$  and difference% that we consider acceptable. Combinations of  $x$  and  $y$  that fall below the line are considered to give poor estimations of the KBIs. Specifically, the following threshold values are considered, 8 for  $-\ln(1 - c^2)$  and 5 for  $-\ln(\text{difference})$  corresponding to  $c^2 = 0.9997$  and difference= 0.7%, respectively.

Using a system with  $L_{\text{box}} = 40$  provides sufficient linear regime that results in very low values of the difference % between KBIs computed and KBIs from very large systems. Figure 3.5 shows that choosing lower  $x$  and  $y$  is favorable. The same observations are made when studying  $L_{\text{box}} = 20$  (Figure 3.6). Using a large value of  $y$  could result in an extrapolation range that includes the diverging part of the  $G^V$  and  $1/R$  scaling. Figure 3.5 and 3.6 show that, for a given  $x$ , the values of  $-\ln(1 - c^2)$  and  $-\ln(\text{difference})$  start decreasing after a specific  $y$  value. In general,  $y$  should not be larger than  $4\sigma$ . Finally, for  $L_{\text{box}} = 10$ , Figure 3.7 indicates that there is a range that is sufficiently linear. However, very few possibilities of  $x$  and  $y$  combinations provide low differences, i.e. the linear regime extends to a value that deviates from the integral at the thermodynamic limit.

Besides examining the linear range for each system size individually, it is possible to investigate the effect of the system size by comparing Figure 3.5, 3.6 and 3.7. Mainly, larger box sizes provide a longer linear regime and smaller differences between KBIs computed from these boxes and integrals from very large boxes. When performing the study for the other integrals as well,  $G_{11}^V$  and  $G_{12}^V$ , a few general findings could be observed. First, it is recommended to choose  $x$  larger than 1, to avoid any fluctuations at small distances. This corresponds to discarding at least  $1\sigma$  from the  $G^V$  and  $1/V$  scaling line. As for the end of the extrapolation range,  $y$ , taking short distances ensures that the extrapolation range does not extend to the end of the line, where the values of  $G_{\alpha\beta}^V$  are diverging. We find that not exceeding  $b = a + 4\sigma$  ensures that the linear fit is acceptable. This applies to boxes with  $L_{\text{box}}$  larger than 20. Finally, while it is important to use linear ranges with high values of the correlation coefficient,  $c^2$  that does not always result in correct estimations of the KBIs at the thermodynamic limit, especially for small simulation boxes. In the following section, inaccuracies in KBIs resulting from finite sizes related to RDFs are discussed.

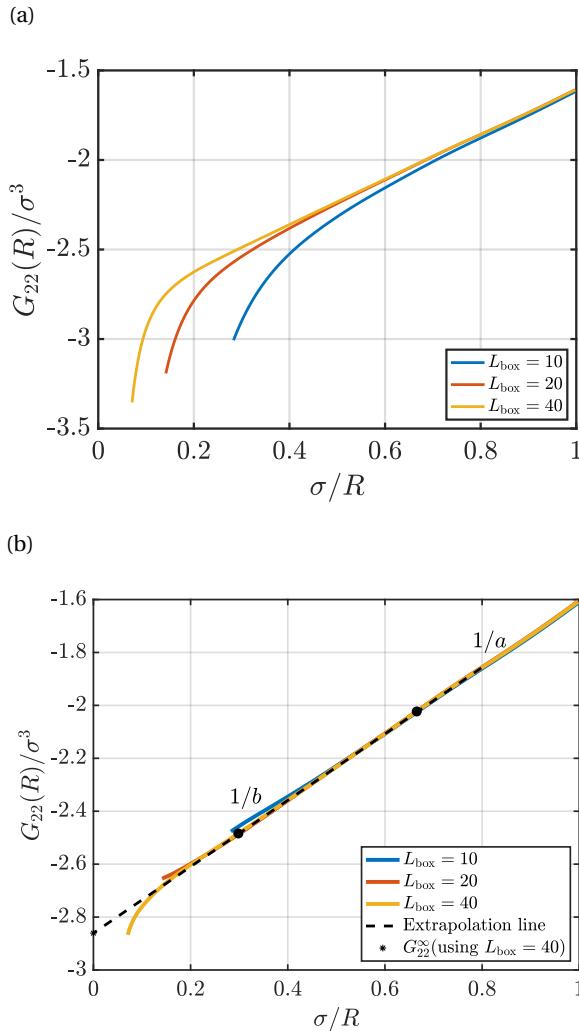


Figure 3.4: KBIs for subvolumes from MD simulations of a binary WCA mixture. The system conditions for all system sizes are:  $x_1 = 0.75$ ,  $T^* = 1.8$ , and  $\rho^* = 0.7$ .  $G_{22}^V/\sigma^3$  is obtained from integrating  $g_{22}(r)$  (Eq. (1.25)) at each subvolume size,  $R$ . The KBIs are computed for simulation boxes with  $L_{\text{box}} = 10, 20$ , and  $40$ . In (a), the RDF is not corrected while in (b) the van der Vegt correction is applied and  $g_{22}^{\text{vdV}}$  is used (Eq. (3.3)). In (b), the range used for extrapolation for the case of  $L_{\text{box}} = 40$  is shown as well as the extrapolation line from which  $G_{22}^\infty/\sigma^3$  is obtained.

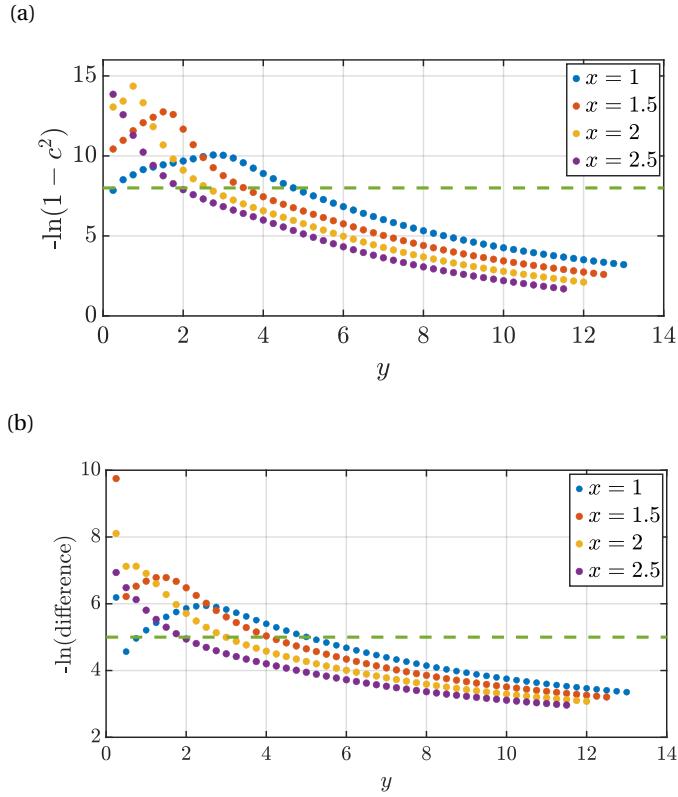


Figure 3.5: The natural logarithm of (a) the linearity  $1 - c^2$  and (b) the differences between KBIs of finite simulation boxes and KBIs of a very large system, as a result of varying the linear range used to extrapolate to  $G_{22}^\infty$ . The KBIs are computed from MD simulations of the binary WCA mixture (details in section 3.3.4). The size of the simulation box is set to  $L_{\text{box}} = 40$ . Points with different colors correspond to different starting points of the extrapolation range,  $x$ , and as a result  $a$  (Eq. (3.13)). The variable  $y$  correspond to how far should the range be extended (Eq. (3.14)).

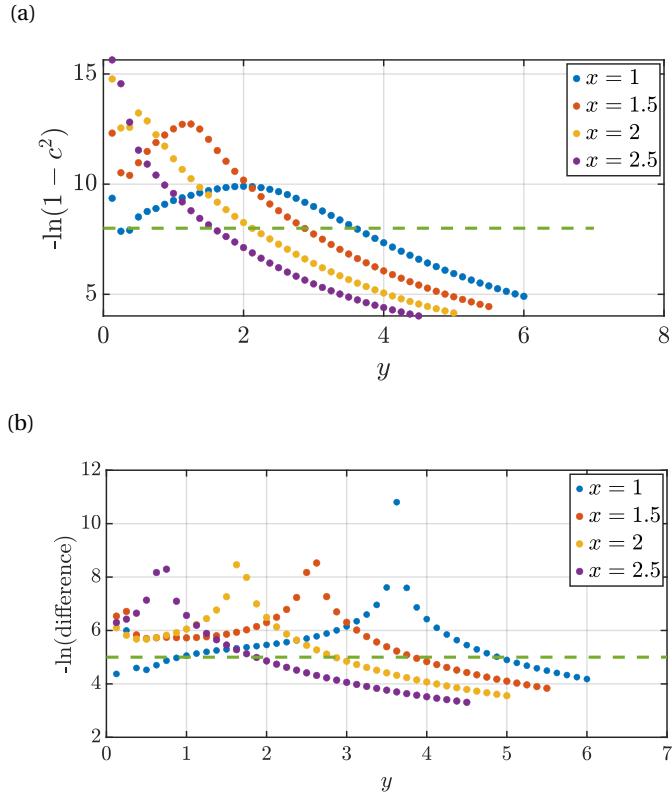


Figure 3.6: The natural logarithm of (a) the linearity  $1 - c^2$  and (b) the differences between KBIs of finite simulation boxes and KBIs of a very large system, as a result of varying the linear range used to extrapolate to  $G_{22}^\infty$ . The KBIs are computed from MD simulations of the binary WCA mixture (details in section 3.3.4). The size of the simulation box is set to  $L_{\text{box}} = 20$ . Points with different colors correspond to different starting points of the extrapolation range,  $x$ , and as a result  $a$  (Eq. (3.13)). The variable  $y$  correspond to how far should the range be extended (Eq. (3.14)).

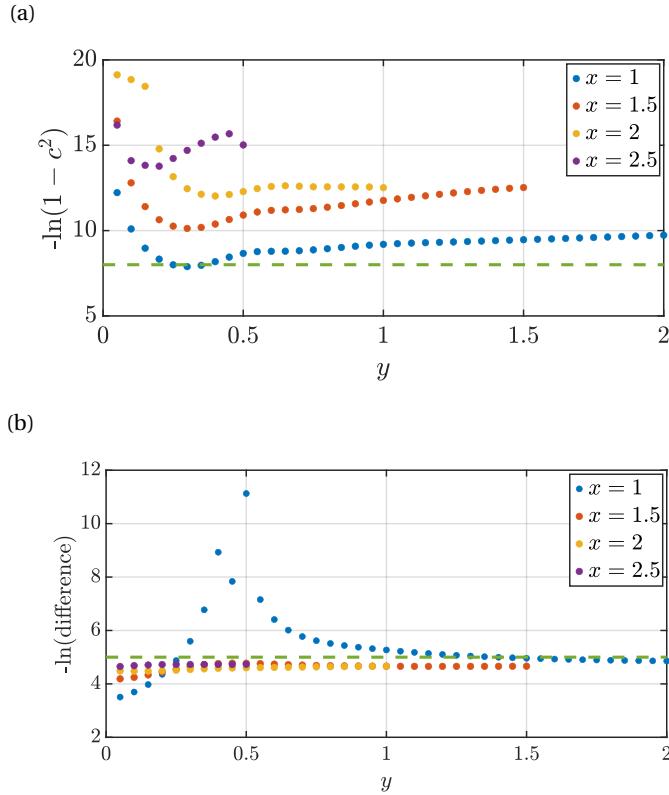


Figure 3.7: The natural logarithm of (a) the linearity  $1 - c^2$  and (b) the differences between KBIs of finite simulation boxes and KBIs of a very large system, as a result of varying the linear range used to extrapolate to  $G_{22}^\infty$ . The KBIs are computed from MD simulations of the binary WCA mixture (details in section 3.3.4). The size of the simulation box is set to  $L_{\text{box}} = 10$ . Points with different colors correspond to different starting points of the extrapolation range,  $x$ , and as a result  $a$  (Eq. (3.13)). The variable  $y$  correspond to how far should the range be extended (Eq. (3.14)).

### COMPARING CORRECTION METHODS

Using RDFs computed from finite and closed simulation boxes leads to a systematic error in KBIs computations. These RDF finite-size effects should be corrected. In Figure 3.4 we already showed the scaling of  $G_{22}^V$  with  $\sigma/R$  (a) when using RDFs that are not corrected and (b) when applying the van der Vegt correction [113]. For all the system sizes used, implementing the correction enhances the linear scaling of the finite-volume KBIs. In addition to the van der Vegt correction, we considered the  $1/N$  correlation method [74], and the method proposed by Cortes-Huerto et al. [83]. We compare between the correction methods by considering the differences between the corrected RDF obtained from small systems and the RDF computed using a large system ( $L_{\text{box}} = 80$ ). Figure 3.8 shows the quantity  $\ln|g(r) - g^{L_{\text{box}}=80}(r)|$  for all distances, where the RDFs are computed for the WCA binary mixture using  $L_{\text{box}} = 10, 20$ , and  $40$ . Generally, the differences decrease with larger box sizes, and the differences are larger for small  $r$ . Also, for all system sizes, the van der Vegt correction and the correction of Cortes-Huerto et al. result in lower deviations from the finite RDF than in the case of the  $1/N$  correlation.

The KBIs  $G_{11}^\infty$ ,  $G_{12}^\infty$ , and  $G_{22}^\infty$  are computed for the binary WCA mixture using three different simulation boxes, with  $L_{\text{box}} = 10$ ,  $L_{\text{box}} = 20$ , and  $L_{\text{box}} = 40$ . In the case of the  $1/N$  correlation, the values of  $G_{\alpha\beta}^\infty$  are obtained by using two simulation box sizes that are not very different. For instance, to compute KBIs from a simulation box with  $L_{\text{box}} = 10$ , RDFs from simulations of boxes with  $L_{\text{box}} = 10$  and  $L_{\text{box}} = 11$ , at the same density and temperature, are required. The obtained KBIs from each correction method are then compared to integrals computed using a system with  $L_{\text{box}} = 80$ . The differences (%) between KBIs computed using  $L_{\text{box}} = 80$  and KBIs  $G_{11}^\infty$ ,  $G_{12}^\infty$ , and  $G_{22}^\infty$  from finite simulation boxes ( $L_{\text{box}} = 10, 20, 40$ ) are presented in Table 3.2. The differences are computed when the RDF is not corrected and when the correction methods discussed in this work are applied. Overall, the differences decrease with larger simulation boxes, with the exception of the  $1/N$  correlation. Additionally, for box sizes up to  $L_{\text{box}} = 20$ , the van der Vegt method results in the lowest differences. For large boxes, Table 3.2 shows that a RDF correction is not needed and in the case of  $L_{\text{box}} = 40$  the differences resulting from not using a correction and the differences resulting from the correction methods are not comparable. Other than the individual KBIs, we also consider the quantity,  $G^f = G_{11} + G_{22} - 2G_{12}$ , which is useful when computing the thermodynamic factor  $\Gamma$  from KBIs [64]. For  $L_{\text{box}} = 80$ , we find that  $G^f/\sigma^3 = -3.38 \pm 0.005$ . Deviations from this value are presented in Table 3.3 for the case of not using a RDF correction as well as when using the three

Table 3.2: Differences (%) between  $G_{\alpha\beta}^{\infty}/\sigma^3$  computed using MD simulations of a large system ( $L_{\text{box}}/\sigma = 80$ ) and KBIs from simulations of small simulation boxes,  $L_{\text{box}} = 10, 20$ , and  $40$ . The MD results are for a binary WCA mixture ( $x_1 = 0.75$ ) at  $T^* = 1.8$  and  $\rho^* = 0.7$  (reduced units). Each column includes differences as a result of implementing a different RDF correction method. For the 1/N correlation, two MD simulations were used,  $L_{\text{box}} = 10, 11$ ,  $L_{\text{box}} = 20, 21$ , and  $L_{\text{box}} = 40, 41$ .

Box length ( $L_{\text{box}}$ )	Without Correction	van der Vegt [113]	1/N correlation [123]	Cortes-Huerto method [83]		
				(a) $G_{11}^{\infty}/\sigma^3$	(b) $G_{12}^{\infty}/\sigma^3$	(c) $G_{22}^{\infty}/\sigma^3$
10	5.6	0.3	0.29	3.7	4	4
20	1.2	0.2	0.9	3.3	3	3
40	0.01	0.2	1.8	2.2	2	2
<hr/>						
10	31	3	4	4	4	4
20	7	0.1	0.3	3	3	3
40	0.2	0.7	17	2	2	2
<hr/>						
10	29	0.9	2	2	4	4
20	8	0.6	3	3	3	3
40	0.4	0.6	13	2	2	2

Table 3.3: Differences (%) between  $G_f^\infty/\sigma^3$  ( $G_f = G_{11} + G_{22} - 2G_{12}$ ) computed using MD simulations of a large system ( $L_{\text{box}} = 80$ ) and the KBIs from simulations of small simulation boxes,  $L_{\text{box}} = 10, 20$ , and  $40$ . The MD results are for a binary WCA mixture ( $\chi_1 = 0.75$ ) at  $T^* = 1.8$  and  $\rho^* = 0.7$  (reduced units). Each column includes differences as a result of implementing a different RDF correction method. For the  $1/N$  correlation, two MD simulations were used,  $L_{\text{box}} = 10, 11$ ,  $L_{\text{box}} = 20, 21$ , and  $L_{\text{box}} = 40, 41$ .

Box length ( $L_{\text{box}}$ )	Without Correction	van der Vegt [113]	$1/N$ correlation [123]	Cortes-Huerto method [83]
10	37	1.8	2.4	5.2
20	9.4	0.3	2.5	3.9
40	0.4	0.8	17.4	2.8

RDF correction methods. In the table, the differences (%) are computed using Eq. (3.12). Using the RDF without a correction results in considerable differences in the values of  $G^f$ , especially for the smallest simulation box with  $L_{\text{box}} = 10$ .

Out of the three methods, the van der Vegt correction [113] leads to the lowest differences in the KBIs results in the thermodynamic limit. In Figure 3.8 we show that the method of Cortes-Huerto et al. [83] is similar to the van der Vegt correction [113] when estimating the function  $g_{\alpha\beta}^\infty(r)$ . Still, when computing the KBIs the van der Vegt correction provides lower differences than the method of Cortes-Huerto et al. [83], which assumes that the finite-size correction of the RDF is independent of  $r$ . Also, the van der Vegt correction is fairly simple to implement, and the corrections are applied to one simulation for each size, unlike the  $1/N$  correlation, where for each size two simulations are required. Another shortcoming of the latter method is its numerical inaccuracy when correcting the RDF, resulting in inconsistency with regard to predicting the KBIs. Furthermore, we apply the correction method proposed by Cortes-Huerto et al. [83] to the same WCA mixture. The method enhances the computation of the integrals from finite systems, but the differences (%) in the KBI computations are not lowered as in the case of the van der Vegt correction.

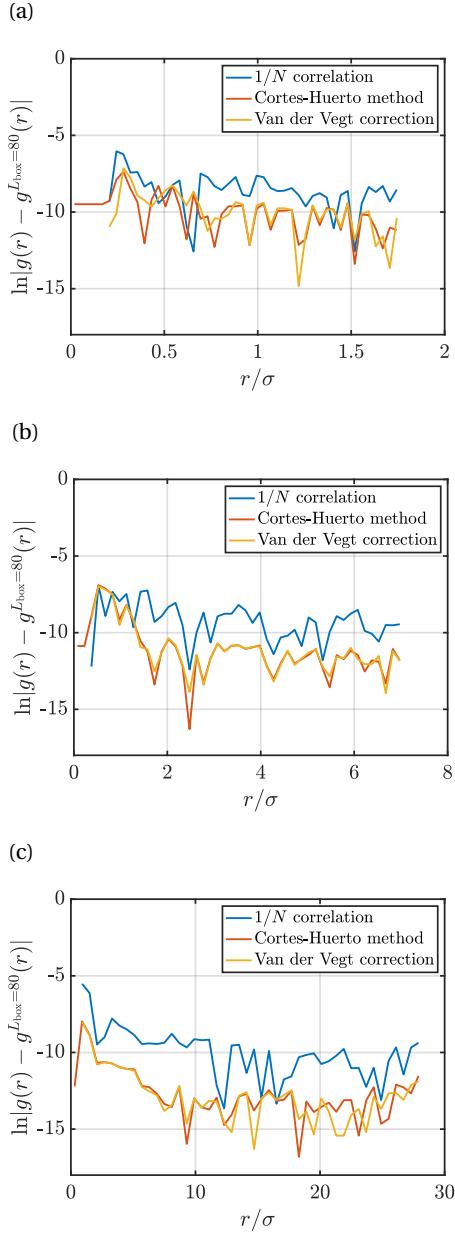


Figure 3.8: The natural logarithm of the difference between RDFs from MD simulation of systems with finite sizes,  $g(r)$ , and RDF from simulation of system with very large size,  $g^{L=80}(r)$ . The RDFs are computed for the binary WCA mixture with:  $x_1 = 0.75$ ,  $T^* = 1.8$ , and  $\rho^* = 0.7$ . The differences are computed for three different simulation box sizes (a)  $L_{\text{box}} = 10$ , (b)  $L_{\text{box}} = 20$ , and (c)  $L_{\text{box}} = 40$ . For each system size, the difference between  $g(r)$  and  $g^{L_{\text{box}}=80}(r)$  is computed at each  $r$ . The different colors represent different RDF correction methods (section 3.3).

### 3.5. CONCLUSIONS

In this chapter, we studied finite size effects related to the computation of KBIs from molecular simulations of finite subvolumes. We presented the uncertainties in KBIs due to: (1) effects due to the finite size of the subvolume, and simulation box, used to compute the KBIs; and (2) effects related to computing RDFs from molecular simulations of closed systems, in contrast to open systems as defined in the KB theory. We showed that uncertainties in the computations of the KBIs decrease when increasing the size of the simulation box, and hence the embedded subvolume. We varied the system size and find that simulation boxes with lengths larger than  $15\sigma$  are sufficient to reduce errors in computed KBIs to below 0.1%. We vary the size of the subvolume, or the maximum distance at which the RDF are computed. We find that a larger distance does not always ensure higher accuracy. In fact, for a given simulation box size, the radius of the spherical subvolume should not be extended beyond half the length of the simulation box. When using an analytic RDF model for the computations of the KBIs, it is relatively straightforward to identify the linear regime in the scaling of finite subvolumes integrals with the inverse size of the subvolume. However, using RDFs computed from MD simulation of WCA molecules did not necessarily result in easily identifiable linear regimes. We presented some guidelines for extrapolating the scaling of finite subvolumes KBIs to the thermodynamic limit. While in some cases small simulation boxes provided a sufficient linear regime, finite-effects caused the resulting KBIs to deviate from those obtained from very large systems. Uncertainties arising from using RDFs of closed systems were evaluated for multiple simulation box sizes, as well as for various RDF correction methods. We demonstrate that using a RDF correction can significantly enhance the convergence of the KBIs and eventually the accuracy of the computations of the integrals. We compare between the RDF correction methods and find that the van der Vegt correction of Ref. [113] achieves the lowest error and is easy to apply.

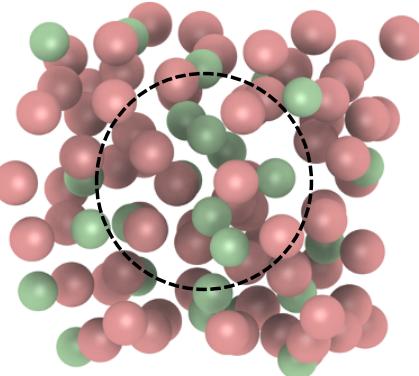


# 4

## SURFACE EFFECTS OF KIRKWOOD–BUFF INTEGRALS FROM MOLECULAR SIMULATIONS

$$G^V = G^\infty + \frac{1}{L} F^\infty$$

Surface term



---

This chapter is based on the paper: N. Dawass, P. Krüger, S. K. Schnell, O. A. Moulton, I. G. Economou, T. J. H. Vlugt, and J. M. Simon, *Kirkwood–Buff Integrals Using Molecular Simulation: Estimation of Surface Effects*. *Nanomaterials*, **10**, 771 (2020) [128]

## 4.1. INTRODUCTION

When studying small systems, of the order of few molecular diameters, thermodynamics of small systems is applied [78, 79]. Using Hill's formulation of small-system thermodynamics [79], it is shown that properties of small systems can be written in terms of volume and surface contributions [122]. In Ref. [122], Hill's thermodynamics was applied to several properties, such as pressure. From the volume contribution of pressure, the homogeneous pressure is obtained, while the Gibbs surface relation was obtained from the surface contribution [122]. This last contribution is proportional to the surface tension. In the case of KBIs, the surface term, or contribution,  $F^\infty$ , can also be defined from the Gibbs surface equation [122]. From a microscopic point of view, it originates from interactions between molecules inside the subvolume and molecules across the boundary of the subvolume [74, 80] (see also section 1.3.2). These surface effects vanish in the thermodynamic limit, but for systems used in MD simulations these effects cannot be neglected [22]. As a result, the quantitative and qualitative study of surface contributions is important for estimating  $G_{\alpha\beta}^\infty$  from integrals of finite subvolumes  $G_{\alpha\beta}^V$ .

The scaling of finite integrals  $G^V$  (Eq. (1.25)) with the size of the subvolumes  $L$  is used to compute KBIs in the thermodynamic limit  $G^\infty$  (for convenience, indices  $\alpha$  and  $\beta$  indicating the different components will be dropped from this point onwards). Specifically,  $G^\infty$  is computed from extrapolating the linear part of the scaling of  $G^V$  with  $1/L$  to the limit  $1/L \rightarrow 0$  [74, 80, 122]. A disadvantage of this approach is that a linear regime is not always easily identified [80].

To avoid extrapolating  $G^V$ , Krüger and Vlugt [81] proposed a direct estimation of KBIs in the thermodynamic limit:

$$G^\infty \approx G_k(L) = \int_0^L [g(r) - 1] u_k(r) dr \quad (4.1)$$

The accuracy of the estimation depends on the function  $u_k(r)$  [129], where the index  $k$  indicates the level of estimation. Krüger and Vlugt [81] considered three different estimations ( $k = 0, 1$  and  $2$ ) and found that integrals computed using the function  $u_2(r)$  provided the best estimation of  $G^\infty$ ,

$$u_2(r) = 4\pi r^2 \left( 1 - \frac{23}{8}x^3 + \frac{3}{4}x^4 + \frac{9}{8}x^5 \right) \quad (4.2)$$

where  $x$  is the dimensionless distance  $x = r/L$ . KBIs computed using Eq. (4.1) and Eq. (4.2) will be denoted by  $G_2$ . To derive the expression for  $G_2$ , the starting point was the scaling of KBIs with  $1/L$ . First, an explicit estimation of  $F^\infty$  in Eq. (1.28) was derived. In the work of Krüger and Vlugt [81],  $F^\infty$  has the following form

$$F^\infty \approx \int_0^\infty [g(r) - 1] 4\pi r^2 \left( -\frac{3}{2}r \right) dr \quad (4.3)$$

It is important to note that the structure of Eq. (4.3) is similar to KBIs in the thermodynamic limit (Eq. (1.3)). So, analogous to Eq. (1.28),  $F^V$  can be defined as,

$$F^V(L) = F^\infty + \frac{C}{L} \quad (4.4)$$

where  $C$  is a constant. For finite systems,  $F^V$  can be computed using

$$F^V \approx \int_0^L [g(r) - 1] \left( -\frac{3}{2}r \right) w(x) dr \quad (4.5)$$

4

where the function  $w(x)$  is provided in Table 1.1 for a spherical subvolume. The similarity between the expression for KBIs (Eq. (1.25)) and surface term (Eq. (4.3)) in the thermodynamic limit allows for deriving an estimation for surface effects as in Eq. (4.1). Using Eq. (4.1), and Eq. (4.3) an explicit expression for surface effects in the thermodynamic limit, denoted here by  $F_2^\infty$ , is obtained from

$$F_2^\infty \approx \int_0^L [g(r) - 1] \left( -\frac{3}{2}r \right) u_2(r) dr \quad (4.6)$$

with  $u_2(r)$  defined in Eq. (4.2).

An alternative method to extrapolate KBIs  $G^V$  to the thermodynamic limit is to use the scaling of  $LG^V$  with  $L$ , rather than the scaling of  $G^V$  with  $1/L$ . The scaling of  $G^V$  in Eq. (1.28) can be rewritten as

$$LG^V(L) = G^\infty L + F^\infty \quad (4.7)$$

By fitting the linear part of the scaling of  $LG^V$  with  $L$ , it is possible to obtain  $G^\infty$  and  $F^\infty$ . Finding the slope and intercepts of a straight line is easier than extrapolating the linear regime of the scaling of  $G^V$  with  $1/L$ . Another advantage of this approach is that an estimation of the surface effects is automatically computed. This estimation can be compared to other available methods for computing  $F^\infty$ . In summary, it is shown that three methods are available for estimating  $G^\infty$  from integrals of finite subvolumes:

1. Using the scaling of  $G^V$  (Eq. (1.25)) with  $1/L$ . To estimate  $G^\infty$ , the linear regime of the scaling is extrapolated to the limit  $1/L \rightarrow 0$ .
2. Using the direct extrapolation formula  $G_2$  (Eq. (4.1)) combined with the function  $u_2(r)$  (Eq. (4.2)). This will converge to  $G^\infty$  for large  $L$ .

3. Computing  $G^\infty$  from fitting the linear regime of the scaling of  $LG^V$  with  $L$  (Eq. (4.7)). The values of the integrals  $G^V$  are computed using Eq. (1.25).

To simplify the estimation of KBIs, it would be useful to evaluate the performance of these methods in terms of accuracy and practicality. Similarly, different methods are available to compute the surface term in the thermodynamic limit  $F^\infty$ :

1. Using the expression in Eq. (4.6).
2. From the scaling of  $LF^V$  with  $L$  (Eq. (4.4)).  $F^V$  is computed using Eq. (4.5). The value of  $F^\infty$  is obtained from the slope of the scaling as  $LF^V(L) = F^\infty L + C$ , in which  $C$  is a constant.
3. From the scaling of  $LG^V$  with  $L$  (Eq. (4.4)). The value of  $F^\infty$  is obtained from the intercept of the scaling.

## 4

The objective of this chapter is to test the estimation of KBIs  $G^\infty$  and the surface effects  $F^\infty$  using the approaches discussed earlier. For both  $G^\infty$  and  $F^\infty$ , the effect of the size of the system is studied. These effects are investigated for both LJ and WCA fluids [130] at different densities. Finally, this work aims at quantifying the contributions of the surface term when computing KBIs of LJ fluid at various densities.

This chapter is organized as follows. In section 4.2, the methods used to compute RDFs, KBIs, and the surface term of KBIs of LJ and WCA fluids are presented. Section 4.2 includes the details of the MD simulations. In section 4.3, the results are presented, which include KBIs and the surface term for WCA and LJ systems at different sizes and densities. Section 4.4 summarises the main findings of this chapter.

## 4.2. METHODS

RDFs of systems of particles interacting via the LJ potential are computed using MD simulations in the *NVT* ensemble. Systems with different densities and number of particles are studied. Also, systems of particles interacting via the WCA potential [112], where only the repulsive part of the LJ potential is included, are considered. For each system, the computed RDF is used to compute KBIs  $G^\infty$  and the surface term  $F^\infty$  in the thermodynamic limit. For both quantities, the methods discussed in section 4.1 are used. In this section, the numerical details of computing RDFs and the required integrals are briefly discussed.

For all systems studied in this work, RDFs are corrected using the Ganguly and van der Vegt correction [113], see section 3.3.1. The corrected RDFs are numerically integrated to obtain  $G^V$ ,  $G_2$ ,  $F^V$ , and  $F_2^\infty$ . Once these quantities are obtained, various methods are implemented to estimate KBIs  $G^\infty$  and the surface

Table 4.1: A brief description of the methods used in this work to estimate KBIs in the thermodynamic limit  $G^\infty$  using RDFs computed from finite systems.

Method	Equations	Description
1. Scaling of $G^V$ with $1/L$	(1.25)	$G^\infty$ is obtained from extrapolating the linear regime of the scaling to $1/L \rightarrow 0$ .
2. Direct estimation of $G_2$	(4.1), (4.2)	A plateau in $G_2$ is identified when plotted as a function of $L$ . To estimate $G^\infty$ , values of $G_2$ in this plateau are averaged.
3. Scaling of $LG^V$ with $L$	(1.25), (4.7)	To find $G^\infty$ , the slope of the linear part of the scaling is computed.

Table 4.2: A brief description of the methods used in this work to estimate the surface term in the thermodynamic limit  $F^\infty$  using RDFs computed from finite systems.

Method	Equations	Description
1. Direct estimation of $F_2^\infty$	(4.6)	A plateau in $F_2^\infty$ is identified when plotted as a function of $L$ . To estimate $F^\infty$ , values of $F_2^\infty$ in this plateau are averaged.
2. Scaling of $LF^V$ with $1/L$	(4.4), (4.5)	To find $F^\infty$ , the slope of the linear part of the scaling is computed.
3. Scaling of $LG^V$ with $1/L$	(1.25), (4.7)	To find $F^\infty$ , the intercept of the linear part of the scaling is computed.

terms  $F^\infty$  in the thermodynamic limit. Table 4.1 provides the relations and description of the methods considered to estimate  $G^\infty$ . Similarly, Table 4.2 presents information regarding the methods used to estimate  $F^\infty$ .

#### 4.2.1. SIMULATION DETAILS

RDFs of LJ and WCA fluids were computed using MD simulations and then used to estimate KBIs and surface effects. The LJ potential is truncated at  $2.5\sigma$  while the WCA potential is a LJ potential that is shifted and truncated at  $2^{1/6}\sigma$ . The MD simulations were carried out using an in-house FORTRAN code. Periodic boundary conditions were applied in all directions. All RDFs were computed from simulations in the  $NVT$  ensemble. The systems were simulated at a dimensionless

temperature  $T = 2$ , dimensionless densities  $\rho$  ranging from 0.2 to 0.8 and using number of particles  $N$  equals to 100, 500, 1000, 5000, 10000, 30000, and 50000. For each system size, the length of the simulation box  $L$  was set according to the required density. The desired temperature was imposed using the Andersen thermostat [22].

All MD simulations started from a randomly-generated configuration for which an energy minimization was used to eliminate particle overlaps. A sufficient number of time steps was used to initialize the system. After initialization, RDFs were sampled every 100 time steps. For both, initialization and production, a dimensionless time step equal to 0.001 was used. The simulation length was chosen depending on the size of the system and the available computational resources. For instance, for systems with  $N = 100$ ,  $1 \times 10^9$  production time steps were carried out, while for the maximum size  $N = 50000$ ,  $7 \times 10^5$  steps were used. Multiple independent simulations were performed for each point  $(\rho, N)$ . The resulting RDFs were then averaged and used to compute  $G^\infty$  and  $F^\infty$ . At high densities ( $\rho > 0.4$ ), RDFs from at least 10 runs are used. At lower densities, at least 20 runs are performed to enhance statistics.

## 4

### 4.3. RESULTS

#### 4.3.1. ESTIMATION OF KIRKWOOD–BUFF INTEGRALS

KBIs in the thermodynamic limit  $G^\infty$  are obtained using the three different approaches discussed earlier. To compare the estimation methods, WCA systems were studied while fixing temperature and density. These parameters define the thermodynamic state of the system. Values of KBIs, computed using different methods, for other densities for LJ and WCA fluids are provided in the Supporting Information of Ref. [128]. After comparing estimation methods of KBIs, the relation between density of the system and KBIs for LJ and WCA system is discussed.

Figure 4.1 shows RDFs for systems of different sizes of a WCA fluid at  $T = 2$  and  $\rho = 0.6$  (dimensionless units). Figure 4.1 (b) shows that using small system sizes, specifically  $N = 100$  and  $N = 500$ , results in RDFs with higher oscillations than large systems, where  $N$  equals to or is larger than 1000. As will be shown later, this causes implications in the computation of  $G^\infty$ . In Figure 4.2, the scaling of KBIs of finite subvolumes  $G^V$  with  $1/L$  is presented. For large systems, where  $N > 500$ , a linear range is identified which can be extrapolated to the limit  $1/L \rightarrow 0$ . Instead of computing  $G^V$ ,  $G^\infty$  can be directly estimated from RDFs using Eqs. (4.1) and (4.2). Figure 4.3 shows the estimation of  $G_2$  for systems with varying sizes. When plotted as a function of  $L$ , the values of the integrals  $G_2$  show a plateau at a constant value which corresponds to  $G^\infty$ . However, Fig-

Table 4.3: KBIs in the thermodynamic limit  $G^\infty$  for a WCA system at  $T = 2$  and  $\rho = 0.6$  (dimensionless units). Values of  $G^\infty$  are computed from systems with various number of particles  $N$  and using the different methods listed in Table 4.1.

$N$	Scaling of $G^V$ with $1/L$	Direct estimation of $G_2$	Scaling of $LG^V$ with $L$
500	$-1.5063 \pm 0.0003$	n/a	$-1.5057 \pm 0.0008$
1000	$-1.5027 \pm 0.0000$	n/a	$-1.5028 \pm 0.0002$
5000	$-1.5012 \pm 0.0000$	$-1.5017 \pm 0.0004$	$-1.5013 \pm 0.0002$
10000	$-1.5012 \pm 0.0000$	$-1.5015 \pm 0.0004$	$-1.5012 \pm 0.0001$
30000	$-1.5004 \pm 0.0001$	$-1.5007 \pm 0.0007$	$-1.5003 \pm 0.0006$
50000	$-1.4999 \pm 0.0001$	$-1.5002 \pm 0.0009$	$-1.500 \pm 0.001$

ure 4.3 (b) shows that this is not true for all system sizes. In fact, the values of  $G^\infty$  can be accurately estimated for systems with a minimum number of particles of 5000, which is larger than the minimum size required in the previous extrapolation method (Figure 4.2). The third method to find  $G^\infty$  is to use the scaling of  $LG^V$  with  $L$  (Eq. (1.28)). Figure 4.4 shows that plotting the integrals of finite subvolumes as  $LG^V$  vs  $L$  results in a clear linear regime that is easily identified. The value of the slope of the fitted line corresponds to the value of  $G^\infty$ . For this method, systems with number of particles equal to or larger than 500 can already be used to compute  $G^\infty$ . In principle, all methods for estimating  $G^\infty$  should result in the same answer in the thermodynamic limit. In Table 4.3, values of KBIs  $G^\infty$  obtained using the three methods studied in this work are listed. For KBIs reported in Table 4.3, only uncertainties larger than 0.01 % are shown. The values are obtained from systems with various sizes. For each system size, a linear range was used to compute  $G^\infty$ . In chapter 3, guidelines were provided for selecting a range for the extrapolation of  $G^V$  vs.  $1/L$ . Essentially, the first few molecular diameters after  $r = 0$ , and distances beyond  $L/2$  should be avoided. Fitting lines of the scaling of  $LG^V$  vs.  $L$  is more convenient. In general, fitting regions are chosen such that the coefficient of determination is equal to or very close to 1. The values of  $G^\infty$  in Table 4.3 show that the three methods provide very similar estimations with statistical uncertainties below 0.1 %. Moreover, the results in Table 4.3 show that computing  $G^\infty$  using the direct estimation of  $G_2$  requires larger systems compared to the other methods. This was found to be true for other densities as well as for systems with LJ particles (see the Supporting Information of Ref. [128]). From studying other systems, it was found that the scaling of  $LG^V$  with  $L$  is the easiest method to apply.

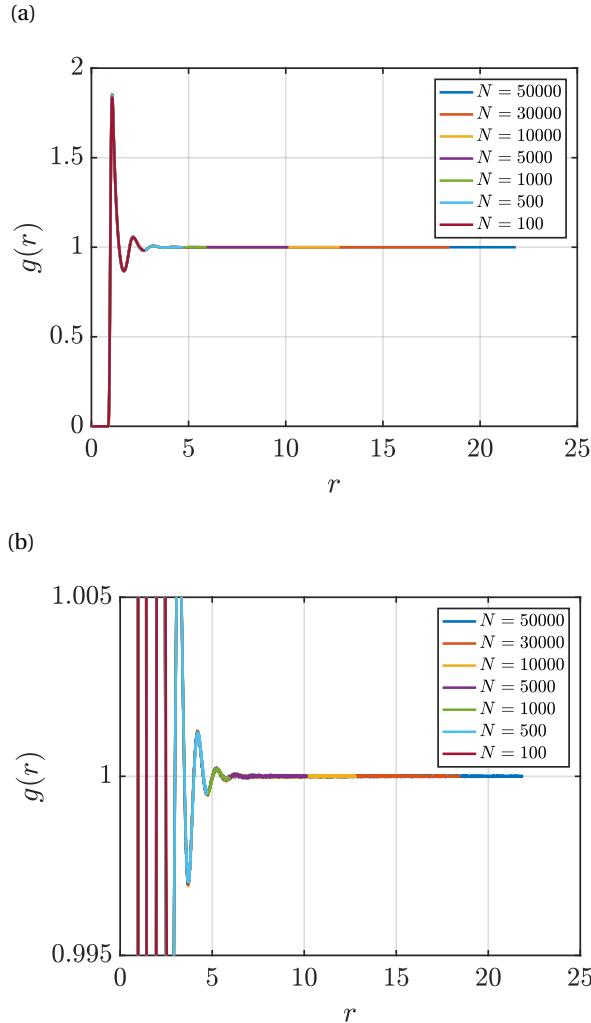


Figure 4.1: (a) RDFs for systems of different sizes of a WCA fluid at  $T = 2$  and  $\rho = 0.6$  (dimensionless units). MD simulations in the  $NVT$  ensemble were used to compute  $g(r)$ , and the Ganguly and van der Vegt correction [113] was applied (Eq. (3.3)), (b) a close-up of the plots in (a) is shown.

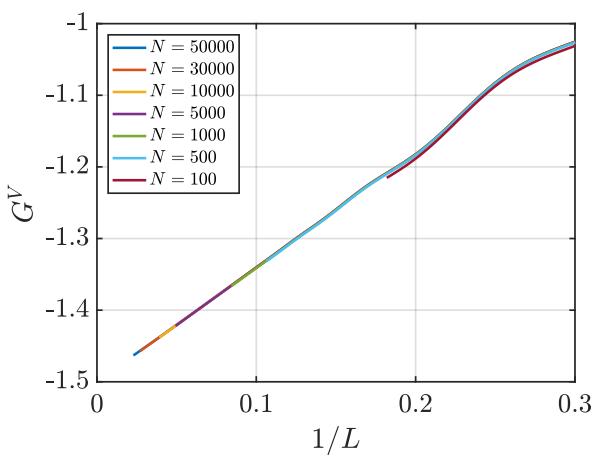


Figure 4.2 KBIs of finite spherical subvolumes  $G^V$  (Eq. (1.25)) vs.  $1/L$  ( $L$  is the diameter of the subvolume) for the WCA fluid at  $T = 2$  and  $\rho = 0.6$  (dimensionless units). The values of  $G^V$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.1.

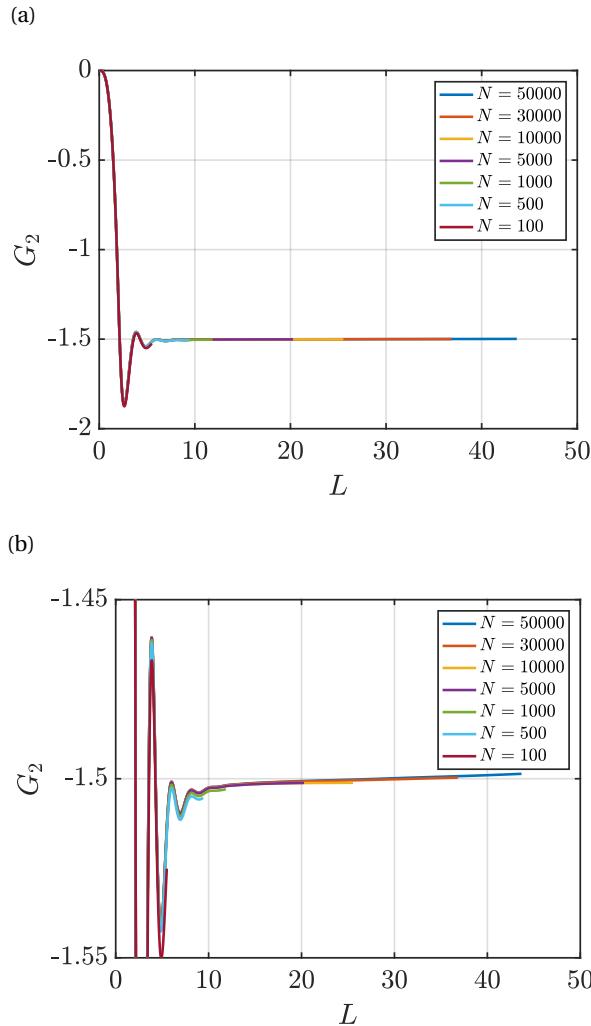


Figure 4.3: (a) Estimation of KBIs in the thermodynamic limit,  $G_2$  (Eqs. (4.1) and (4.2)) vs.  $L$  for the WCA fluid at  $T = 2$  and  $\rho = 0.6$  (dimensionless units), (b) a close-up in of the plots in (a) is shown. The values of  $G_2$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.1.

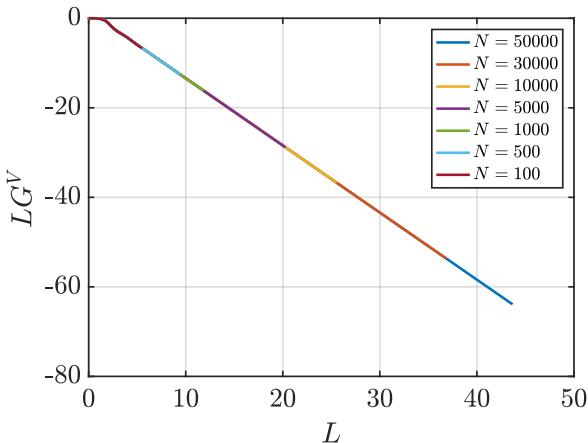


Figure 4.4 KBIs of finite subvolumes multiplied by  $L$ ,  $LG^V$  (Eq. (1.25)) vs.  $L$  for a WCA fluid at  $T = 2$  and  $\rho = 0.6$  (dimensionless units). The values of  $G^V$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.1.

The differences between the estimation methods can be further demonstrated by using a system of LJ particles at  $\rho = 0.4$ , which is more difficult to sample compared to the previously studied system. In Figure 4.5, RDFs of systems of varying sizes of a LJ fluid at  $\rho = 0.4$  and  $T = 2$  are shown. In Table 4.4, KBIs  $G^\infty$  computed using the RDFs in Figure 4.5 are provided. The scaling of  $G^V$  with  $1/L$  is shown in Figure 4.6. For this method, a linear range is not obtained for all system sizes. Systems with at least  $N = 1000$  particles can be used to extrapolate to the thermodynamic limit. In Figure 4.7,  $G_2$  is plotted as a function of the size of the subvolume  $L$ . The figure shows that even larger systems are needed to find a reasonable estimate of  $G^\infty$  using  $G_2$ . Figure 4.7 (b) shows that a plateau is only achieved for large systems where  $N$  equals to or larger than 5000. For this system (LJ fluid at  $\rho = 0.4$ ), it is possible to use the scaling of  $LG^V$  with  $L$  to compute KBIs from small sizes. Figure 4.8 demonstrates that straight lines that are easily fitted are achieved when using the scaling of  $LG^V$  with  $L$ , even for sizes where an estimation can not be made with the other two methods.

Table 4.4: KBIs in the thermodynamic limit  $G^\infty$  for a LJ system at  $T = 2$  and  $\rho = 0.4$  (dimensionless units). Values of  $G^\infty$  are computed from systems with various number of particles  $N$  and using the different methods listed in Table 4.1.

$N$	Scaling of $G^V$ with $1/L$	Direct estimation of $G_2$	Scaling of $LG^V$ with $L$
500	n/a	n/a	$-1.1593 \pm 0.0001$
1000	$-1.1395 \pm 0.0001$	n/a	$-1.1390 \pm 0.0008$
5000	$-1.1161 \pm 0.0006$	$-1.13 \pm 0.02$	$-1.114 \pm 0.004$
10000	$-1.1156 \pm 0.0005$	$-1.13 \pm 0.02$	$-1.114 \pm 0.004$
30000	$-1.1064 \pm 0.0009$	$-1.12 \pm 0.02$	$-1.10 \pm 0.01$

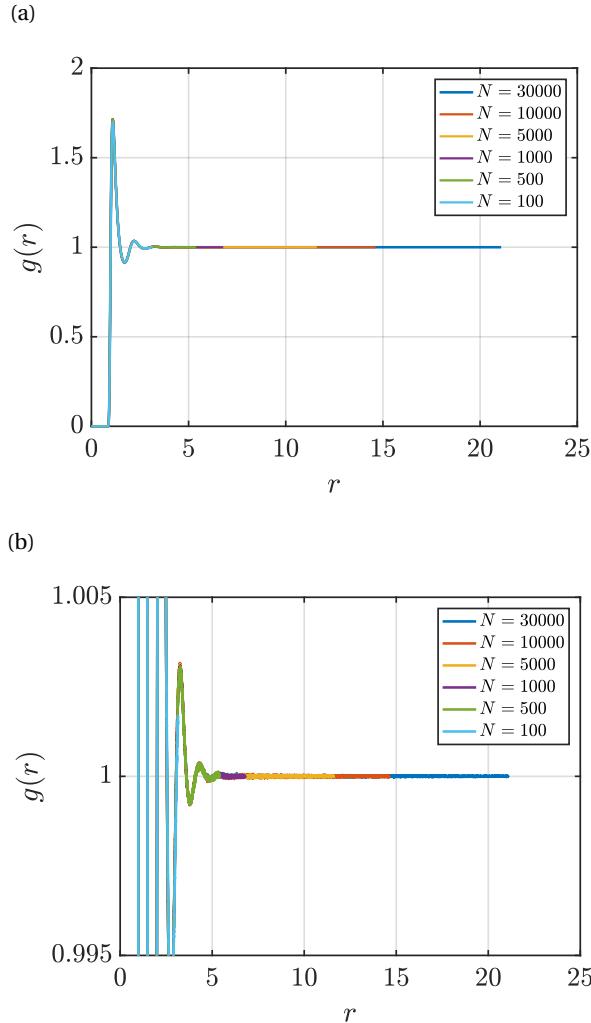


Figure 4.5: (a) RDFs for systems of different sizes of a LJ fluid at  $T = 2$  and  $\rho = 0.4$  (dimensionless units), (b) a close-up in of the plots in (a). MD simulations in the  $NVT$  ensemble were used to compute  $g(r)$ , and the Ganguly and van der Vegt correction [113] was applied (Eq. (3.3)).

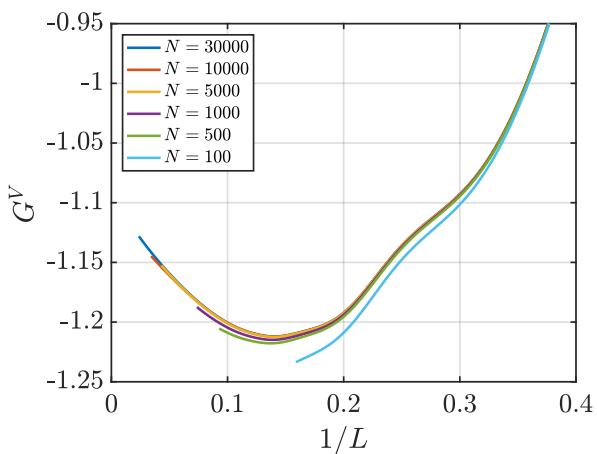


Figure 4.6 KBIs of finite spherical subvolumes  $G^V$  (Eq. (1.25)) vs.  $1/L$  ( $L$  is the diameter of the subvolume) for the LJ fluid at  $T = 2$  and  $\rho = 0.4$  (dimensionless units). The values of  $G^V$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.5.

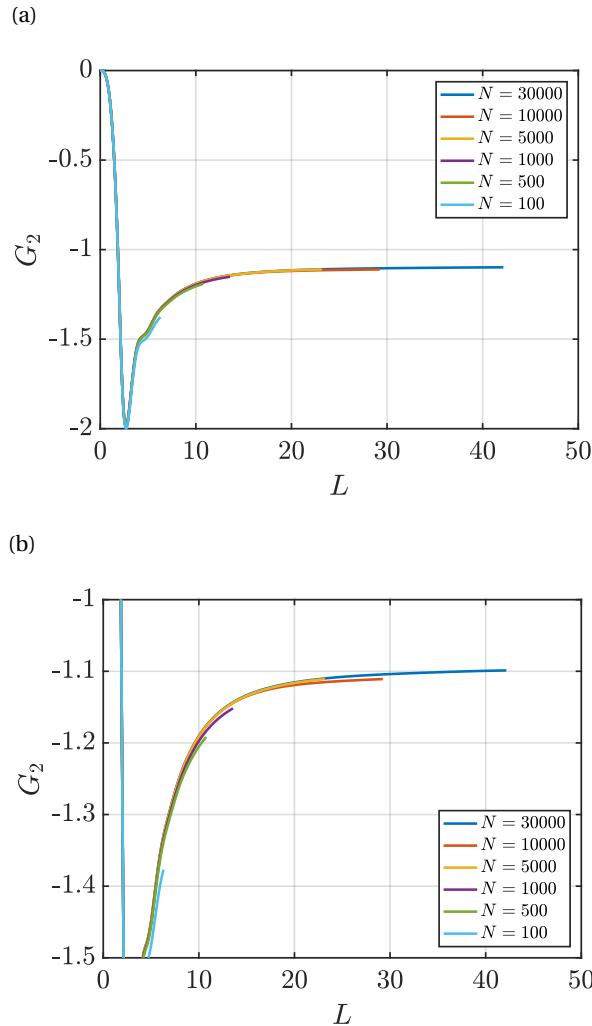


Figure 4.7: (a) Estimation of KBIs in the thermodynamic limit,  $G_2$  (Eqs. (4.1) and (4.2)) vs.  $L$  for the LJ fluid at  $T = 2$  and  $\rho = 0.4$  (dimensionless units), (b) a close-up in of the plots in (a). The values of  $G_2$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.5.

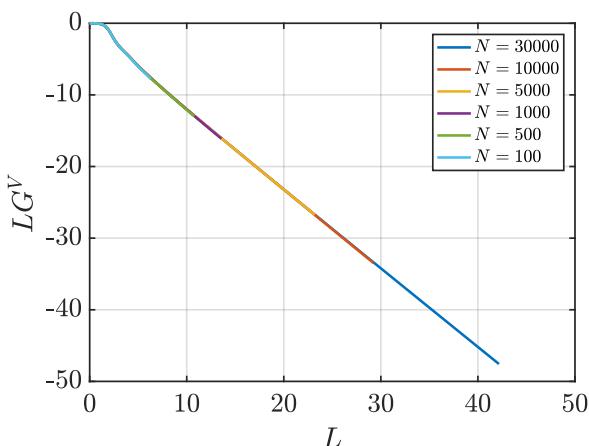


Figure 4.8 KBIs of finite subvolumes multiplied by  $L$ ,  $LG^V$  (Eq. (1.25)) vs.  $L$  for a LJ fluid at  $T = 2$  and  $\rho = 0.4$  (dimensionless units). The values of  $G^V$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.5.

### EFFECT OF SYSTEM SIZE AND DENSITY

Figure 4.9 (a) shows the effect of the size of the system on the values of  $G^\infty$  computed using the scaling of  $LG^V$  with  $L$ . The obtained values of  $G^\infty$  are practically constant. For the LJ fluid, a weak decrease, roughly linear in  $N^{-1/3}$ , is observed. Figure 4.9 (a), Table 4.3, and Table 4.4 demonstrate that statistical uncertainties are small for systems with intermediate sizes ( $N = 5000$  and  $N = 10000$ ). Smaller systems do not provide a sufficient linear regime and very large systems require longer sampling. In Figure 4.9 (b), KBIs at different densities are shown for the LJ and WCA fluids. To estimate  $G^\infty$ , the scaling of  $LG^V$  with  $L$  was used. MD simulations were performed to study systems with dimensionless densities ranging from 0.1 to 0.8.

4

The behaviour of KBIs in the limit  $\rho \rightarrow 0$  can be checked by using the fact that in this limit, the RDF is known analytically,  $g(r) = \exp[-\beta u(r)]$ , where  $u(r)$  is the pair potential [26] and  $\beta = 1/(k_B T)$ . Figure 4.9 (b) shows that for both interaction potentials, the values of  $G^\infty$  computed using molecular simulation approach the correct value in the low density limit. In the high density limit, the differences between  $G^\infty$  of LJ and WCA fluids seem to disappear. At high densities, the repulsive part of the interaction potential, which is the same for WCA and LJ, becomes more important. Hence, the two fluids are expected to behave in the same way as the density increases. This is shown in Figure 4.9 (b).

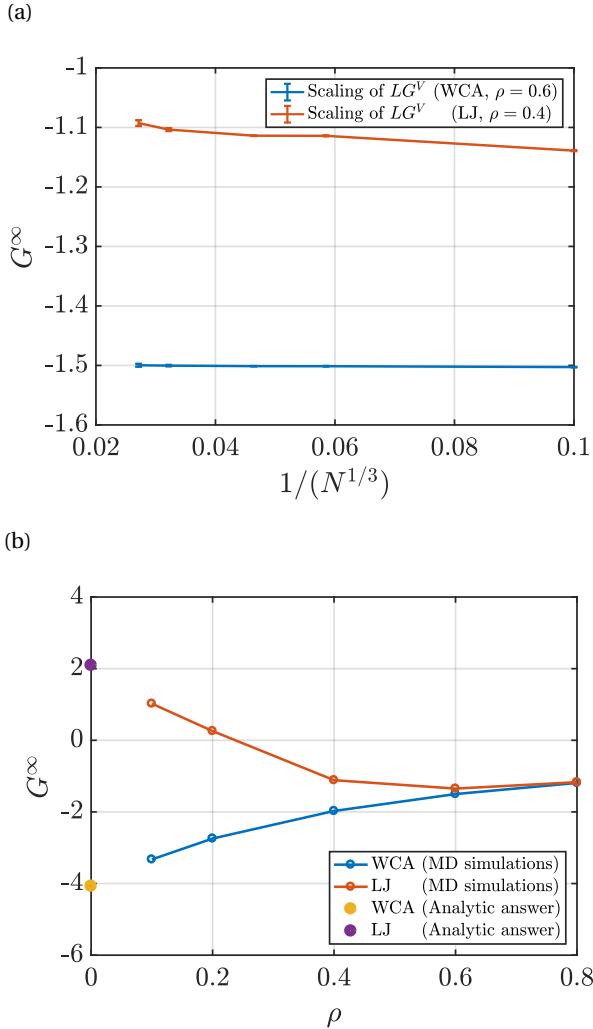


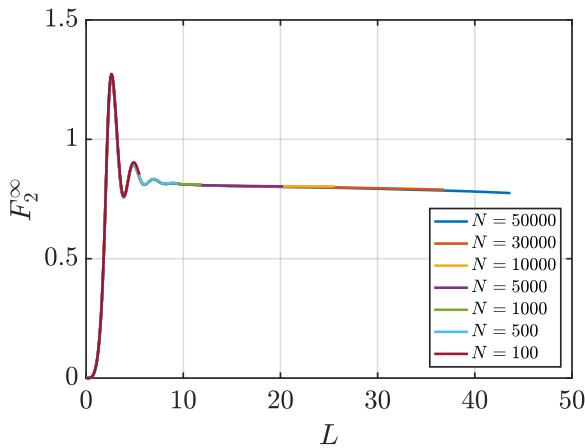
Figure 4.9: (a) KBIs in the thermodynamic limit  $G^\infty$  as a function of the size of the system for the WCA fluid at  $\rho = 0.6$ , and the LJ fluid at  $\rho = 0.4$ . Both fluids are simulated at  $T = 2$  (dimensionless units). (b)  $G^\infty$  as a function of dimensionless density  $\rho$  of LJ and WCA systems at  $T = 2$ . For all densities, the same number of particles is used,  $N = 10000$ . At the limit  $\rho \rightarrow 0$ , the analytic answer is computed using  $g(r) = \exp[-\beta u(r)]$ , where  $u(r)$  is the pair potential [26].

### 4.3.2. ESTIMATION OF SURFACE EFFECTS

An important objective of this work is to investigate surface effects of finite systems used to compute KBIs. As mentioned earlier, there are three possible approaches to compute the surface term in the thermodynamic limit  $F^\infty$ . Similar to the estimation of  $G^\infty$ , the surface term of the WCA fluid is computed from systems with varying number of particles  $N$  at the same thermodynamic state.

In Figure 4.10, estimations of the surface term in the thermodynamic limit  $F_2^\infty$  (Eq. (4.6)) are presented as a function of  $L$ . Unlike the values of  $G_2^\infty$ , a plateau where the values can be averaged is not easily identified. Alternatively, it is possible to consider the scaling of the values of the surface term of finite subvolumes  $F^V$  (Eq. (4.5)). Figure 4.11 shows the scaling of  $LF^V$  with  $L$  for the same WCA fluid. As in the case of the scaling of  $LG^V$ , a linear regime to be fitted is easily identified. The slope corresponds to the value of  $F^\infty$ . Additionally, the value of  $F^\infty$  can be estimated from the intercept with the vertical axis of the line formed from the scaling of  $LG^V$  with  $L$ . The latter two approaches require smaller system sizes than the direct estimation of  $F_2^\infty$ . For instance, Figure 4.11 shows that systems with as few as  $N = 1000$  provide a clear linear range that can be used to estimate  $F^\infty$ . It is of interest to investigate whether the different available methods to find  $F^\infty$  result in matching estimations. The values of  $F^\infty$  computed using the three methods considered in this work are listed in Table 4.5. Results are shown for systems with varying number of particles. While acceptable statistics are achieved for most methods and system sizes, the values of  $F^\infty$  from the three different methods agree less well than the corresponding values of  $G^\infty$ . This can be attributed to the larger statistical errors obtained when compared to estimating  $G^\infty$ .

As in the case of computing  $G^\infty$ , using the scaling of  $LG^V$  provides estimations of the surface term using systems smaller than those required by the other methods. This effect is more significant when looking at a system of LJ particles at a relatively low density. Table 4.6 provides the values of  $F^\infty$  for a LJ fluid at  $\rho = 0.4$ , computed using the methods studied in this work. In Figures 4.12 and 4.13 the scaling of  $F_2^\infty$  with  $L$  as well as the scaling of  $LF^V$  with  $L$  are shown. These plots illustrate that linear regions are not easily identified to compute  $F^\infty$ , in contrast to the higher density case with  $\rho = 0.6$  (Figures 4.10 and 4.11). As a result, when computing surface effects, the scaling of  $LG^V$  with  $L$  is recommended.



4

Figure 4.10 Estimation of the surface term in the thermodynamic limit  $F_2^\infty$  (Eq. (4.6)) as a function of  $L$  for the WCA fluid at  $T = 2$  and  $\rho = 0.6$  (dimensionless units). The values of  $F_2^\infty$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.1.

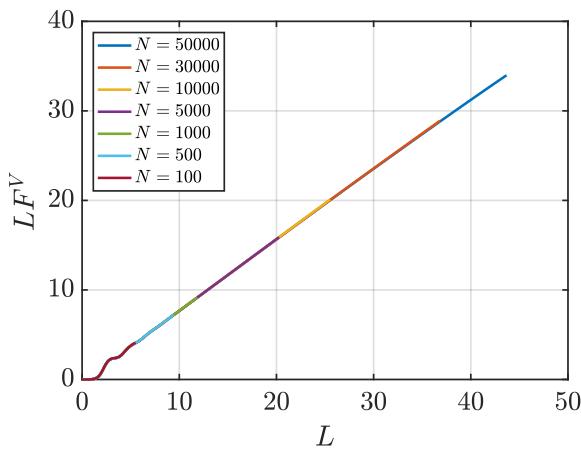


Figure 4.11 Surface effects of finite subvolumes multiplied by the radius of the subvolume  $LF^V$  (Eq. (4.5)) as a function of  $L$  for the WCA fluid at  $T = 2$  and  $\rho = 0.6$  (dimensionless units). The values of  $G^V$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.1.

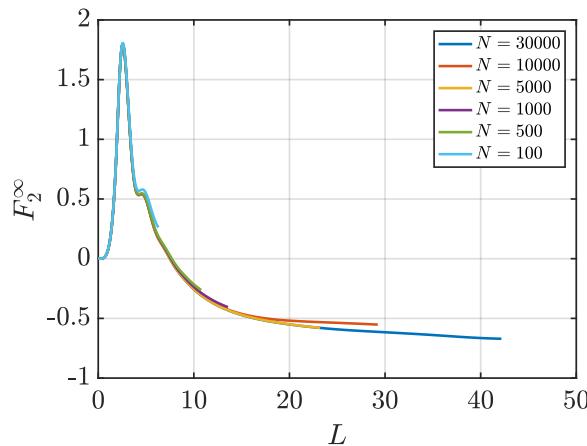


Figure 4.12 Estimation of the surface term in the thermodynamic limit  $F_2^\infty$  (Eq. (4.6)) vs.  $L$  for the LJ fluid at  $T = 2$  and  $\rho = 0.4$  (dimensionless units). The values of  $F_2^\infty$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.5.

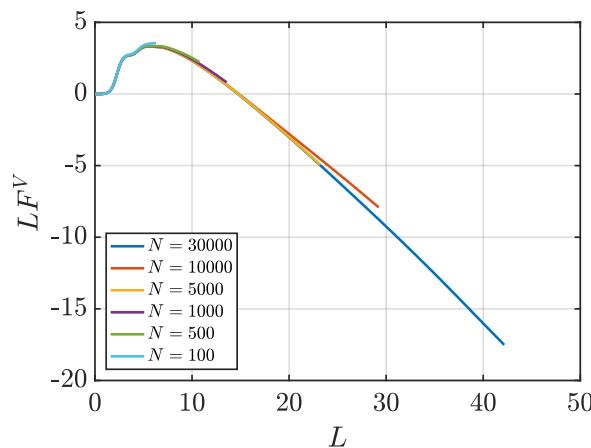


Figure 4.13 Surface effects of finite subvolumes multiplied by the radius of the subvolume  $LF^V$  (Eq. (4.5)) vs.  $L$  for the LJ fluid at  $T = 2$  and  $\rho = 0.4$  (dimensionless units). The values of  $G^V$  are computed for systems with varying number of molecules  $N$ . The used RDFs are provided in Figure 4.5.

Table 4.5: Surface term in the thermodynamic limit  $F^\infty$  for a WCA system at  $T = 2$  and  $\rho = 0.6$  (dimensionless units). Values of  $F^\infty$  are computed from systems with various number of particles  $N$  and using the different methods listed in Table 4.2.

$N$	Direct estimation of $F_2^\infty$	Scaling of $LG^V$ with $L$	Scaling of $LF^V$ with $L$
500	n/a	$0.8168 \pm 0.0008$	n/a
1000	n/a	$0.8082 \pm 0.0002$	$0.804 \pm 0.004$
5000	$0.801 \pm 0.002$	$0.8036 \pm 0.0002$	$0.8004 \pm 0.0002$
10000	$0.8013 \pm 0.0004$	$0.8034 \pm 0.0001$	$0.8013 \pm 0.0003$
30000	$0.795 \pm 0.005$	$0.7979 \pm 0.0006$	$0.79 \pm 0.01$
50000	$0.79 \pm 0.01$	$0.793 \pm 0.001$	$0.78 \pm 0.02$

Table 4.6: Surface term in the thermodynamic limit  $F^\infty$  for a LJ system at  $T = 2$  and  $\rho = 0.4$  (dimensionless units). Values of  $F^\infty$  are computed from systems with various number of particles  $N$  and using the different methods listed in Table 4.2.

$N$	Direct estimation of $F_2^\infty$	Scaling of $LG^V$ with $L$	Scaling of $LF^V$ with $L$
500	n/a	$-0.2483 \pm 0.0001$	n/a
1000	n/a	$-0.3320 \pm 0.0008$	n/a
5000	$-0.53 \pm 0.04$	$-0.460 \pm 0.004$	$-0.5718 \pm 0.0001$
10000	$-0.52 \pm 0.03$	$-0.464 \pm 0.004$	$-0.5433 \pm 0.0004$
30000	$-0.60 \pm 0.06$	$-0.543 \pm 0.008$	$-0.6315 \pm 0.0009$

### EFFECT OF SYSTEM SIZE AND DENSITY

Figure 4.13(a) shows the dependence of  $F^\infty$  on the inverse of the system size. Specifically,  $F^\infty$  decreases as  $(1/N)^{1/3}$ . This is observed for surface terms computed using the scaling of  $LF^V$  with  $L$  as well as values computed using the scaling of  $LG^V$  with  $L$ . In Figure 4.13(a), the error bars of the values of  $F^\infty$  vary with  $N$  in a similar manner to the values of  $G^\infty$ . Statistical uncertainties of the systems studied in this work are provided in Tables 4.5 and 4.6.

As for KBIs, the surface terms  $F^\infty$  can be determined accurately as a function of the density. Figure 4.14 (b) shows the values of  $F^\infty$  with density for LJ and WCA fluids. The surface term is estimated for the range  $\rho = 0.1 - 0.8$ . At  $\rho \rightarrow 0$ ,  $F^\infty$  is computed analytically using Eq. (4.5) and  $g(r) = \exp[-\beta u(r)]$  [26]. In the low density limit, the surface terms computed in this work approach the theoretical value. In the high density limit, differences between surface term of LJ and WCA disappear due to dominating repulsive interactions, which are the same for the LJ and WCA potentials. From Figure 4.9 (b) and Figure 4.14 (b), a comparison between the values of  $G^\infty$  and  $F^\infty$  can be made. Both values change in the

same manner to the density of the system. For all densities, surface terms  $F^\infty/\sigma$  seem to have the same order of magnitude as KBIs  $G^\infty$ . This indicates the significant contribution of surface effects of finite systems used to compute KBIs in the thermodynamic limit. The last observation applies to both LJ and WCA fluids.

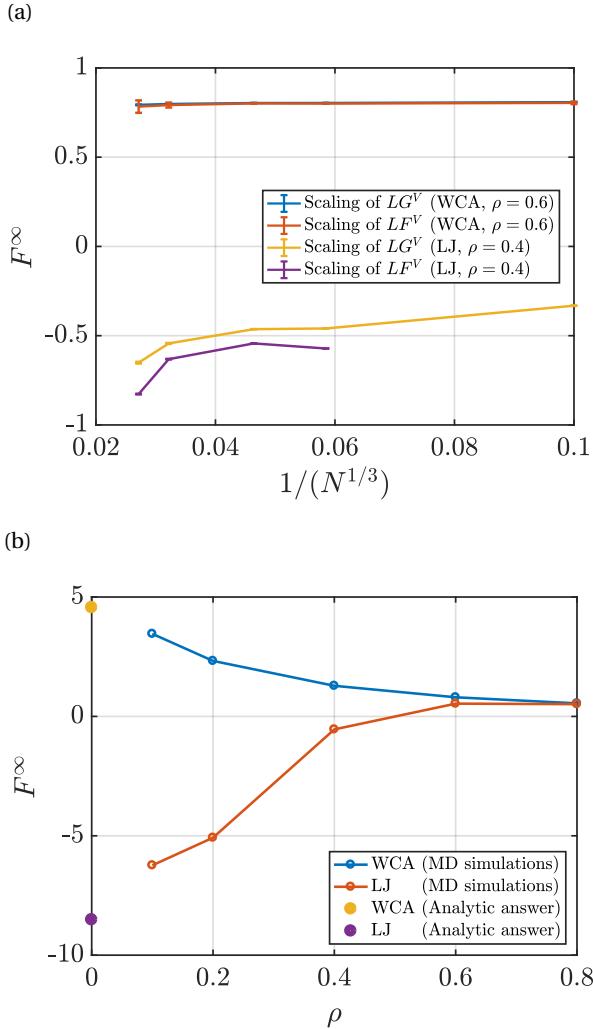


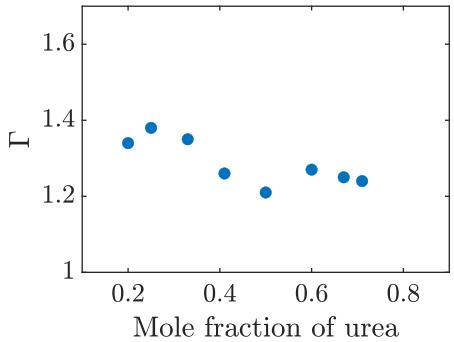
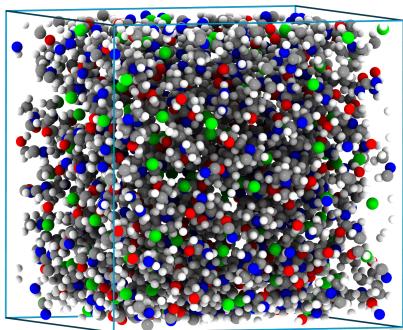
Figure 4.14: (a) Surface term in the thermodynamic limit  $F^\infty$  as a function of the size of the system for the WCA fluid at  $\rho = 0.6$ , and the LJ fluid at  $\rho = 0.4$ . Both fluids are simulated at  $T = 2$  (dimensionless units). (b)  $F^\infty$  as a function of dimensionless density  $\rho$  of LJ and WCA systems at  $T = 2$ . For all densities, the same number of particles is used,  $N = 10000$ . At the limit  $\rho \rightarrow 0$ , the analytic answer is computed using  $g(r) = \exp[-\beta u(r)]$ , where  $u(r)$  is the pair potential [26].

#### 4.4. CONCLUSIONS

In this work, KBIs and surface effects in the thermodynamic limit were computed for systems of LJ and WCA fluids. RDFs of the LJ and WCA systems of different sizes were computed using MD simulations. Different methods were used to estimate KBIs  $G^\infty$  from RDFs of finite systems: scaling of  $G^V$  with  $1/L$ , direct estimation integrals  $G_2$ , and the scaling of  $LG^V$  with  $L$ . The three methods were found to provide reliable estimates of  $G^\infty$ . Differences between the three methods mainly arise from the size of the system required to obtain an accurate estimation. The scaling of  $LG^V$  with  $L$  was found to require smaller systems when compared to other methods. The scaling of  $LG^V$  was found the easiest to implement for estimating KBIs and it provides a suitable estimate of surface effects. Estimating the surface term in the thermodynamic limit  $F^\infty$  is possible from: the finite integral  $F_2^\infty$ , the scaling of  $LF^V$  with  $L$ , as well as the scaling of  $LG^V$  with  $L$ . For all methods, the surface term  $F^\infty$  was found to decrease with increasing system size. The magnitude of the values of  $F^\infty/\sigma$  were found to be the same as the magnitude of the KBIs  $G^\infty$ . Both quantities were found to change in the same manner with the density of the system. KBIs and surface terms were computed for LJ and WCA fluids at different densities. The differences between KBIs of the two systems, LJ and WCA, vanish for high densities as the structure is dominated by repulsive interactions at high density.

# 5

## THERMODYNAMIC AND TRANSPORT PROPERTIES OF UREA-CHOLINE CHLORINE MIXTURES



Kirkwood-Buff Integrals

This chapter is based on the paper: A. T. Celebi, N. Dawass, O. A. Mourtos, and T. J. H. Vlugt, *How sensitive are physical properties of choline chloride - urea mixtures to composition changes: Molecular Dynamics simulations and Kirkwood Buff theory*. Journal of Chemical Physics, submitted. [131]

## 5.1. INTRODUCTION

DESs are produced by mixing quaternary ammonium salts and a hydrogen bond donor (HBD) in a ratio that results in a homogenous solvent with a melting temperature much lower than the melting temperatures of the constituting components [132]. DESs have many similarities with ILs such as the thermophysical behaviour, and tunability [133]. One of the most popular type of DESs is based on the ammonium salt choline chloride (ChCl), which is biodegradable, nontoxic and readily available [133]. To form a DES, ChCl can be easily mixed with various HBDs such as urea, ethylene glycol, or carboxylic acids [133]. ChCl-based DESs are increasingly considered to be a cheaper alternative to organic solvents and ILs [132, 133]. To use DESs in industrial applications, knowledge of their thermodynamic and transport properties is required. While there is an abundant number of publications on ILs [5, 9, 134–138], attention is moving towards studying DESs [132, 133, 139–141]. In recent years, many experimental and computational studies focused on understanding the chemistry of DESs, providing thermodynamic and transport data [6, 11, 132, 133, 142–146].

Molecular simulation is a powerful tool for predicting the properties of complex fluids such as ILs and DESs, while also studying the microscopic structure and interactions [5, 144, 147–149]. As shown in previous chapters, the KB theory provides a theoretically sound framework to predict macroscopic properties from microscopic structure. When studying realistic and complex liquids, KBIs from molecular simulations can be used to: (1) predict a number of thermodynamic properties such as partial molar volumes, and derivatives of the chemical potential with respect to composition, (2) connect properties computed from molecular simulation to experimental measurements as in the case of Fick and MS diffusion coefficients (see section 1.4.3), and (3) compute thermodynamic factors, which indicate the non-ideality of a mixture and the affinity between the different components.

KBIs of mixtures involving ILs and DESs [10, 150, 151] have been mainly computed from experimental data using the inversion of the KB theory. In the limited number of studies that use molecular simulation [44], KBIs were computed by truncating KBIs of infinite systems. Truncating KBIs to a cutoff distance results in poor estimates and is not physical (see chapter 2). In general, computing KBIs of salt solutions, such as ILs and DESs, is not straightforward since the electroneutrality of the system has to be maintained, while KBIs are defined in an open ensemble. This theoretical disparity was discussed in section 1.4.2 of this thesis. The approach of Krüger and co-workers [74] offers a method to compute KBIs from closed systems while using open subvolumes that mimic the grand-canonical ensemble. In this approach, ions and cations can be treated as indistinguishable. In this way, KBIs of pseudo-binary mixture can be computed.

In this chapter, we aim at applying the KB theory to compute thermodynamic properties of mixtures of urea and ChCl. From MD simulations in the *NPT* ensemble, RDFs of mixtures of urea and ChCl are computed at  $T=343.15\text{ K}$  and  $P=1\text{ atm}$ . Mixtures are simulated at eight different mole fractions of urea ranging from 0.20 to 0.71. At each molar concentration, KBIs of a pseudo-binary mixture are estimated, from which the thermodynamic factors and the partial molar volumes are computed. In addition to thermodynamic properties, transport behaviour of mixtures of urea and ChCl are investigated at different molar concentrations using MD simulations. The thermodynamic factors and MS diffusion coefficients are used to compute Fick diffusivities.

This chapter is organized as follows. In section 5.2.1, theoretical expressions to compute RDFs of a pseudo-binary mixture with indistinguishable ions from RDFs of a ternary mixture of anions–cations–solvent are derived. In section 5.2.2, we explain how to compute MS diffusion coefficients of pseudo-binary mixtures. In section 5.2.4, the details of MD simulations and computing KBIs are provided. In section 5.3, results of the scaling of KBIs with the size of the system of mixtures of urea and ChCl are presented, along with values of the thermodynamic factor and partial molar volumes as a function of the mole fraction of urea. Transport properties of mixtures of urea and ChCl are also provided in section 5.3. In section 5.4, the main findings of this chapter are summarised.

## 5.2. METHODS

To compute KBIs of mixtures of urea and ChCl, it is possible to apply expressions of ternary mixtures [13]. Alternatively, mixtures of urea and ChCl can be treated as a pseudo-binary, where the salt components (choline and chloride) are indistinguishable. For a pseudo-binary mixture, mole fractions of urea  $x_{\text{urea}}$  are defined as:

$$x_{\text{urea}} = \frac{N_{\text{urea}}}{N_{\text{urea}} + N_{\text{Ch}} + N_{\text{Cl}}} \quad \text{and} \quad x_{\text{ChCl}} = 1 - x_{\text{urea}} \quad (5.1)$$

where  $N_{\text{urea}}$ ,  $N_{\text{Ch}}$ , and  $N_{\text{Cl}}$  are the number of molecules of urea, choline and chloride molecules, respectively. These mole fractions are used for computing thermodynamic properties of a pseudo-binary mixtures. For instance, when computing thermodynamic factors from KBIs (Eq. (1.14)), the mole fractions are computed as in Eq. (5.1).

### 5.2.1. RADIAL DISTRIBUTION FUNCTIONS OF PSEUDO-BINARY MIXTURES

In this section, we describe how to combine RDFs for components that should be treated as indistinguishable. We are specifically looking at the case where we merge the identity of two RDFs, to create a new "virtual" component. First, we

show the derivation for a ternary mixture  $\alpha$ ,  $\beta$  and  $\gamma$ , where the identities of  $\beta$  and  $\gamma$  are combined to  $\omega$ . Second, we will provide the expressions for the case of combining an arbitrary number of components. Finally, we show that the obtained RDFs converge to the correct answer in the case of an ideal gas.

The radial distribution function of molecules of the same type equals [13]

$$g_{ii}(r) = \frac{n_{ii}(r)/V_{\text{shell}}(r)}{N_i/V_{\text{box}}} = \frac{n_{ii}(r)}{N_i} c(r) \quad (5.2)$$

where  $n_{ii}(r)/V_{\text{shell}}(r)$  is the local density of component  $i$  inside a small radial shell with volume  $V_{\text{shell}}$  at distance  $r$  from a central molecule of type  $i$ ,  $N_i/V_{\text{box}}$  is the overall number density of component  $i$  in the system, and  $c(r) = V_{\text{shell}}(r)/V_{\text{box}}$ , where  $V_{\text{box}}$  is the volume of the simulation box. In the case of molecules of two different types, RDFs are computed from

$$g_{ij}(r) = \frac{n_{ij}(r)}{N_j} c(r) \quad (5.3)$$

where  $n_{ij}(r)$  is the number of atoms of type  $j$  in a radial shell formed around a central molecule of type  $i$ . Based on these general expressions, we can write the following RDFs for the ternary system composed of  $\alpha$ ,  $\beta$  and  $\gamma$ :

$$g_{\alpha\alpha}(r) = \frac{n_{\alpha\alpha}(r)}{N_\alpha} c(r) \quad (5.4)$$

$$g_{\beta\beta}(r) = \frac{n_{\beta\beta}(r)}{N_\beta} c(r) \quad (5.5)$$

$$g_{\gamma\gamma}(r) = \frac{n_{\gamma\gamma}(r)}{N_\gamma} c(r) \quad (5.6)$$

$$g_{\alpha\beta}(r) = \frac{n_{\alpha\beta}(r)}{N_\beta} c(r) \quad (5.7)$$

$$g_{\alpha\gamma}(r) = \frac{n_{\alpha\gamma}(r)}{N_\gamma} c(r) \quad (5.8)$$

$$g_{\beta\alpha}(r) = \frac{n_{\beta\alpha}(r)}{N_\alpha} c(r) \quad (5.9)$$

$$g_{\gamma\alpha}(r) = \frac{n_{\gamma\alpha}(r)}{N_\alpha} c(r) \quad (5.10)$$

$$g_{\gamma\beta}(r) = \frac{n_{\gamma\beta}(r)}{N_\beta} c(r) \quad (5.11)$$

$$g_{\beta\gamma}(r) = \frac{n_{\beta\gamma}(r)}{N_\gamma} c(r) \quad (5.12)$$

Note that RDFs are symmetric, i.e.  $g_{ij}(r) = g_{ji}(r)$ . The expressions above are used to find RDFs of the pseudo-binary mixture ( $\alpha$  and  $\omega$ ), resulting from combining the identity of  $\beta$  and  $\gamma$  into  $\omega$ . For this new system, we need to find expressions for  $g_{\omega\omega}(r)$  and  $g_{\omega\alpha}(r)$ .

To derive an expression for  $g_{\omega\omega}(r)$ , we start with the general RDF expression for similar molecules (Eq. (5.2))

$$g_{\omega\omega}(r) = \frac{n_{\omega\omega}(r)}{N_\omega} c(r) \quad (5.13)$$

The local number of molecules  $n_{\omega\omega}(r)$  is composed of different contributions:  $n_{\beta\beta}(r)$ ,  $n_{\beta\gamma}(r)$ ,  $n_{\gamma\beta}(r)$ , and  $n_{\gamma\gamma}(r)$ . The probability that the central molecule is of type  $\beta$  or of type  $\gamma$  is  $N_\beta/N_\omega$  and  $N_\gamma/N_\omega$ , respectively.  $N_\omega$  is the total number of indistinguishable molecules. In this case,  $N_\omega = N_\beta + N_\gamma$ . Substituting these terms in Eq. (5.13) yields

$$\begin{aligned} g_{\omega\omega}(r) &= \frac{\left(\frac{N_\beta}{N_\omega} n_{\beta\beta}(r) + \frac{N_\gamma}{N_\omega} n_{\gamma\gamma}(r) + \frac{N_\gamma}{N_\omega} n_{\gamma\beta}(r) + \frac{N_\beta}{N_\omega} n_{\beta\gamma}(r)\right)}{N_\omega} c(r) \\ &= \frac{\left(N_\beta n_{\beta\beta}(r) + N_\gamma n_{\gamma\gamma}(r) + N_\gamma n_{\gamma\beta}(r) + N_\beta n_{\beta\gamma}(r)\right)}{N_\omega^2} c(r) \end{aligned} \quad (5.14)$$

Multiplying and diving the nominator by  $\frac{N_\beta N_\gamma}{N_\beta N_\gamma}$ , yields

$$g_{\omega\omega}(r) = \frac{\left(\frac{N_\beta^2 N_\gamma}{N_\beta N_\gamma} n_{\beta\beta}(r) + \frac{N_\gamma^2 N_\beta}{N_\beta N_\gamma} n_{\gamma\gamma}(r) + \frac{N_\gamma^2 N_\beta}{N_\beta N_\gamma} n_{\gamma\beta}(r) + \frac{N_\beta^2 N_\gamma}{N_\beta N_\gamma} n_{\beta\gamma}(r)\right)}{N_\omega^2} c(r) \quad (5.15)$$

Using Eqs. (5.5), (5.6), (5.11) and (5.12) results in

$$g_{\omega\omega}(r) = \frac{\left(\frac{N_\beta^2 N_\gamma}{N_\gamma} g_{\beta\beta}(r) + \frac{N_\gamma^2 N_\beta}{N_\beta} g_{\gamma\gamma}(r) + \frac{N_\gamma^2 N_\beta}{N_\gamma} g_{\gamma\beta}(r) + \frac{N_\beta^2 N_\gamma}{N_\beta} g_{\beta\gamma}(r)\right)}{N_\omega^2} \quad (5.16)$$

The functions  $g_{\beta\gamma}(r)$  and  $g_{\gamma\beta}(r)$  are equal, and Eq. (5.16) can be further simplified to

$$g_{\omega\omega}(r) = \frac{N_\beta^2 g_{\beta\beta}(r) + N_\gamma^2 g_{\gamma\gamma}(r) + 2N_\beta N_\gamma g_{\beta\gamma}(r)}{N_\omega^2} \quad (5.17)$$

Similarly, to find  $g_{\omega\alpha}(r)$ , we apply the general expression for RDFs of two different molecules (Eq. (5.3))

$$g_{\omega\alpha}(r) = \frac{n_{\omega\alpha}(r)}{N_\alpha} c(r) \quad (5.18)$$

where  $n_{\omega\alpha}(r)$  accounts for two contributions:  $n_{\beta\alpha}(r)$  and  $n_{\gamma\alpha}(r)$ . The probability that the central molecule is of type  $\beta$  is  $N_\beta/N_\omega$ . Similarly, the probability that the central molecule is of type  $\gamma$  is  $N_\gamma/N_\omega$ . As a result Eq. (5.18) can be rewritten as

$$\begin{aligned} g_{\omega\alpha}(r) &= \frac{\left(\frac{N_\beta}{N_\omega} n_{\beta\alpha}(r) + \frac{N_\gamma}{N_\omega} n_{\gamma\alpha}(r)\right)}{N_\alpha} c(r) \\ &= \frac{N_\beta}{N_\omega N_\alpha} n_{\beta\alpha}(r) c(r) + \frac{N_\gamma}{N_\omega N_\alpha} n_{\gamma\alpha}(r) c(r) \end{aligned} \quad (5.19)$$

## 5

Using Eqs. (5.9) and (5.10) results in

$$g_{\omega\alpha}(r) = \frac{N_\beta g_{\beta\alpha}(r) + N_\gamma g_{\gamma\alpha}(r)}{N_\omega} \quad (5.20)$$

In the same way, Eqs. (5.17) and (5.20) can be generalized for the case of a system of component  $\alpha$  and  $n$  indistinguishable components 1, 2, 3, ...,  $n$  denoted as  $\omega$ . The RDFs of the pseudo-binary mixture composed of  $\alpha$  and  $\omega$  ( $N_\omega = N_1 + N_2 + \dots + N_n$ ) can be written as:

$$g_{\omega\omega}(r) = \frac{\sum_{i=1}^n \sum_{j=1}^n N_i N_j g_{ij}(r)}{\left(\sum_{i=1}^n N_i\right)^2} \quad (5.21)$$

$$g_{\omega\alpha}(r) = \frac{\sum_{i=1}^n N_i g_{i\alpha}(r)}{\sum_{i=1}^n N_i} \quad (5.22)$$

In the case of an ideal gas, RDFs of the combined molecules  $g_{\omega\omega}(r)$  and  $g_{\omega\alpha}(r)$  should converge to  $(N_\omega - 1)/N_\omega$  and 1, respectively. Considering an ideal gas mixture that consists of  $\alpha$  and  $n$  indistinguishable components, we start with Eq. (5.21) and substitute the RDFs in the expressions with the ideal gas answer:  $g_{ii}(r) = (N_i - 1)/N_i$  and  $g_{ij}(r) = 1$  (where  $i \neq j$ ). This results in:

$$\begin{aligned}
g_{\omega\omega}(r) &= \frac{\sum_{i=1}^n \sum_{j=1(j \neq i)}^n N_i N_j g_{ij}(r) + \sum_{i=1}^n N_i^2 g_{ii}(r)}{\left(\sum_{i=1}^n N_i\right)^2} \\
&= \frac{\sum_{i=1}^n \sum_{j=1(j \neq i)}^n N_i N_j + \sum_{i=1}^n N_i^2 \frac{N_i - 1}{N_i}}{\left(\sum_{i=1}^n N_i\right)^2} \\
&= \frac{\sum_{i=1}^n \sum_{j=1}^n N_i N_j - \sum_{i=1}^n N_i^2 + \sum_{i=1}^n N_i(N_i - 1)}{\left(\sum_{i=1}^n N_i\right)^2} \\
&= \frac{\sum_{i=1}^n N_i \sum_{j=1}^n N_j - \sum_{i=1}^n N_i^2 + \sum_{i=1}^n N_i^2 - \sum_{i=1}^n N_i}{\left(\sum_{i=1}^n N_i\right)^2} \\
&= \frac{\left(\sum_{i=1}^n N_i\right)^2 - \sum_{i=1}^n N_i}{\left(\sum_{i=1}^n N_i\right)^2} \\
&= \frac{\sum_{i=1}^n N_i - 1}{\sum_{i=1}^n N_i} \tag{5.23}
\end{aligned}$$

Similarly, for  $g_{\omega\alpha}(r)$

$$g_{\omega\alpha}(r) = \frac{\sum_{i=1}^n N_i g_{i\alpha}(r)}{\sum_{i=1}^n N_i} = \frac{\sum_{i=1}^n N_i}{\sum_{i=1}^n N_i} = 1 \tag{5.24}$$

Hence, for an ideal gas, Eqs. (5.21) and (5.22) provide the correct answer.

### 5.2.2. MAXWELL-STEFAN DIFFUSION COEFFICIENTS OF PSEUDO-BINARY MIXTURES

The MS diffusion coefficient in a three-dimensional system can be computed from the Onsager coefficients ( $\Lambda_{ij}$ ). Onsager coefficients are defined as the cross-correlation of the displacement of the molecules of species  $i$  and  $j$  in a multicomponent mixture: [66, 152]

$$\Lambda_{ij} = \lim_{t \rightarrow \infty} \frac{1}{2t} \frac{1}{3N} \left\langle \left( \sum_{k=1}^{N_i} [\mathbf{r}_{k,i}(t) - \mathbf{r}_{k,i}(0)] \right) \cdot \left( \sum_{l=1}^{N_j} [\mathbf{r}_{k,j}(t) - \mathbf{r}_{k,j}(0)] \right) \right\rangle \quad (5.25)$$

where  $N$ ,  $N_i$  and  $N_j$  are the total number of molecules, number of molecules of species  $i$  and  $j$ , respectively. The matrix of Onsager coefficients has a symmetric nature. Thus, we can correlate the Onsager coefficients of a binary mixture in terms of the molar masses of the two components  $\alpha$  and  $\beta$  ( $M_\alpha$  and  $M_\beta$ ): [153]

$$\Lambda_{\alpha\beta} = - \left[ \frac{M_\alpha}{M_\beta} \right] \Lambda_{\alpha\alpha} = - \left[ \frac{M_\beta}{M_\alpha} \right] \Lambda_{\beta\beta} \quad (5.26)$$

For binary mixtures, there is a single MS and Fick diffusion coefficient defined. The MS diffusion coefficient  $D$  is related to the Onsager coefficients by:[69]

$$D = \left[ \frac{(M_\beta + x_\alpha(M_\alpha - M_\beta))^2}{x_\alpha x_\beta M_\beta^2} \right] \Lambda_{\alpha\alpha} \quad (5.27)$$

In a conventional ternary system consisting of molecules of  $\alpha$ ,  $\theta$  and  $\gamma$ , there are six Onsager coefficients as  $\Lambda_{\alpha\alpha}$ ,  $\Lambda_{\alpha\theta}$ ,  $\Lambda_{\alpha\gamma}$ ,  $\Lambda_{\theta\theta}$ ,  $\Lambda_{\theta\gamma}$ , and  $\Lambda_{\gamma\gamma}$ . In a pseudo-binary system, there are only three Onsager coefficients that are interdependent according to Eq. (5.26). A pseudo-binary system consists of an independent and an indistinguishable species, which could be described as molecules of  $\alpha$  and  $\beta=(\theta+\gamma)$ , respectively. For this pseudo-binary system, the three Onsager coefficients are  $\Lambda_{\alpha\alpha}$ ,  $\Lambda_{\alpha\beta}$ , and  $\Lambda_{\beta\beta}$ .

To compute Onsager coefficients of the pseudo-binary system  $\Lambda_{\alpha\alpha}$ ,  $\Lambda_{\alpha\beta}$ , and  $\Lambda_{\beta\beta}$ , we can use the Onsager coefficients of the ternary system following the steps below:

1. Perform MD simulations of the ternary systems consisting of molecules of types  $\alpha$ ,  $\theta$  and  $\gamma$ .
2. Compute Onsager coefficients for this ternary system. ( $\Lambda_{\alpha\alpha}$ ,  $\Lambda_{\alpha\theta}$ ,  $\Lambda_{\alpha\gamma}$ ,  $\Lambda_{\theta\theta}$ ,  $\Lambda_{\theta\gamma}$ , and  $\Lambda_{\gamma\gamma}$ ) using MSDs obtained from the OCTP tool in LAMMPS [154]. Eq. (5.25) requires information on the total number of molecules, which is  $N = N_\alpha + N_\theta + N_\gamma$ .

3. The value of  $\Lambda_{\alpha\alpha}$  is identical in the ternary and pseudo-binary systems. Use  $\Lambda_{\alpha\alpha}$  in Eq. (5.27) to compute MS diffusion coefficient of the pseudo-binary mixture.
4. In Eq. (5.27), molecular weights ( $M_\alpha$  and  $M_\beta$ ) of the pseudo-binary mixture are essential.  $M_\beta$  is the average molecular weight of the indistinguishable molecule, computed from:

$$M_\beta = \frac{N_\theta M_\theta + N_\gamma M_\gamma}{N_\theta + N_\gamma} \quad (5.28)$$

In the case of a 1:1 salt solution we have  $N_\gamma = N_\theta$  so therefore

$$M_\beta = \frac{M_\theta + M_\gamma}{2} \quad (5.29)$$

5. For pseudo-binary mixtures, the mole fractions required in Eq. (5.27) can be computed using:

$$x_\alpha = \frac{N_\alpha}{N_\alpha + N_\theta + N_\gamma} \quad \text{and} \quad x_\beta = 1 - x_\alpha \quad (5.30)$$

5

### 5.2.3. FORCE FIELD

ChCl and urea molecules were modelled using the general amber force field (GAFF [155]). Partial charges were derived using the Restrained Electrostatic Potential (RESP) method based on the Hartree-Fock HF/6.31G\* level of theory [147, 156, 157]. As discussed in the earlier works by Perkins et al. [147, 158], Liu et al. [159], Shah and Mjalli [160], and Chaban et al. [161], charge scaling is essential when simulating ILs and DESs due to overestimated electrostatic interaction potentials. Blazquez and co-workers [162] have shown that even for simple electrolytes such as NaCl, charge scaling improves electrostatic interactions. We scaled down the partial charges of ChCl molecules by a factor of 0.8. The GAFF force field combined with reduced charges have been used in MD studies to accurately predict structural, thermodynamic and transport properties of many ChCl-based DESs such as reline, ethaline and glyceline [143, 144, 147, 158, 160, 163, 164]. All force field parameters used in this chapter are listed in the Supporting Information of Ref. [131].

### 5.2.4. SIMULATION DETAILS

MD simulations were performed using the Large-Scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) version released on August, 2018 [127]. Eight different mole fractions of urea ranging from 0.20 to 0.71 were considered. For

Table 5.1: The mole fractions, number of molecules, and initial box lengths of all simulated systems

$N_{\text{ChCl}}/N_{\text{Urea}}$	$x_{\text{Urea}}$	Small System			Large System		
		$N_{\text{ChCl}}$	$N_{\text{Urea}}$	$L_S / [\text{\AA}]$	$N_{\text{ChCl}}$	$N_{\text{Urea}}$	$L_L / [\text{\AA}]$
2/1	0.20	160	80	34.7	800	400	59.4
3/2	0.25	144	96	34.0	720	480	58.1
1/1	0.33	120	120	32.8	600	600	56.2
5/7	0.41	100	140	31.8	500	700	54.4
1/2	0.50	80	160	30.7	400	800	52.6
1/3	0.60	60	180	29.6	300	900	50.6
1/4	0.67	48	192	28.8	240	960	49.3
1/5	0.71	40	200	28.3	200	1000	48.4

## 5

each mole fraction, we considered a small and a large system consisting of 240 and 1200 molecules, respectively. The former was used for computing transport properties (i.e. viscosity, self-diffusivity, MS diffusion), while the latter was used for computing RDFs required for KBIs. It is important to note that larger system sizes were essential to obtain a sufficiently large linear regime of the scaling of KBIs [63]. Table 5.1 shows the number of molecules and the size of the simulation box at each molar ratio of ChCl to urea for all systems simulated in this work.

Initial configurations were generated by randomly inserting molecules in a cubic simulation box using the PACKMOL package [165]. The generated simulation boxes were first relaxed using the conjugate gradient method for 10000 steps. The energy minimization was followed by MD runs in the isothermal-isobaric (*NPT*) ensemble at 343.15 K and 1 atm for 10 ns. In the *NPT* ensemble, average volumes and densities were computed. Starting from the average density obtained from *NPT* runs, each system was then equilibrated at 343.15 K and 1 atm for 1 ns in the canonical (*NVT*) ensemble. Consecutively, transport properties were computed in the next 120 ns. For small systems (see Table 5.1), a 120 ns run typically takes 96 hours using 24 CPUs. During production runs, the OCTP (On-the-fly Computation of Transport Properties) plugin in LAMMPS was used to compute transport properties [154]. The OCTP plugin uses Einstein relations combined with the order-*n* algorithm [22, 166]. For more details about the OCTP plugin, readers are referred to the original study by Jamali et al. [154]. RDFs were computed from separate MD simulations of 80 ns of the large systems in the *NPT* ensemble. For large systems (see Table 5.1), a 80 ns run typically takes 720 hours using 24 CPUs. Finite-size effects of the reported RDFs are cor-

rected using the method proposed by Gaunly and van der Vegt [167]. The RDFs were numerically integrated to obtain KBIs of small subvolumes  $G_{\alpha\beta}^V$  (Eq. (1.25)). To estimate KBIs in the thermodynamic limit, the linear range of the scaling of  $LG_{\alpha\beta}^V$  with  $L$  was used. For more details on the computations of the KBIs, the reader is referred to chapter 5. Using KBIs, thermodynamic factors, and partial molar volumes were computed.

All simulations were carried out at 343.15 K and 1 atm. For *NPT* and *NVT* ensembles, temperature and pressure were maintained using the Nosé-Hoover thermostat and barostat with coupling constants of 0.1 ps and 1 ps, respectively [22]. Long range electrostatic interactions between charged species were calculated based on the particle-particle, particle-mesh (pppm) solver with a relative precision of  $10^{-6}$ . The cut-off radius was set to 12 Å, for both the Lennard-Jones (LJ) and the real-space part of Coulombic potentials. LJ parameters between the dissimilar species were determined based on the Lorentz-Berthelot mixing rules [23]. Equations of motions were integrated using the Verlet algorithm with a time step of 1 fs. Standard deviations in the transport properties and KBIs were computed based on 10 and 25 independent simulations, respectively. Each independent simulation started from a different initial configuration.

## 5.3. RESULTS AND DISCUSSION

### 5.3.1. KIRKWOOD–BUFF INTEGRALS OF UREA–CHOLINE CHLORIDE MIXTURES

Figure 5.1 presents the RDFs of urea–urea, urea–ChCl and ChCl–ChCl interactions at various molar ratios of ChCl to urea. In Figure 5.1, it is shown that the position of the first peak of all RDFs remains unchanged with composition. However, the addition of urea to the system has an effect on the height of the first peak, especially for urea–ChCl and ChCl–ChCl distances. Figure 5.1 (a) shows that urea–urea distances remain relatively unchanged with increasing the molar ratio of urea to ChCl. Figures 5.1 (b) and 5.1 (c) show that the height of the first peak of the  $g(r)$  of urea–ChCl and ChCl–ChCl decreases with increasing molar ratios of urea to ChCl. This indicates weakening of interactions between pairs of urea–ChCl and ChCl–ChCl.

The RDFs shown in Figure 5.1 are used to compute KBIs of urea and ChCl at various molar compositions. In Figure 5.2, KBIs in the thermodynamic limit for pairs of urea–urea, ChCl–urea, and ChCl–ChCl at 343.15 K and 1 atm are shown as a function of the mole fraction of urea. The reader should note that in Figure 5.2 as well as in the following figures we report properties as a function of the mole fraction defined in Eq. (5.1). KBIs of urea–urea pairs are affected the most by the increase of the urea content. Figure 5.2 shows that the values of  $G_{\text{urea–urea}}$

are increasing with the mole fraction of urea. This indicates that urea-urea interactions become stronger as more urea is added to the system. Figure 5.2 also shows that the values of the KBIs  $G_{\text{urea}-\text{ChCl}}$  and  $G_{\text{ChCl}-\text{ChCl}}$  slightly decrease with increasing the content of urea. Consequently, urea-ChCl and ChCl-ChCl interactions become weaker with larger mole fractions of urea. Such trends obtained in KBIs are consistent with the results of RDFs of ChCl mixtures.

It is interesting to examine how interactions between dissimilar components (i.e. ChCl-urea) differ from the interactions between similar components (i.e. urea-urea and ChCl-ChCl) at various mole fractions of urea. To study this, the term  $G_f = G_{\alpha\alpha} + G_{\beta\beta} - 2G_{\alpha\beta}$  can be used. The term  $G_f$  is zero for an ideal mixture in which interactions of  $\alpha$  and  $\beta$  are equal to the average interactions of molecules of the same type. The term  $G_f$  can be computed by integrating the combined RDF  $g_f(r) = g_{\alpha\alpha}(r) + g_{\beta\beta}(r) - 2g_{\alpha\beta}(r)$  as shown in Eq. 1.25. The values of  $G_f$  shown in Figure 5.2 are computed using this approach. Similar to the KBIs of urea-urea pairs, the values of  $G_f$  increase dramatically between  $x_{\text{urea}} = 0.25$  and  $x_{\text{urea}} = 0.4$ . This means that the molecular interactions of the system are shifting for these mole fractions. While the affinity between urea and ChCl is always stronger than the average affinity between similar molecules, attractive interactions between urea and ChCl are stronger when  $x_{\text{urea}} < 0.25$ . Interestingly, the values of  $G_f$  do not significantly change beyond  $x_{\text{urea}} = 0.41$  as more urea is added to the system. When studying other microscopic properties of mixtures of ChCl and urea, Sun et al. [168] has correlated the change of interactions with the content of urea to the eutectic behaviour. The authors reported that the eutectic composition is at the molar ratio of ChCl to urea of 1:2 ( $x_{\text{urea}} = 0.5$ ).

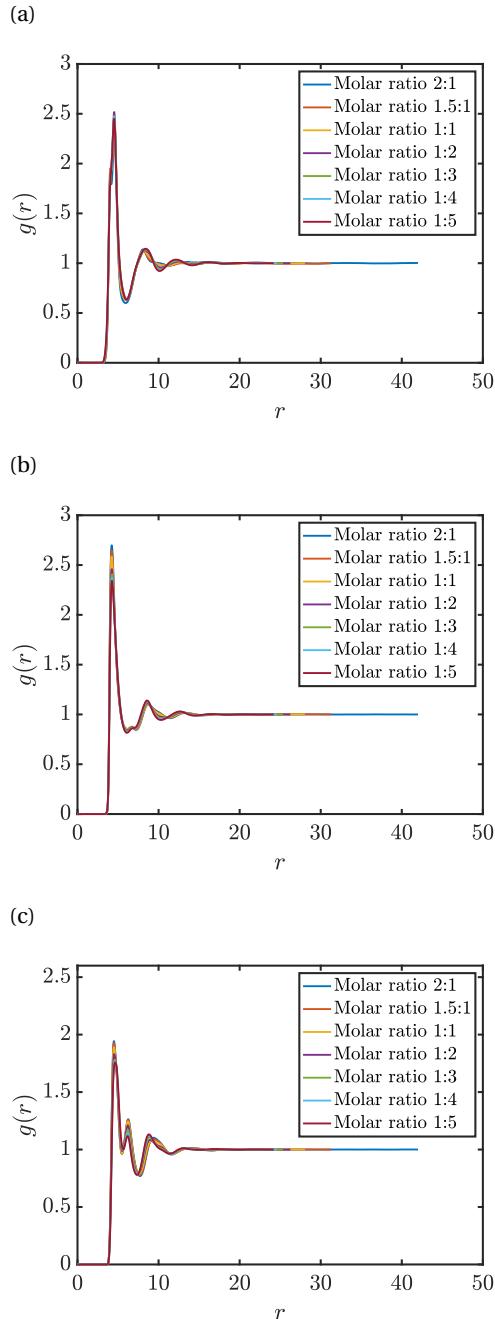


Figure 5.1: The functions  $g(r)$  for (a) urea–urea (b) urea– $\text{ChCl}$ , and (c)  $\text{ChCl–ChCl}$  for urea– $\text{ChCl}$  interactions at  $T = 343.15\text{ K}$ ,  $P = 1\text{ atm}$ , and various molar ratios of  $\text{ChCl}$  to urea. Details of the MD simulations are provided in section 5.2.4 and Table 5.1.

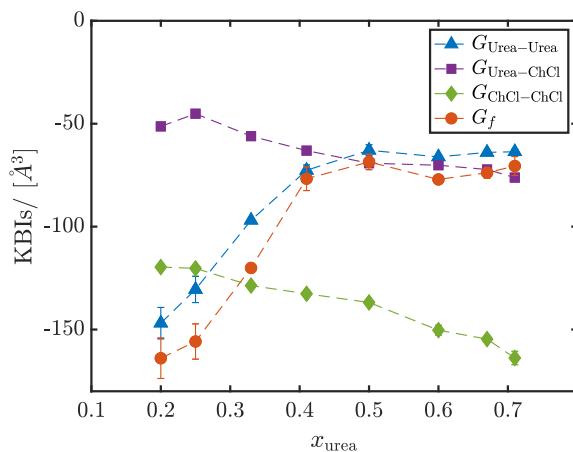


Figure 5.2: KBIs of urea–ChCl as a function of the mole fraction of urea at  $T = 343.15 \text{ K}$  and  $P = 1 \text{ atm}$ . Details of the MD simulations are provided in section 5.2.4 and Table 5.1. To extrapolate to the thermodynamics limit, the scaling of  $LG_{\alpha\beta}^V$  with  $L$  was used.

### 5.3.2. THERMODYNAMIC PROPERTIES OF UREA-CHOLINE CHLORIDE MIXTURES

#### DENSITY

Densities of ChCl and urea mixtures at 343.15 K and 1 atm are shown in Figure 5.3 as a function of the mole fraction of urea (Eq. 5.1). The density of the mixture increases as the mole fraction of urea increases. This trend was expected due to the high density of pure urea. We performed MD simulations of pure ChCl (400 molecules) and pure urea (400 molecules) systems at 343.15 K and 1 atm. The computed density of pure ChCl is 1021.2 kg/m<sup>3</sup> and the density of pure urea is 1453.2 kg/m<sup>3</sup>. Values of the densities of mixtures of ChCl and urea are bounded between the densities of the pure components. For a mixture with a molar ratio of 1:2 ( $x_{\text{Urea}} = 0.5$ ), the computed density exhibits an excellent agreement with the density measured experimentally by Yadav et al. [169], showing deviation of less than 1.5%. To the best of our knowledge, experimental data for other mole fractions of urea are not available in literature.

#### THERMODYNAMIC FACTORS

To compute the thermodynamic factors of ChCl and urea mixtures, we use three different descriptions of KBIs: (1) the individual KBIs of  $G_{\alpha\alpha}$ ,  $G_{\alpha\beta}$  and  $G_{\beta\beta}$  computed by integrating the corresponding RDFs, (2) the term  $G_f$  computed by integrating the RDF  $g_f(r) = g_{\alpha\alpha}(r) + g_{\beta\beta}(r) - 2g_{\alpha\beta}(r)$ , and (3) KBIs using the Cortes-Huerto finite-size correction [83] (see section 3.3.3). In the first two approaches, RDFs are corrected for finite-size effects using the method reported by Ganguly and van der Vegt [167]. In the Cortes-Huerto approach, RDFs are corrected using a correction that is independent of the interparticle distance. Based on the computed RDFs, KBIs of finite subvolumes are used to estimate KBIs in the thermodynamic limit ( $G_{\alpha\beta}^{\infty}$ ) [83]. This is important to study the effect of the correction method on the computed KBIs and thermodynamic factors. While RDFs are sampled using long and multiple simulations, still, uncertainties of the computed KBIs and the term  $G_f$  are not very small. Naturally, these uncertainties reflect on the accuracy of the estimation of thermodynamic factors. In Figure 5.4, we show the thermodynamic factors of ChCl and urea at  $T = 343.15$  K,  $P = 1$  atm and various mole fractions of urea. The different methods, used here to compute thermodynamic factors, present relatively similar results. This could be explained by the fact that systems of large sizes were used to compute KBIs, which makes the choice of the RDF correction not very important. Figure 5.4 shows that all  $\Gamma$  values are found larger than 1 for all urea contents. This indicates that all mixtures of ChCl and urea studied here are not ideal, and that the interactions between urea and ChCl are more favorable than interactions between molecules of the same type (i.e. urea-urea and ChCl-ChCl). Computed thermodynamic

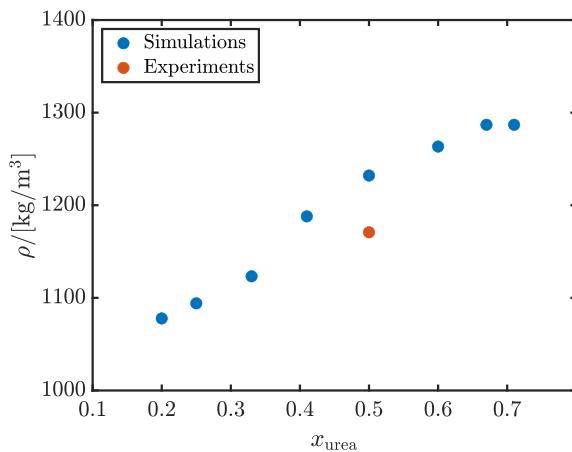


Figure 5.3: Densities of urea–ChCl as a function of the mole fraction of urea computed using MD simulations at  $T = 343.15 \text{ K}$  and  $P = 1 \text{ atm}$ . Details of the MD simulations are provided in section 5.2.4 and Table 5.1. The experimental density [169] is reported at molar ratio of urea to ChCl of 1:2 ( $x_{\text{urea}} = 0.5$ ).

5

factors exhibit no distinct trend as a function of the mole fraction of urea. The variations are small and error bars are large. Although there is no certain trend shown in Figure 5.4, we can see that thermodynamic factors become minimum at a mole fraction of urea of 0.5. This can be attributed to the relatively increased affinity between urea–urea molecules at this mole fraction.

### PARTIAL MOLAR VOLUMES

The partial molar volume of a component represents the change of volume as a result of the addition of the same component in the mixture at a fixed temperature, pressure and number of molecules. The relation between KBIs and partial molar volumes is provided in Eq. (1.6). In Figure 5.5, we show the partial molar volumes of urea and ChCl as a function of the mole fraction of urea. As the mole fraction of urea increases, the partial molar volume of urea approaches the molar volume of pure urea. For both components, the change in partial molar volumes with the increasing mole fraction of urea is not significant. This is interesting because the results of RDFs and KBIs show that the interactions between urea and ChCl are significantly affected by the change of the urea content. In Figure 5.5, we also show the molar volumes of pure urea and pure ChCl, calculated from separate MD simulations of pure systems. Molar volumes for mixtures are consistent with those of pure components.

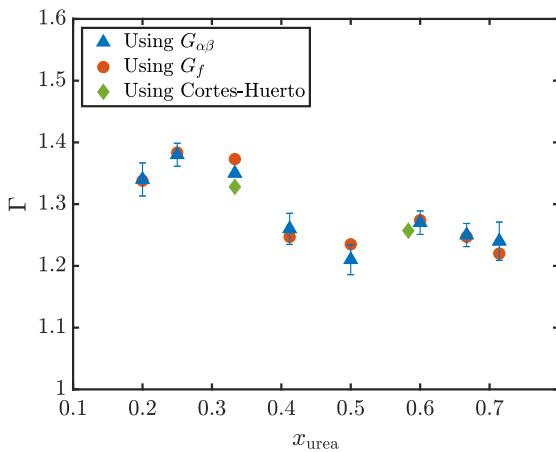


Figure 5.4: Thermodynamic factors  $\Gamma$  (Eq. (1.14)) of mixtures of urea–ChCl as a function of the mole fraction of urea at  $T = 343.15$  K and  $P = 1$  atm. Details of the MD simulations are provided in section 5.2.4 and Table 5.1. To compute  $\Gamma$ , KBIs were estimated using: individual KBIs  $G_{\alpha\beta}$ , the term  $G_f$  (in the last two methods, the Ganguly and van der Vegt [167] was applied), and the method of Cortes-Huerto [83].

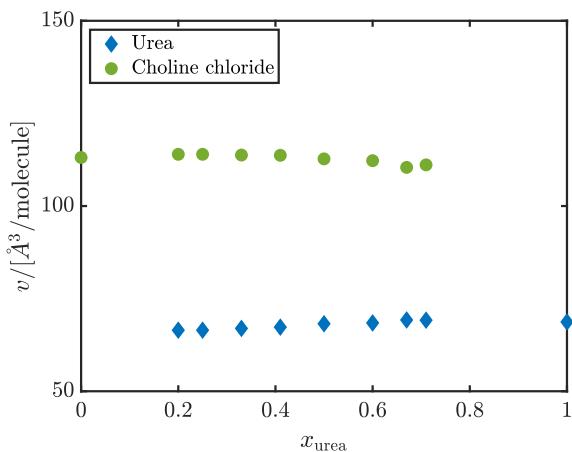


Figure 5.5: Partial molar volumes of mixtures of urea–ChCl as a function of the mole fraction of urea at  $T = 343.15$  K and  $P = 1$  atm. Details of the MD simulations are provided in section 5.2.4 and Table 5.1.

### 5.3.3. TRANSPORT PROPERTIES OF UREA-CHOLINE CHLORIDE MIXTURES

In Figure 5.6, viscosities computed from MD simulations are compared to the experimental viscosity [169], which is available for a molar ratio of urea to ChCl of 1:2 ( $x_{\text{urea}} = 0.5$ ). For this molar ratio, the viscosity from MD simulations deviates from the experimental viscosity by ca. 9%. Also, Figure 5.6 shows the viscosities of mixtures of urea and ChCl at various mole fractions of urea computed using MD simulations. The viscosity of the mixture decreases almost linearly with the mole fraction of urea. Low viscosities result in a higher mobility and as a result, larger diffusivities are observed with the addition of more urea to the mixture. From MD simulations, self diffusion coefficients  $D_{\text{self}}$  of urea, choline and chlorine were computed and corrected for finite-size effects using the Yeh-Hummer correction [72, 73, 97]. In Figure 5.7 (a), the self diffusion coefficients  $D_{\text{self}}$  of urea, choline and chlorine are shown as a function of the mole fraction of urea. As expected, self diffusion coefficients of all species increase with the addition of more urea to the mixture. This is due to the fact that urea diffuses faster than choline and chloride as shown in Figure 5.7 (a). From the computed viscosities and thermodynamic factors, it is possible to correct the binary diffusion of urea-ChCl mixtures for finite-size effects using the correction proposed by Jamali and co-workers [72, 94]. The corrected MS and Fick diffusivities are presented in Figure 5.7 (b) for urea and ChCl at different mole fractions. The Fick diffusivities are computed from the MS diffusion coefficients and the thermodynamic factors. As in the case of self diffusivities, the binary diffusion of urea and ChCl mixtures increases as the mole fraction of urea increases.

Another transport property that is of interest for salt solutions is the ionic conductivity. Ionic conductivities can be estimated from the Nernst-Einstein (NE) equation [170],

$$\kappa = \frac{e^2}{k_B T V} \sum_i N_i q_i^2 D_{\text{self},i} \quad (5.31)$$

where  $q_i$  is the charge of the molecule and  $e$  is the elementary charge.  $D_{\text{self},i}$  is the self diffusion coefficient of component  $i$  computed from MD simulations. In Figure 5.8, ionic conductivities of urea and ChCl mixtures are presented. Larger diffusion was observed at high mole fractions of urea, but ionic conductivities remain relatively constant with composition. This is due to the fact that fewer ions are present in the system when more urea is added.

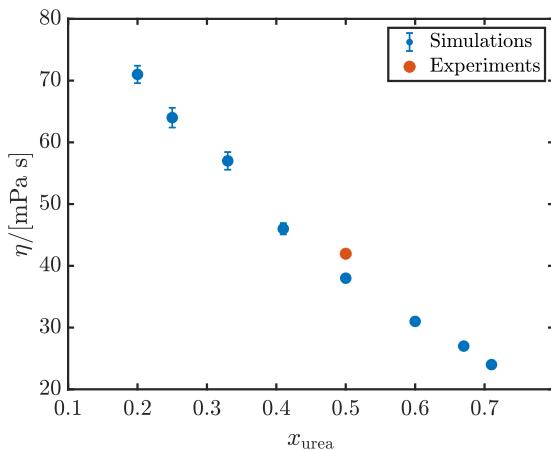


Figure 5.6: Viscosities of mixtures of urea–ChCl at  $T = 343.15 \text{ K}$  and  $P = 1 \text{ atm}$  as a function of the mole fraction of urea. Details of the MD simulations are provided in section 5.2.4 and Table 5.1. The experimental viscosity [169] is reported at a molar ratio of urea to ChCl of 1:2 ( $x_{\text{urea}} = 0.5$ ).

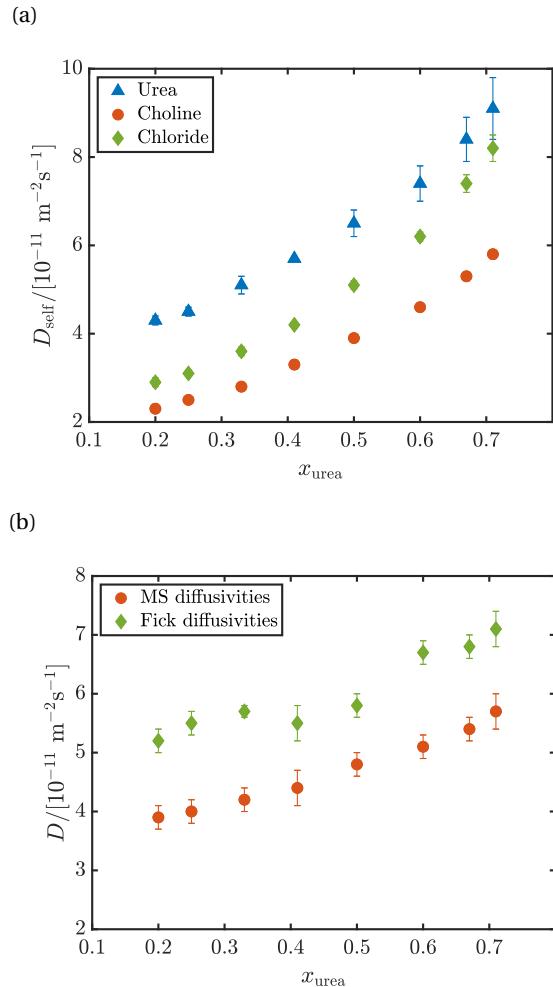


Figure 5.7: (a) Self diffusion coefficients of urea, choline, and chloride and (b) Collective diffusivities of pseudo-binary mixtures of urea–ChCl as a function of the mole fraction of urea at  $T = 343.15$  K and  $P = 1$  atm. In both (a) and (b), diffusion coefficients are corrected for finite-size effects [72]. Details of the MD simulations are provided in section 5.2.4 and Table 5.1.

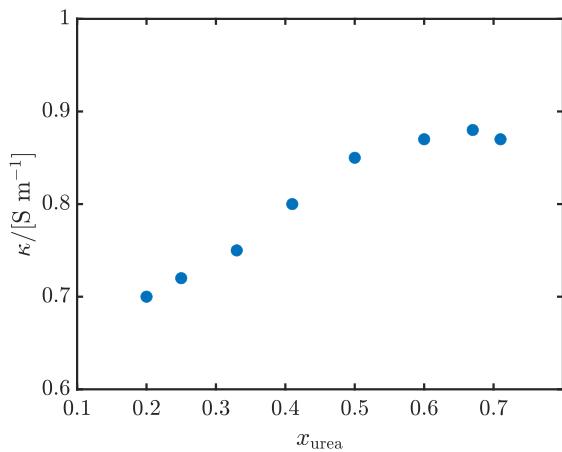


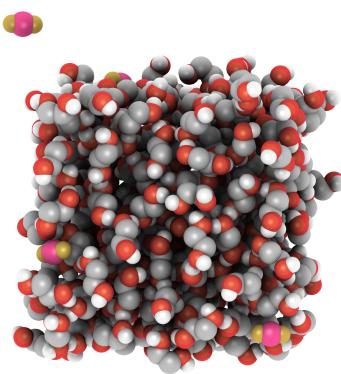
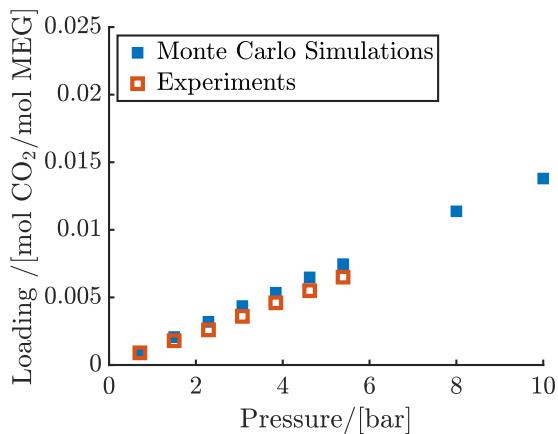
Figure 5.8: Ionic conductivities of mixtures of urea–ChCl as a function of the mole fraction of urea at  $T = 343.15 \text{ K}$  and  $P = 1 \text{ atm}$ . Errorbars are smaller than the marker size. Details of the MD simulations are provided in section 5.2.4 and Table 5.1.

## 5.4. CONCLUSIONS

In this chapter, properties of mixtures of ChCl and urea were computed at different molar ratios. MD simulations in the *NPT* ensemble were carried out to compute RDFs of systems of urea, choline and chloride. RDFs were corrected for finite size effects using the van der Vegt and Gaungly correction [167]. To compute KBIs, the system was treated as pseudo-binary were ions are indistinguishable. KBIs in the thermodynamic limit were estimated and used to examine the affinity between the components of the mixture. As the mole fraction of urea increases, the affinity between pairs of urea-urea increases. Also, urea-ChCl and ChChl interactions are weakened as a result of increasing the content of urea. Other than studying molecular interactions, KBIs were used to compute thermodynamic factors and partial molar volumes. Partial molar volumes were hardly affected by the change in the mole fraction of urea in sharp contrast to thermodynamic factors and transport properties. From MD simulations, the following transport properties were also computed as a function of composition: viscosity, self diffusion coefficients, MS diffusivities and ionic conductivities. With higher mobility of urea, choline and chloride, the viscosity of the mixtures decreased with increasing mole fraction of urea. Self and binary diffusivities were found to increase by increasing the content of urea.

# 6

## SOLUBILITY OF GASES IN MONOETHYLENE GLYCOL



This chapter is based on the paper: N. Dawass, R. R. Wanderley, M. Ramdin, O. A. Moulton, H. K. Knuutila and T.J.H. Vlugt, *Solubility of Carbon Dioxide, Hydrogen Sulfide, Methane, and Nitrogen in Monoethylene Glycol; Experiments and Molecular Simulation*. Journal of Chemical & Engineering Data, **66**, 524–534 (2020) [171].

## 6.1. INTRODUCTION

Monoethylene glycol (MEG) is a colorless, low-volatility, and stable liquid. MEG is fully miscible in water as well as in many organic liquids such as acetone and methanol [172]. In 2020, the global market size of MEG is valued at USD 20 billion [173]. MEG is widely used as an anti-freeze agent, coolant, heat transfer agent, and as a raw material for the manufacturing of polyester fibers. [174]. In the oil-and-gas industry, MEG is widely used for the prevention of gas hydrate formation [174, 175]. In the course of mitigating gas hydrate formation, MEG has been reported to absorb acid gases such as carbon dioxide ( $\text{CO}_2$ ) and hydrogen sulfide ( $\text{H}_2\text{S}$ ) [176]. Due to the absorption capability, stability, and miscibility of MEG in many organic liquids, it is also considered for use in separation processes for acid gases [176–179].

A number of MEG-based solvents, such as deep eutectic solvents, are considered for  $\text{CO}_2$  capture [179–183]. More recently, mixtures made from MEG, amines and water are investigated for simultaneously preventing hydrate formation and removing  $\text{H}_2\text{S}$  in offshore oil-and-gas applications [184]. To achieve these purposes, triethylene glycol (TEG)-amine-water mixtures were previously used. Replacing TEG with the less viscous MEG is expected to improve the absorption capability of glycols-amine-water solvents since absorption rates increase with lower viscosities [185]. To design and optimise processes in which MEG acts as a hydrate formation inhibitor or as an absorbent, knowledge of phase equilibria is essential [178, 185]. To this purpose, a number of experimental measurements of binary mixtures of MEG and gases, i.e.,  $\text{CO}_2$  and  $\text{H}_2\text{S}$  have been performed [177, 178, 186]. For a review of experimental studies on the solubility of acid gases in MEG the reader is referred to Refs. [178, 187].

While traditionally phase equilibria data are obtained from experimental measurements, such an approach is not always feasible, especially if high pressures and/or temperatures are required, and when dangerous gases, such as  $\text{H}_2\text{S}$  are involved. For this reason, theoretical approaches for computing the phase equilibria of mixtures of gases and liquids have been widely-used [1, 188–190]. Unlike classical thermodynamic models, molecular-based methods account for the strong molecular interactions present in associating liquids [176]. The past few decades, molecular simulation has emerged as a powerful tool for using microscopic information of associating liquids to predict their macroscopic behavior [22, 23]. In addition to providing thermodynamic and transport data, molecular simulation can also be used to investigate the microscopic structure of solvents, and to understand absorption mechanisms [5, 148].

The KB theory can be applied to study the absorption of gases in solvents. KBIs relate to fluctuations in the grand-canonical ensemble, from which a number of thermodynamic properties are obtained such as derivatives of the chemi-

cal potential which can be used to compute isotherms. Alternatively, solubilities of gases in solvents are directly computed using MC simulations in the grand-canonical and osmotic ensembles. MC simulations have been used to predict the solubility of gases in associating liquids [59, 148, 149]. MC simulations have been also widely-used to study the absorption of gases in solvents such as alcohols [149, 191, 192], ionic liquids [193, 194], and deep eutectic solvents [143]. To the best of our knowledge, studies reporting MC simulations of the phase equilibria of small gas molecules and MEG are lacking. A possible reason for the absence of such studies is the fact that the simulation of dense liquids with strong intermolecular interactions, as in the case of MEG, is computationally demanding. MC simulations in open ensembles are often used to compute the solubility of solutes in liquids. In these ensembles, the solutes are added to or removed from the simulation box. For dense liquids and/or with the presence of strong interactions, such insertions can be challenging [58, 60]. To enhance the efficiency of molecular transfers in MC simulations, Shi and Maginn [59, 195] developed the CFCMC (Continuous Fractional Component Monte Carlo) simulation method. In this method, the system is expanded using a so-called fractional molecule with a coupling parameter  $\lambda$ , which is used to vary the interactions between the fractional molecule and the surrounding molecules. In solubility calculations, a fractional molecule is used to gradually add/remove molecules to/from the solvent [196]. The presence of a fractional molecule does not affect the prediction of thermodynamic properties of the system [60, 197]. For a detailed discussion of the CFCMC method the reader is referred to the recent review by Rahbari et al. [62].

A prerequisite for successful MC simulations of pure and multi-component mixtures is the use of force fields that can adequately represent inter- and intramolecular interactions. Thus, another challenge of simulating gases in associating liquids, such as MEG, is the availability of force fields that provide accurate predictions of the desired properties. One of the most commonly-used force fields for a large number of gases and liquids is the Transferable Potentials for Phase Equilibria (TraPPE) force field [198, 199]. The TraPPE force field has been successfully used for the prediction of thermodynamic and transport properties of gases and liquids [149, 192, 193, 200–202, 202–205]. Cardona et al. [206] used TraPPE and other classical force fields to compute thermodynamic properties of pure MEG. The authors found that the united-atom version of TraPPE (TraPPE-UA) [199] is able to accurately predict thermodynamic properties of pure MEG, such as the density, isothermal compressibility, and heat of vaporisation.

In this chapter, we aim to predict the solubilities of small gas molecules ( $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{N}_2$ , and  $\text{H}_2\text{S}$ ) in MEG using CFCMC simulations in the osmotic ensemble. To model MEG and the gases studied in this work (Ref. [171]), the TraPPE-UA

force field is used, without any modifications. In the case of the solubility of H<sub>2</sub>S, we test the performance of an additional force field developed by Kristóf and Liszi [207] which was used in a number of solubility studies [208, 209]. Solubility calculations are validated by comparing with experimental data. For this purpose, we have collaborated with the research group of Prof. Hanna Knuutila at The Norwegian University of Science and Technology to experimentally measure the solubility of CO<sub>2</sub> in MEG. In their labs, calorimetric measurements were performed using a CPA 122 calorimeter purchased from ChemiSens AB. In this device, gas from a compartment with a fixed known volume is absorbed to the sample. By accurately measuring the changes in pressure and temperature of the gas in this compartment and using the Peng-Robinson equation of state [210], the amount of absorbed material in the sample can be calculated. For details related to the experimental setup and methods, the reader is referred to Ref. [171]. For the binary system CO<sub>2</sub>–MEG, absorption isotherms from MC simulations and experiments are compared for the temperatures 333.15 K, 353.15 K, and 373.15 K. The solubilities of CH<sub>4</sub>, N<sub>2</sub>, and H<sub>2</sub>S in MEG are computed at 373.15 K and compared to experimental data from literature, when available. In addition to absorption isotherms, Henry coefficients are computed using CFCMC simulations. From the knowledge of Henry coefficients of different solutes in MEG, selectivities are also computed.

## 6

The chapter is organised as follows. In section 6.2, the MC simulation methods used to compute the solubilities in MEG are explained. Results are presented and discussed in section 6.3. In section 6.3.1, MC calculations of absorption isotherms of CO<sub>2</sub> in MEG are compared to the experimental data obtained from Ref. [171]. MC simulations results of the solubility of CH<sub>4</sub>, N<sub>2</sub>, and H<sub>2</sub>S in MEG are shown in section 6.3.2. In section 6.4, the main findings of this chapter are summarized.

## 6.2. METHODS

### 6.2.1. FORCE FIELDS

Classical force fields were used to describe the interactions of the molecules studied in this work. For MEG, all interaction potentials and parameters follow from TraPPE-UA force field [199]. The TraPPE force field adequately predicts densities and vapor-liquid equilibria (VLE) of many species such as normal alkanes [198], branched alkanes [211], glycols, and ketones [199]. To accurately represent the molecular structure of MEG, Stubbs et al. [199] added an additional repulsive term ( $r^{-12}$ ) for interactions between a hydroxyl hydrogen and an oxygen atom separated by four bonds. In our study, the TraPPE–UA force field was also used to represent CO<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>S, and N<sub>2</sub> as rigid objects. We also tested an-

Table 6.1: Chemical formulas and force fields of the components simulated in this work.

Component	Chemical formula	Force field
Monoethylene glycol	$\text{HO}(\text{CH}_2)_2\text{OH}$	TrAPPE-UA [199]
Carbon dioxide	$\text{CO}_2$	TrAPPE-UA [212]
Methane	$\text{CH}_4$	TrAPPE-UA [198]
Hydrogen sulfide	$\text{H}_2\text{S}$	TrAPPE-UA [213]
Hydrogen sulfide	$\text{H}_2\text{S}$	Kristóf and Liszi [207]
Nitrogen	$\text{N}_2$	TrAPPE-UA [212]

other four-site model presented by Kristóf and Liszi [207] for  $\text{H}_2\text{S}$ , here referred to as  $\text{H}_2\text{S}$ -KL. The main differences between the two force fields are the non-bonded LJ parameters and the atomic charges. Table 6.1 lists the components simulated in this study and the force field used for each component. All force field parameters are listed in the Supporting Information of Ref. [171].

In our simulations, two types of intermolecular interactions are computed: LJ and Coulombic interactions. LJ interactions were truncated at 12 Å and the uncertainty due to truncation is handled by applying analytic tail corrections [22, 23]. The Lorentz–Berthelot mixing rules were used for LJ interactions between dissimilar interaction sites [22, 23]. The Ewald summation method was applied to handle electrostatic interactions with a relative precision of  $10^{-6}$ . The real-space part of the electrostatic interactions was truncated at 12 Å. All simulations were carried out in the osmotic ensemble (see section 6.2.2). The PC-SAFT equation of state was used to compute the fugacity of the solutes at the desired temperatures and pressures [214, 215].

### 6.2.2. THE CFCMC METHOD IN THE OSMOTIC ENSEMBLE

The osmotic ensemble [22] is used to compute the solubility of small solute molecules in non-volatile solvents. In this open ensemble, the following parameters are fixed: the temperature ( $T$ ), the imposed hydrostatic pressure ( $P$ ), the number of molecules of MEG ( $N_{\text{MEG}}$ ), and the fugacity of the solute ( $f$ ). The number of molecules of the solute ( $N_s$ ), and the volume of the system ( $V$ ) are varied to achieve equilibrium. The hydrostatic pressure  $P$  inside the simulation box corresponds to the imposed fugacity  $f$  of the reservoir. An essential part of the calculations is the insertion and/or deletion of solute molecules in the simulation box. When studying dense solvents, as in the case of MEG, molecule insertions can be challenging [58, 60]. To improve the probability of accepting insertion/deletion trial moves, the CFCMC method was used [59, 60, 216, 217]. The osmotic ensemble was expanded using a so-called fractional molecule [59].

As opposed to a whole molecule, the strength of interactions of a fractional molecule is varied using a coupling parameter  $\lambda$ . When  $\lambda = 0$ , the fractional molecule acts as an ideal gas molecule and does not interact with the surrounding molecules. When  $\lambda = 1$ , the fractional molecule fully interacts with the surrounding molecules. By varying the strength of interactions of the fractional molecule with the surrounding molecules, whole molecules can be gradually added or removed. Besides the conventional MC thermalization trial moves [22, 23], trial moves attempting to change  $\lambda$  are required. Shi and Maginn [59] derived Metropolis-like acceptance rules for changing the values of  $\lambda$  in the osmotic ensemble. For each solute type, a fractional molecule is used to insert/delete molecules in the simulation box. When  $\lambda$  drops below zero, the fractional molecule is removed and a randomly selected whole molecule is transformed into a fractional molecule. When  $\lambda$  is larger than 1, the fractional molecule is transformed into a whole molecule and a new fractional molecule is inserted [59, 195]. For the solvent, a fractional molecule is used to improve sampling, while keeping the total number of molecules of the solvent fixed. For a fractional molecule of the solvent type,  $\lambda$  trial moves involve changing the values of  $\lambda$ , random reinsertions of the fractional molecule, and identity changes between a fractional molecule and a whole molecule [197]. In CFCMC simulations, the system is biased to ensure a flat distribution of the observed probability of  $\lambda$ . From the probability distribution of  $\lambda$ , the excess chemical potential and hence the Henry coefficient is obtained. For more details the reader is referred to Refs. [59, 60, 196, 197].

### SCALING OF THE INTERMOLECULAR INTERACTIONS

For each fractional molecule, a weightfunction  $W(\lambda)$  is constructed to achieve a flat probability distribution of  $\lambda$  and ensure smooth transitions between  $\lambda = 0$  and  $\lambda = 1$  [60]. Essentially, at a certain  $\lambda$ , the value of  $W(\lambda)$  counteracts the value of  $\langle \partial U / \partial \lambda \rangle$ , which is the average potential energy change with  $\lambda$ . Fluctuations in the value of  $\partial U / \partial \lambda$  can be large, which can hinder the efficient sampling of the  $\lambda$  space. As a result, a pathway that minimizes the variance of  $\partial U / \partial \lambda$  has to be chosen [218]. Changes in the values of  $\lambda$  depend on how intermolecular interactions, are scaled when  $\lambda$  is varied [218] (intramolecular interactions are not scaled). Electrostatic interactions are scaled by using a scaling parameter  $\lambda_{\text{el}}$  that is a function of  $\lambda$  [196, 219]. For non-bonded LJ interactions, the following soft-core scaling potential is used [218, 220, 221]:

$$u(r_{ij}, \lambda) = 4\epsilon_{ij}\lambda^a \left[ \left( \frac{1}{\alpha(1-\lambda)^b + \left( \frac{r_{ij}}{\sigma_{ij}} \right)^c} \right)^{12/c} - \left( \frac{1}{\alpha(1-\lambda)^b + \left( \frac{r_{ij}}{\sigma_{ij}} \right)^c} \right)^{6/c} \right] \quad (6.1)$$

where  $i$  and  $j$  are the interaction sites.  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the LJ parameters.  $r_{ij}$  is the distance between  $i$  and  $j$ . The parameters  $a$ ,  $b$ ,  $c$  and  $\alpha$  are adjusted to achieve an efficient sampling of the  $\lambda$  space. For systems composed of MEG molecules, a number of scaling potentials were tested. Figure 6.1 shows the values of  $\lambda$  of an MEG fractional as a function of the number of MC cycles, for three scaling potentials. The commonly used  $(a-b-c) = 1-2-6$  potential results in poor sampling of the  $\lambda$ -space. Figure 6.1 (a) shows that at certain periods, the values of  $\lambda$  are confined to a limited range. Changing the parameter  $b$  from 2 to 1 improves the sampling as demonstrated in Figure 6.1 (b). Figure 6.1 (c) shows that the 1–1–48 potential with  $\alpha = 0.0025$  that was recommended earlier by Pham et al. [218], results in the best sampling.

### 6.2.3. SIMULATION DETAILS

Molecular simulations were performed using the recently developed open-source software package Brick-CFCMC [196]. The density of pure MEG was computed in the *NPT* ensemble at 1 bar, and at temperatures 333.15 K and 353.15 K. The solubility of CO<sub>2</sub> in MEG was computed at three temperatures,  $T = 333.15$  K, 353.15 K, and 373.15 K. The solubilities of CH<sub>4</sub>, N<sub>2</sub>, and H<sub>2</sub>S in MEG were computed at  $T = 373.15$  K. For all gases, solubilities were computed at pressures ranging from 1 to 10 bar, but in the case of N<sub>2</sub> pressures up to 100 bar were considered since N<sub>2</sub> has very low solubilities in MEG at low pressures. Simulation boxes were set up with 250 to 350 MEG molecules, depending on the number of absorbed solute molecules. Two fractional molecules were used: a fractional molecule to insert/remove solute molecules into the simulation box, and a MEG fractional molecule. The following MC trial moves were used: translations, rotations, and volume change trial moves. MC trial moves that attempt to change the values of  $\lambda$  were used for both fractional molecules. Simulations in the osmotic ensemble were carried out with the following probabilities for selecting trial moves: 25% translations, 25% rotations, 32% trial moves to change the internal configuration of MEG [196], 1% volume changes, 17% trial moves to change  $\lambda$  divided equally between the solute and MEG fractional molecules. A minimum of  $10^6$  cycles was carried out for equilibration. At each MC cycle, the number of the trial moves performed equals the number of molecules of the system.

During equilibration, an iterative scheme was used to obtain a weight function  $W(\lambda)$  that results in a flat probability distribution of  $\lambda$ . For production runs,  $10^6$  cycles were carried out. To minimise the statistical error of the computed averages, a number of independent production simulations was performed at a specified  $T$  and  $P$ . The number of simulations performed was selected such that

the uncertainty is less than 5%. For each system simulated in this work, at least 25 production runs were carried out. Error bars were computed by dividing these runs into five groups, and calculating the standard deviation.

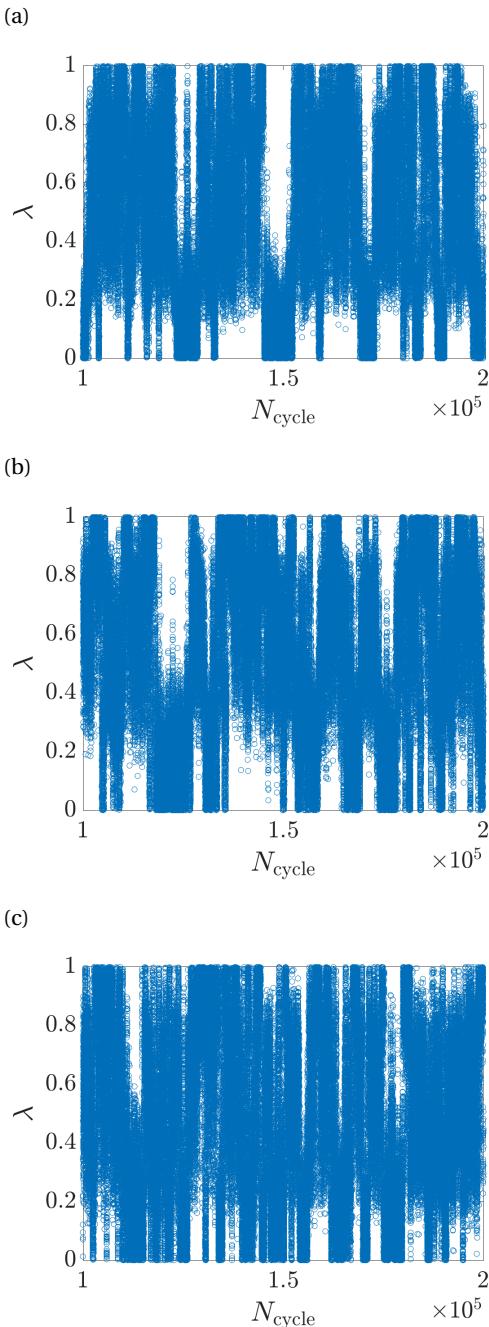


Figure 6.1: The values of  $\lambda$  vs. the number of MC cycles of a MEG fractional molecule in *NPT* MC simulations. To scale the interaction of the fractional molecules, the following scaling potentials are used (a–b–c) (see Eq. (6.1)): (a) 1–2–6, (b) 1–1–6, and (c) 1–1–48.

### 6.3. RESULTS AND DISCUSSION

#### 6.3.1. SOLUBILITY OF CO<sub>2</sub> IN MONOETHYLENE GLYCOL

Densities of pure MEG were computed using MC simulations in the *NPT* ensemble at  $P = 1$  bar, and at  $T = 333.15$  K and  $T = 353.15$  K. In Table 6.2, densities (reported in units of kg/m<sup>3</sup>) from simulations are compared to experimental values from the work of Skylogianni et al. [184]. Table 6.2 shows that when using the TraPPE-UA force field, simulations underpredict densities of MEG. The differences between experiments and simulations are around 5%. Simulating a solvent with an underestimated density can result in higher absorption capacity. Deviations between experiments and simulations will be discussed in detail later in this section.

As a fractional MEG molecule is present in the simulation, we can calculate the excess chemical potential of MEG  $\mu^{\text{ex}}$  from the probability distribution of its  $\lambda$  parameter [60, 196]. The chemical potential of MEG in the liquid phase equals [196]:

$$\mu^L = \mu^o + RT \ln \frac{\rho^L}{\rho^o} + \mu^{\text{ex}} \quad (6.2)$$

6

where  $\rho^L$  is the number density of MEG and  $\mu^o$  is the reference chemical potential, which only depends on the temperature. Eq. (6.2) also applies to MEG in the gas phase. At equilibrium, the chemical potentials in the liquid phase and gas phase are equal. If we assume an ideal gas phase, then  $\mu^{\text{ex}}$  in the gas phase equals zero and  $\rho^L = P^{\text{sat}} / k_B T$ . From this, the saturated vapor pressure  $P^{\text{sat}}$  of MEG can be estimated by

$$P^{\text{sat}} = k_B T \rho^L \exp \left[ \frac{\mu^{\text{ex}}}{k_B T} \right] \quad (6.3)$$

For a detailed derivation of Eq. (6.3), see the Supporting Information of Ref. [171]. In Table 6.2, we report the excess chemical potential, vapor densities and saturated vapor pressures of pure MEG. The vapor pressures computed from MC simulations are compared to experimental values obtained from NIST database [222]. Table 6.2 shows that both methods are in good agreement. The pressures reported in Table 6.2 can be considered very small, which validates the assumption made by the experimental method regarding the non-volatility of MEG.

Figure 6.3 shows absorption isotherms of CO<sub>2</sub> in MEG, from experiments and MC simulations in the osmotic ensemble, at temperatures  $T = 333.15$  K,  $T = 353.15$  K, and  $T = 373.15$  K. In Figure 6.2, a typical MC simulation snapshot of MEG and CO<sub>2</sub> molecules is shown. Figure 6.2 shows that CO<sub>2</sub> molecules are dispersed in MEG and not clustered. In Figure 6.3, the ratio of the number of moles

Table 6.2: Properties of pure MEG at different temperatures from experiments and MC simulations. Density of pure MEG in kg/m<sup>3</sup> obtained from experiments [184] and MC simulations at  $P = 1$  bar. The vapor densities and saturated vapor pressures of MEG are computed from  $\mu^{\text{ex}}/(k_B T)$  (see section 6.3.1). Experimental saturated vapor pressures are obtained from the NIST database [222].

$T/\text{[K]}$	$\rho_{\text{exp}}^{\text{L}}/\text{[kg/m}^3]$	$\rho_{\text{sim}}^{\text{L}}/\text{[kg/m}^3]$	$\mu^{\text{ex}}/k_B \text{[K]}$	$\rho_{\text{sim}}^{\text{V}}/\text{[kg/m}^3]$	$P_{\text{sim}}^{\text{sat}}/\text{[bar]}$	$P_{\text{NIST}}^{\text{sat}}/\text{[bar]}$
333.15	1085	$1029.1 \pm 20$	$-4125 \pm 41$	$4.2 \times 10^{-6}$	0.022	0.021
353.15	1070.1	$1016.3 \pm 0.5$	$-3948 \pm 25$	$1.4 \times 10^{-5}$	0.0068	0.0067
373.15	–	$1003.3 \pm 0.5$	$-3749 \pm 27$	$4.3 \times 10^{-5}$	0.0019	0.0018

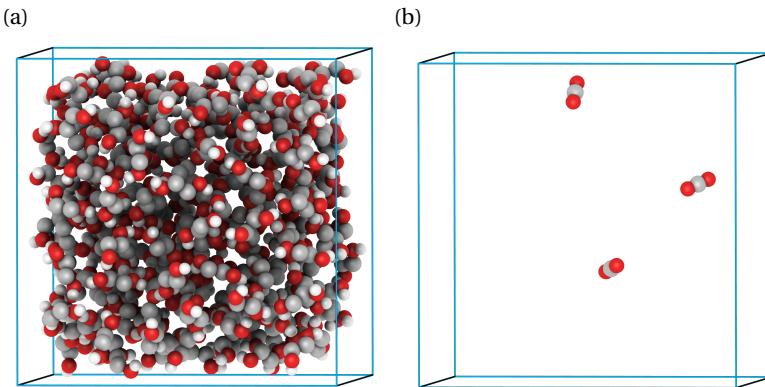


Figure 6.2: (a) Typical snapshot of a simulation of MEG in which CO<sub>2</sub> is absorbed in the osmotic ensemble ( $T = 333$  K,  $P = 8$  bar,  $N_{\text{MEG}} = 220$  molecules, and  $N_{\text{CO}_2} = 3$  molecules), in a simulation box with the dimensions  $28 \times 28 \times 28$  Å. (b) The same snapshot as in (a) but only showing CO<sub>2</sub> molecules. Figures were produced using the software package iRASPA [223].

of CO<sub>2</sub> to the number of moles of MEG (i.e. loading) is plotted as a function of pressure. Solubilities from MC simulations were found to qualitatively agree with experimental measurements. Both experiments and simulations show that for all temperatures studied, the loading is almost linear in this pressure range. Additionally, both methods report that the absorption of CO<sub>2</sub> in MEG decreases at higher temperatures. Figure 6.3 shows that the quantitative agreement between MC simulations and experiments varies with temperature and pressure. At very low pressures (i.e.  $< 2$  bar), loadings of CO<sub>2</sub> obtained from MC simulations agree well with experiments. At higher pressures, MC simulations overpredict the absorption of CO<sub>2</sub> when compared to experiments. The deviation between simulations and experiments systematically increases with pressure, and decreases with temperature. The inherent uncertainties of the experimental loadings of CO<sub>2</sub> in MEG are shown with vertical error bars. Uncertainties of experimental values are calculated using the methodology described in the work of Wanderley

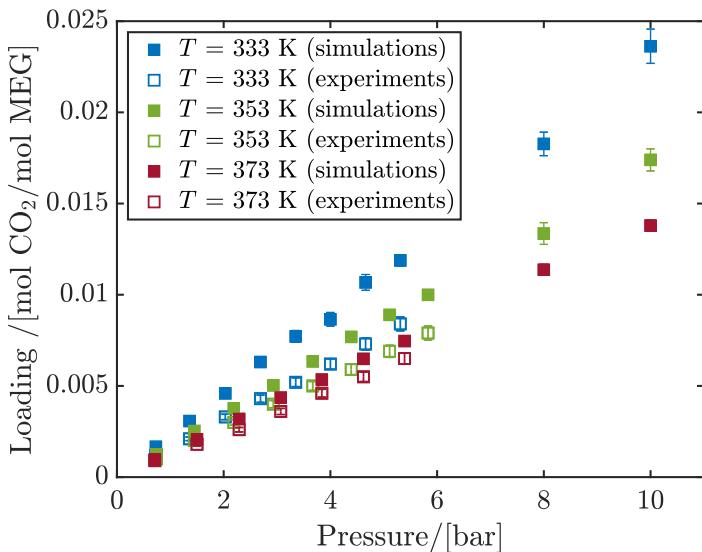


Figure 6.3: Absorption isotherm of  $\text{CO}_2$  in MEG at various temperatures. Details of MC simulations are provided in section 6.2.

6

and co-workers [224]. The inherent uncertainties of the total pressure are limited by the sensitivity of the pressure transducer used for these measurements, which is  $\pm 0.015$  bar. Conversely, this means that the uncertainty of the estimated  $\text{CO}_2$  partial pressures is  $\pm 0.021$  bar due to error propagation. One can see in Figure 6.3 that those are very small uncertainties when considering the span of pressures measured in the series of experiments. In the Supporting Information of Ref. [171], solubility data from experiments and MC simulations are provided in a tabulated form along with their uncertainties.

In Figure 6.4, solubilities measured in this work at  $T = 373.15$  K are compared to solubilities from other experimental studies. The measurements in this work were found to match the data from Jou et al. [186] at low pressures. At higher pressures,  $\text{CO}_2$  solubility from other experimental works slightly differ from results of this work. Figure 6.4 also shows that loadings computed using MC simulations agree the most with our experimental results.

Besides absorption isotherms, it is also possible to describe solubility of gases in solvents through Henry coefficients. The Henry coefficient of solute 2 in solvent 1 is defined as [226]:

$$H_{21} = \lim_{x_2 \rightarrow 0} \frac{P_2}{x_2} = \lim_{x_2 \rightarrow 0} \frac{f_2}{x_2} \quad (6.4)$$

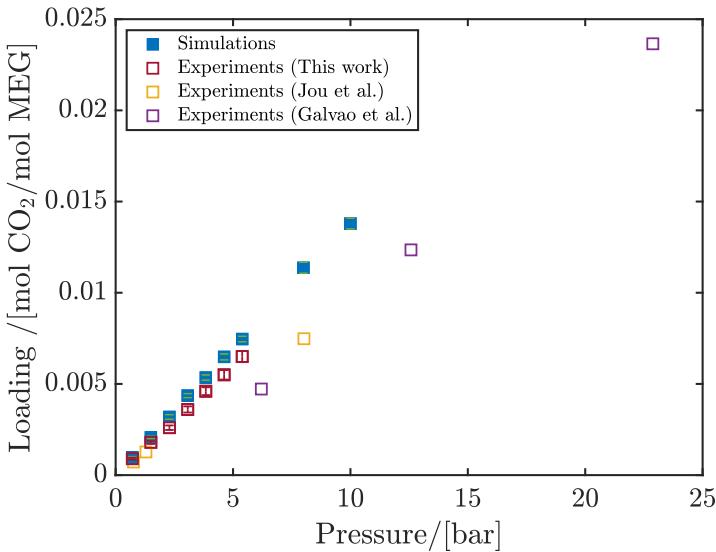


Figure 6.4: Absorption isotherm of CO<sub>2</sub> in MEG at T=373.15 K. Open symbols represent experimental data of this work, from Jou et al. [186] and from Galvao et al. [225]. Closed symbols are MC simulations results, details are provided in section 6.2.

where  $P_2$  and  $x_2$  are the partial pressure and mole fraction of solute 2, respectively.  $f_2$  is the fugacity of the solute. With these experimental values, the Henry coefficient is defined as the partial pressure of CO<sub>2</sub> in bar divided by the molar fraction of CO<sub>2</sub> in MEG. In MC simulations, Henry coefficients  $H_{21}$  are computed from the excess chemical potential of the solute  $\mu_2^{\text{ex}}$  [227],

$$H_{21} = \lim_{x_2 \rightarrow 0} k_B T \rho_1 \exp \left[ \frac{\mu_2^{\text{ex}}}{k_B T} \right] \quad (6.5)$$

In Table 6.3, Henry coefficients of CO<sub>2</sub> in MEG  $H_{\text{CO}_2, \text{MEG}}$  computed using MC simulations are reported at different temperatures and compared to Henry coefficients from experiments. Both methods demonstrate that the value of  $H_{\text{CO}_2, \text{MEG}}$  increases with temperature. The maximum difference between experimental and computed Henry coefficients is 30%. The difference consistently decreases with increasing temperature to reach 13% at  $T = 373.15$  K. Predictions from MC simulations are satisfactory considering that the force fields and the mixing rules used for MEG and CO<sub>2</sub> were not modified. The Henry coefficients reported in Table 6.3 indicate that pure MEG would not be a good absorbent for CO<sub>2</sub>. In a study of Ramdin et al. [228], Henry coefficients at  $T = 333$  K of CO<sub>2</sub> in selexol and the

Table 6.3: Henry coefficient of CO<sub>2</sub> in MEG at different temperatures obtained from experiments (this work) and molecular simulation.

$T / [\text{K}]$	$H_{\text{CO}_2, \text{MEG}}$ , bar/(mol CO <sub>2</sub> /mol EG)	
	Experimental	MC Simulations
333.15	$634 \pm 2$	$445 \pm 20$
353.15	$736 \pm 2$	$576 \pm 15$
373.15	$843 \pm 2$	$730 \pm 21$

ionic liquid [bmim][TF<sub>2</sub>N] were reported to be 68 bar and 66 bar, respectively. At the same temperature, experimental Henry coefficient of CO<sub>2</sub> in MEG is 634 bar (Table 6.3).

From the knowledge of solubility of CO<sub>2</sub> at different temperature, the heat of absorption  $q$  of CO<sub>2</sub> in MEG can be calculated using

$$\frac{-q}{R} = \frac{\partial \ln(H_{\text{CO}_2, \text{MEG}} / P_o)}{\partial(1/T)} \quad (6.6)$$

## 6

where  $R$  is the ideal gas constant and  $P_o$  is a reference pressure to make the argument of the logarithm dimensionless. Using solubilities of CO<sub>2</sub> in MEG from MC simulations of this work,  $q$  was found to be equal to -12.8 kJ/mol, indicating that the absorption of CO<sub>2</sub> is an exothermic process. This value is in good agreement with experimental findings. In a study by Wanderley et al. [229], where the heat of absorption was measured using calorimetric experiments resulting in -14 kJ/mol at 343.15 K.

The differences between theoretical and experimental solubilities can be attributed to the force field used to describe MEG. From our simulations, it is observed that the TraPPE-UA force field underpredicts the density of pure MEG (see Table 6.2). Lower MEG densities can potentially lead to higher absorption capacities of solutes. Moreover, the force field parameters of TraPPE-UA [199] were obtained using VLE experimental data of MEG at high temperatures ( $> 400$  K), as a result inaccuracies at lower temperatures can be expected as we move outside the fitting range of the TraPPE force field.

Figure 6.3 shows that deviations between experiments and simulations are larger at lower temperatures. While deviations can be reduced by optimising the force field parameters of MEG, force field parameters of the solute have to be considered as well. For CO<sub>2</sub>, TraPPE force field parameters are obtained using pure component data and not data of multi-component systems. Predictions of MC simulations can be improved by revising force field parameterisation or considering different force field combinations. Alternatively, one might consider

changes to the used combination rules. To improve the predictions of MC simulations, deviations from the Lorentz–Berthelot rules can be used. For instance, other combination rules can be considered and/or adjustable parameter(s) can be added to fine-tune solute–solvent interactions.

### 6.3.2. SOLUBILITY OF CH<sub>4</sub>, H<sub>2</sub>S AND N<sub>2</sub> IN MONOETHYLENE GLYCOL

MC simulations were used to compute the solubility of other pure gases in MEG at 373.15 K. MC simulation results were compared to experimental data from literature.

In Figure 6.5 the absorption isotherm of CH<sub>4</sub> in MEG is shown for  $T = 373.15\text{ K}$  and pressures ranging from 1 to 10 bar. At this pressure range, low loadings of CH<sub>4</sub> are obtained from MC simulations. To validate computational results, experimental solubilities [225] at higher pressures are shown in Figure 6.5. At  $P = 17.9\text{ bar}$ , MC simulations overpredicts the solubility of CH<sub>4</sub> in MEG by ca. 25%. As discussed earlier in section 6.3.1, higher absorption of solutes is due to the underestimated densities of MEG when using the TraPPE-UA force field.

In Figure 6.6, solubilities of H<sub>2</sub>S in MEG at  $T = 373.15\text{ K}$  from MC simulations using two different H<sub>2</sub>S force fields are compared to experimental solubilities from Ref. [186]. A reasonable agreement between MC simulations and experiments is obtained for the two force fields, but larger deviations appear at high pressures. The H<sub>2</sub>S-TraPPE force field underpredicts loadings of H<sub>2</sub>S, while the H<sub>2</sub>S-KL force field overpredicts loadings at the studied conditions. At atmospheric pressure, solubilities computed using H<sub>2</sub>S-KL were found to be closer to the experimental value reported by Jou et al. [186], compared to the solubility computed using H<sub>2</sub>S-TraPPE.

In Figure 6.7, the absorption isotherm of N<sub>2</sub> in MEG at  $T = 373.15\text{ K}$  is shown. The loading of N<sub>2</sub> is computed using MC simulations and is compared to experimental data from Zheng et al. [177]. Since the absorption of N<sub>2</sub> in MEG is negligible at atmospheric pressures, simulations were performed at pressures ranging from 10 bar to 100 bar. From Figure 6.7 it can be seen that the computed loadings deviate considerably from experimental values, and that the deviations increase systematically with pressure. As discussed earlier, differences between MC simulations and experimental data at high pressures data can be improved by modifying the used force fields, or fine-tune solute–solvent interactions.

In Table 6.4, Henry coefficients computed using MC simulations of CH<sub>4</sub>, H<sub>2</sub>S, and N<sub>2</sub> are listed. Experimental Henry coefficients of CH<sub>4</sub> and H<sub>2</sub>S are also shown. Average differences between experimental and computational values are around 25%. From the Henry coefficients at  $T = 373.15\text{ K}$ , the following order of

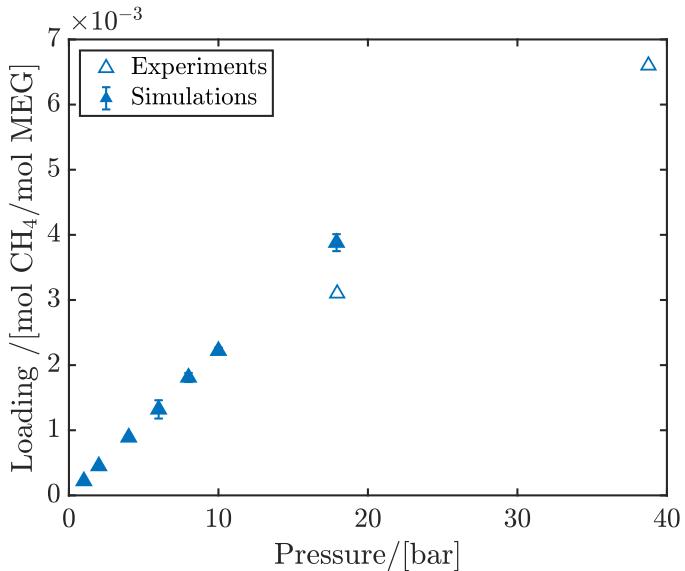


Figure 6.5: Absorption isotherm of  $\text{CH}_4$  in MEG at  $T = 373.15 \text{ K}$ . Closed symbols are solubilities from MC simulations in the osmotic ensemble (details in section 6.2) and open symbols are experimental data from Ref. [225].

6

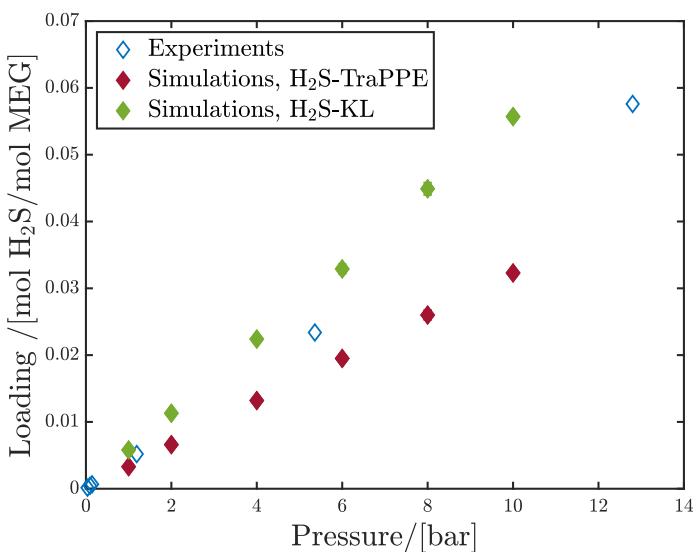


Figure 6.6: Absorption isotherm of  $\text{H}_2\text{S}$  in MEG at  $T = 373.15 \text{ K}$ . Closed symbols are solubilities from MC simulations (details in section 6.2) using two force fields:  $\text{H}_2\text{S}$ -TraPPE and  $\text{H}_2\text{S}$ -KL [207]. Open symbols are experimental data from Ref. [180].

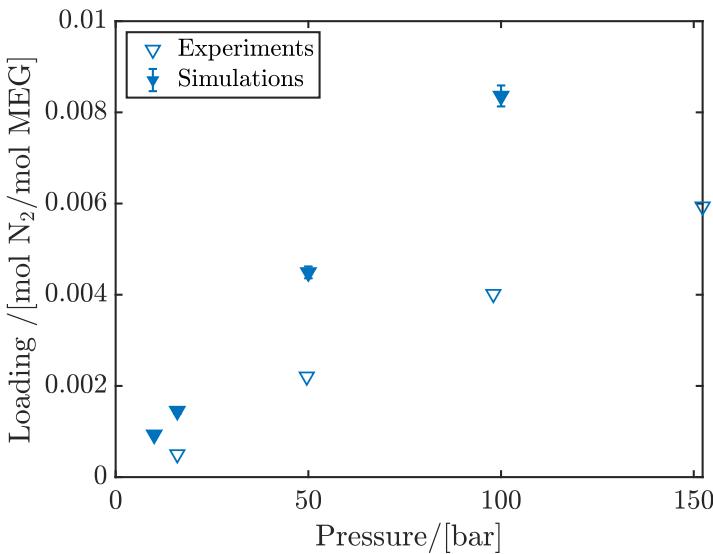


Figure 6.7: Absorption isotherm of N<sub>2</sub> in MEG at  $T = 373.15\text{ K}$ . Closed symbols are solubilities from MC simulations (details in section 6.2) and open symbols are experimental data from Ref. [177].

solubilities is exhibited: H<sub>2</sub>S > CO<sub>2</sub> > CH<sub>4</sub> > N<sub>2</sub>. From the Henry coefficients, the ideal selectivity of the desired component  $i$  from the undesired component undesired component  $j$  can be expressed as [230]

$$S_{i/l\,j} = \left[ \frac{H_j}{H_i} \right]_T \quad (6.7)$$

In Table 6.5, ideal selectivities for the separation of CO<sub>2</sub> using MEG are provided, computed using Henry coefficients from experiments and MC simulations. The results show that CO<sub>2</sub> is more selectively absorbed in MEG than CH<sub>4</sub> and N<sub>2</sub>. However, that is not the case for the separation of CO<sub>2</sub> using MEG in the presence of H<sub>2</sub>S.

Table 6.4: Henry coefficients of CH<sub>4</sub>, H<sub>2</sub>S and N<sub>2</sub> in MEG obtained from experiments and MC simulation. Experimental values are taken from Refs. [225], and [177] for CH<sub>4</sub>, and H<sub>2</sub>S, respectively.

Solute	$H_{i,j}$ , bar/(mol <i>i</i> /mol EG)	
	Experimental	MC Simulations
CH <sub>4</sub>	5673	4504 ± 165
H <sub>2</sub> S-TraPPE	227.4	302 ± 14
H <sub>2</sub> S-KL	227.4	173 ± 5
N <sub>2</sub>	–	10815 ± 248

Table 6.5: Ideal selectivities of CO<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>S and N<sub>2</sub> in MEG computed using Henry coefficients from experiments and MC simulation at *T*=373.15 K. Experimental values are taken from Refs. [225], and [177] for CH<sub>4</sub>, and H<sub>2</sub>S, respectively.

Separation	$S_{i,j}$	
	Experimental	MC Simulations
CO <sub>2</sub> /CH <sub>4</sub>	6.73	6.17
CO <sub>2</sub> /H <sub>2</sub> S-TraPPE	0.27	0.41
CO <sub>2</sub> /H <sub>2</sub> S-KL	0.27	0.24
CO <sub>2</sub> /N <sub>2</sub>	–	14.81

## 6.4. CONCLUSIONS

In this chapter, we predict the absorption of CO<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>S, and N<sub>2</sub> in MEG using MC simulations in the osmotic ensemble. The CFCMC method was used to facilitate insertions/deletions of molecules into the solvent. The TraPPE force field was used to model all species. For H<sub>2</sub>S, two force fields were compared, i.e. H<sub>2</sub>S-TraPPE and H<sub>2</sub>S-KL. The solubility of CO<sub>2</sub> in MEG was measured at the following temperatures: 333.15 K, 353.15 K, and 373.15 K. From experiments and MC simulations, CO<sub>2</sub> was found to be better absorbed at lower temperatures. At *T* = 373.15 K, CO<sub>2</sub> and H<sub>2</sub>S were found to have higher solubilities in MEG than CH<sub>4</sub> and N<sub>2</sub>. Solubilities predicted by MC simulations are in reasonable agreement with experimental data. For all the solutes studied in this work, deviations between MC simulations and experiments were found to increase with pressure. For the solubility of H<sub>2</sub>S, predictions from the H<sub>2</sub>S-KL force field were closer to experimental measurements than H<sub>2</sub>S-TraPPE. Other than absorption isotherms, Henry coefficients were also computed. The order of solubilities of the gases in MEG at 373.15 K was found to be as follows: H<sub>2</sub>S > CO<sub>2</sub> > CH<sub>4</sub> > N<sub>2</sub>. The average difference between Henry coefficients from experiments and Henry coefficients from MC simulations is ca. 20%. These results can be regarded sat-

isfactory, considering that force fields from literature were directly used without fitting binary interaction parameters. To improve predictions at high pressures, force field adjustments are required. For the solubility of CO<sub>2</sub> in MEG, the experimental data provided in this work may be used to generate new force field parameters.



# CONCLUSIONS

The Kirkwood-Buff (BF) theory is one of the most powerful theories connecting the microscopic structure of isotropic fluids to their macroscopic properties. KBIs are defined either in terms of RDFs over infinite and open volumes, or in terms of density fluctuations in the grand-canonical ensemble. KBIs are related to a number of thermodynamic quantities such as partial derivatives of the chemical potential with respect to composition, partial molar volumes, and the isothermal compressibility. Consequently, chemical potentials and other thermodynamic properties can be obtained from KBIs. The main advantage of the KB theory is that it can be applied to any type of interactions, making it an attractive alternative to molecular simulations in open ensembles which are usually carried out to predict phase equilibria. Molecular simulations in open ensembles require insertions/deletions of molecules, which can be challenging for dense liquids with strong molecular interactions. While KBIs are derived for open and infinitely large systems, they can be computed from molecular simulations of finite and closed systems (e.g. simulations in the *NVT* and *NPT* ensembles). Krüger and co-workers [74] derived an expression for KBIs of small and open subvolumes embedded in a larger reservoir (simulation box). These KBIs scale with the inverse size of the subvolume and extrapolating this scaling to the thermodynamic limit yields KBIs of open and infinite systems.

## 6

To accurately apply the method of Krüger and co-workers [63, 74], a number of effects were examined. In chapter 2, the effect of the shape of the subvolume was investigated. When computing KBIs, the weight function  $w(x)$  is the only term in the method of Krüger and co-workers [63, 74] that depends on the shape of the subvolume. In chapter 2, a method to compute  $w(x)$  for subvolumes with arbitrary shape was developed. From computing  $w(x)$  and KBIs of various subvolume shapes, it was demonstrated that KBIs in the thermodynamic limit are independent of the shape of the used subvolume. For small values of  $x$  and for all shapes, all functions  $w(x)$  can be transformed on to a universal function that only depends on the area to volume ratio of the subvolume. Using this universal expression for  $w(x)$ , it was confirmed that the truncation of KBIs for infinitely large systems, which is the most commonly used approach to compute KBIs, is not correct and nonphysical. Two system size effects are observed in MD simulations: (1) effects due to the size of the simulation box and the size of the finite subvolume embedded in the simulation box, and (2) effects due to computing RDFs from a closed and finite system. In chapter 3, finite-size effects of computing KBIs from molecular simulations were investigated using systems of WCA particles. It was demonstrated that calculations of KBIs should not be extended beyond half the size of the simulation box. For finite-size effects related to RDFs, the Ganguly and Van der Vegt correction [113] was found to yield the most accurate results. Numerical inaccuracies may also arise from extrapolating the scaling

of KBIs of small subvolumes with the inverse size of the subvolume. In chapter 4, alternative approaches for extrapolating RDFs of finite systems to compute KBIs in the thermodynamic limit were considered. These methods also allow for the computation of surface effects. KBIs and surface terms in the thermodynamic limit were computed for LJ and WCA fluids. It was found that the methods discussed in chapter 4 converge to the same value in the thermodynamic limit. The main differentiating factor was the quality of the convergence with the size of the subvolume  $L$ . The method that required the smallest size was the one which exploited the scaling of finite volume KBIs multiplied by  $L$ . KBIs and surface terms were computed at different densities. As the density is increased, differences in KBIs of LJ and WCA fluids vanish due to dominating repulsive interactions. The values of the surface terms were of the order of magnitude of KBIs.

The method of Krüger and co-workers [74] provides the advantage of accessing a grand-canonical setup in a closed system. As a result, KBIs of ionic systems can be computed while maintaining the electroneutrality of the system. In chapter 5, KBIs of solutions of choline chloride and urea were computed using MD simulations. The outcome from chapters 3 and 4 were applied to mitigate finite-size effects and obtain KBIs in the thermodynamic limit. KBIs were computed for systems of choline chloride and urea with varying molar ratios, and were used to study the affinity between components of the system at varying composition. Also, thermodynamic and transport properties were computed, and from KBIs, the thermodynamic factors and partial molar volumes were computed. KBIs and the thermodynamic factors were found to be useful in examining the effect of composition on molecular interactions. As more urea is added to the system, urea-urea interactions become stronger while interactions of pairs of urea-ChCl and pairs of ChCl-ChCl slightly decreased. The values of the thermodynamic factors was found to be larger than one, indicating that, on average, urea-ChCl interactions are stronger than interactions of identical components at the studied compositions (ranging from molar ratios of ChCl to urea of 2:1 to 1:5). While KBIs provide a connection to fluctuations in the grand-canonical ensemble and hence to several thermodynamic quantities, in some cases, it is more convenient to directly perform simulations in an open ensemble. For instance, when computing the solubility of gases in liquid solvents, MC simulations in the grand-canonical and osmotic ensembles are commonly used for this. In chapter 6, solubilities of CO<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>S, and N<sub>2</sub> in monoethylene glycol (MEG) were computed using MC simulations in the osmotic ensemble. The CFCMC method was used. Solubilities from experiments and simulations are in good agreement at low pressures, but deviations were observed at high pressures. The order of

solubilities of the gases in MEG at 373.15 K was computed as  $\text{H}_2\text{S} > \text{CO}_2 > \text{CH}_4 > \text{N}_2$ . Force field modifications may be required to improve the prediction of solubilities of gases in MEG at high pressures and low temperatures.

In this thesis, the KB theory and the CFCMC method were used to predict thermodynamic properties of fluids with strong molecular interactions. In both methods, long simulations were required to achieve low statistical errors and experienced challenges were specific to the studied systems. When simulating urea–ChCl mixtures, sampling the system was found to be computationally demanding, while difficulties in scaling the interactions of the fractional molecules were faced when simulating MEG. Further research can be carried out to compare the two methods and assess their performance in predicting thermodynamic properties of complex systems.

## BIBLIOGRAPHY

- [1] B. E. Poling, J. M. Prausnitz, and J. P. O'Connell, *The properties of gases and liquids*, 5th ed., Vol. 5 (Mcgraw-hill, 2001).
- [2] O. Pfohl and R. Dohrn, *Provision of thermodynamic properties of polymer systems for industrial applications*, Fluid Phase Equilibria **217**, 189–199 (2004).
- [3] S. I. Sandler, *Chemical, biochemical, and engineering thermodynamics*, 4th ed. (John Wiley & Sons, New York, USA, 2006).
- [4] E. Hendriks, G. M. Kontogeorgis, R. Dohrn, J.-C. de Hemptinne, I. G. Economou, L. F. Žilnik, and V. Vesovic, *Industrial requirements for thermodynamics and transport properties*, Industrial & Engineering Chemistry Research **49**, 11131–11141 (2010).
- [5] E. J. Maginn, *Molecular simulation of ionic liquids: current status and future opportunities*, Journal of Physics: Condensed Matter **21**, 373101 (2009).
- [6] L. J. B. M. Kollau, M. Vis, A. van den Bruinhorst, A. C. C. Esteves, and R. Tuinier, *Quantification of the liquid window of deep eutectic solvents*, Chemical Communications **54**, 13351–13354 (2018).
- [7] B. Dahlgren, M. M. Reif, P. H. Hünenberger, and N. Hansen, *Calculation of derivative thermodynamic hydration and aqueous partial molar properties of ions based on atomistic simulations*, Journal of Chemical Theory and Computation **8**, 3542–3564 (2012).
- [8] I. Nezbeda, F. Moučka, and W. R. Smith, *Recent progress in molecular simulation of aqueous electrolytes: Force fields, chemical potentials and solubility*, Molecular Physics **114**, 1665–1690 (2016).
- [9] K. N. Marsh, J. A. Boxall, and R. Lichtenthaler, *Room temperature ionic liquids and their mixtures—a review*, Fluid Phase Equilibria **219**, 93–98 (2004).
- [10] J. E. S. J. Reid, A. J. Walker, and S. Shimizu, *Residual water in ionic liquids: clustered or dissociated?* Physical Chemistry Chemical Physics **17**, 14710–14718 (2015).

- [11] F. Gabriele, M. Chiarini, R. Germani, M. Tiecco, and N. Spreti, *Effect of water addition on choline chloride/glycol deep eutectic solvents: characterization of their structural and physicochemical properties*, Journal of Molecular Liquids **291**, 111301 (2019).
- [12] J. G. Kirkwood and F. P. Buff, *The statistical mechanical theory of solutions. I*, Journal of Chemical Physics **19**, 774–777 (1951).
- [13] A. Ben-Naim, *Molecular theory of solutions*, 1st ed. (Oxford University Press, Oxford, UK, 2006).
- [14] A. Ben-Naim, *Inversion of the Kirkwood–Buff theory of solutions: application to the water–ethanol system*, Journal of Chemical Physics **67**, 4884–4890 (1977).
- [15] D. G. Hall, *Kirkwood–Buff theory of solutions. An alternative derivation of part of it and some applications*, Transactions of the Faraday Society **67**, 2516–2524 (1971).
- [16] S. K. Schnell, X. Liu, J.-M. Simon, A. Bardow, D. Bedeaux, T. J. H. Vlugt, and S. Kjelstrup, *Calculating thermodynamic properties from fluctuations at small scales*, Journal of Physical Chemistry B **115**, 10911–10918 (2011).
- [17] P. E. Smith, *On the Kirkwood–Buff inversion procedure*, Journal of Chemical Physics **129**, 124509 (2008).
- [18] J. D. Pandey and R. Verma, *Inversion of the Kirkwood–Buff theory of solutions: application to binary systems*, Chemical Physics **270**, 429–438 (2001).
- [19] E. Matteoli, *A study on Kirkwood–Buff integrals and preferential solvation in mixtures with small deviations from ideality and/or with size mismatch of components. Importance of a proper reference system*, Journal of Physical Chemistry B **101**, 9800–9810 (1997).
- [20] R. Chitra and P. E. Smith, *Molecular association in solution: a Kirkwood–Buff analysis of sodium chloride, ammonium sulfate, guanidinium chloride, urea, and 2,2,2-trifluoroethanol in water*, Journal of Physical Chemistry B **106**, 1491–1500 (2002).
- [21] J. M. Haile, I. Johnston, A. J. Mallinckrodt, and S. McKay, *Molecular dynamics simulation: elementary methods*, Computers in Physics **7**, 625–625 (1993).
- [22] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, 2nd ed., Vol. 1 (Academic press, London, UK, 2002).

- [23] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, 2nd ed. (Oxford University Press, New York, US, 2017).
- [24] D. C. Rapaport, *The art of molecular dynamics simulation*, 1st ed. (Cambridge University Press, Cambridge, England, 2004).
- [25] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. (Springer, New York, US, 2004).
- [26] J.-P. Hansen and I. R. McDonald, *Theory of simple liquids*, 2nd ed. (Elsevier, Amsterdam, Netherlands, 1990).
- [27] D. Borgis, R. Assaraf, B. Rotenberg, and R. Vuilleumier, *Computation of pair distribution functions and three-dimensional densities with a reduced variance principle*, Molecular Physics **111**, 3486–3492 (2013).
- [28] D. de las Heras and M. Schmidt, *Better than counting: Density profiles from force sampling*, Physical Review Letters **120**, 218001 (2018).
- [29] S. H. Jamali, *Transport Properties of Fluids: Methodology and Force Field Improvement using Molecular Dynamics Simulations*, Ph.D. thesis, Delft University of Technology (2020), <https://doi.org/tudelft.idm.oclc.org/10.4233/uuid:c32e495d-5dd9-42f8-9203-b540e1f9f175>.
- [30] K. E. Newman, *Kirkwood–Buff solution theory: derivation and applications*, Chemical Society Reviews **23**, 31–40 (1994).
- [31] W. H. Stockmayer, *Light scattering in multi-component systems*, Journal of Chemical Physics **18**, 58–61 (1950).
- [32] J. G. Kirkwood and R. J. Goldberg, *Light scattering arising from composition fluctuations in multi-component systems*, Journal of Chemical Physics **18**, 54–57 (1950).
- [33] E. Ruckenstein and I. Shulgin, *Entrainier effect in supercritical mixtures*, Fluid Phase Equilibria **180**, 345–359 (2001).
- [34] E. Ruckenstein and I. Shulgin, *Effect of a third component on the interactions in a binary mixture determined from the fluctuation theory of solutions*, Fluid Phase Equilibria **180**, 281–297 (2001).
- [35] L. Lepori and E. Matteoli, *Excess Gibbs energies of the ternary system ethanol+ tetrahydrofuran+ cyclohexane at 298.15 K*, Fluid Phase Equilibria **134**, 113–131 (1997).

- [36] D. M. Pfund, L. L. Lee, and H. D. Cochran, *Application of the Kirkwood–Buff theory of solutions to dilute supercritical mixtures. II. The excluded volume and local composition models*, Fluid Phase Equilibria **39**, 161–192 (1988).
- [37] A. Ben-Naim, *Preferential solvation in two- and in three-component systems*, Pure and Applied Chemistry **62**, 25–34 (1990).
- [38] E. Matteoli and L. Lepori, *Solute–solute interactions in water. II. An analysis through the Kirkwood–Buff integrals for 14 organic solutes*, Journal of Chemical Physics **80**, 2856–2863 (1984).
- [39] K. J. Patil, *Application of Kirkwood–Buff theory of liquid mixtures to water–butanol system*, Journal of Solution Chemistry **10**, 315–320 (1981).
- [40] K. J. Patil, G. R. Mehta, and S. S. Dhondge, *Application of Kirkwood–Buff theory of liquid mixtures to binary aqueous solutions of alcohols*, India Journal of Chemistry **33**, 1069–1074 (1994).
- [41] A. K. Nain, *Inversion of the Kirkwood–Buff theory of solutions: application to tetrahydrofuran+ aromatic hydrocarbon binary liquid mixtures*, Journal of Solution Chemistry **37**, 1541–1559 (2008).
- [42] A. K. Nain, *Application of the Kirkwood–Buff theory of solutions to acetonitrile+ amide binary mixtures by using inversion procedure and regular solution theory*, Journal of Chemical Sciences **121**, 361–367 (2009).
- [43] J. J. Shephard, S. K. Callear, S. Imberti, J. S. O. Evans, and C. G. Salzmann, *Microstructures of negative and positive azeotropes*, Physical Chemistry Chemical Physics **18**, 19227–19235 (2016).
- [44] T. Kobayashi, J. E. S. J. Reid, S. Shimizu, M. Fyta, and J. Smiatek, *The properties of residual water molecules in ionic liquids: a comparison between direct and inverse Kirkwood–Buff approaches*, Physical Chemistry Chemical Physics **19**, 18924–18937 (2017).
- [45] T. Kato, T. Fujiyama, and H. Nomura, *Estimation of parameters,  $G_{11}$ ,  $G_{22}$ , and  $G_{12}$  in the Kirkwood–Buff solution theory on the basis of the concentration fluctuation data obtained from Rayleigh scattering*, Bulletin of the Chemical Society of Japan **55**, 3368–3372 (1982).
- [46] K. Nishikawa, *Simple relationship between the Kirkwood–Buff parameters and the fluctuations in the particle number and concentration obtained by small-angle X-ray scattering: application to tert-butyl alcohol and water mixtures*, Chemical Physics Letters **132**, 50–54 (1986).

- [47] I. Shulgin and E. Ruckenstein, *Kirkwood–Buff integrals in aqueous alcohol systems: comparison between thermodynamic calculations and X-ray scattering experiments*, Journal of Physical Chemistry B **103**, 2496–2503 (1999).
- [48] K. Nishikawa, H. Hayashi, and T. Iijima, *Temperature dependence of the concentration fluctuation, the Kirkwood–Buff parameters, and the correlation length of tert-butyl alcohol and water mixtures studied by small-angle X-ray scattering*, Journal of Physical Chemistry **93**, 6559–6565 (1989).
- [49] L. Almásy, G. Jancsó, and L. Cser, *Application of SANS to the determination of Kirkwood–Buff integrals in liquid mixtures*, Applied Physics A **74**, s1376–s1378 (2002).
- [50] M. A. Blanco, E. Sahin, Y. Li, and C. J. Roberts, *Reexamining protein–protein and protein–solvent interactions from Kirkwood–Buff analysis of light scattering in multi-component solutions*, Journal of Chemical Physics **134**, 06B606 (2011).
- [51] A. Perera, F. Sokolić, L. Almásy, P. Westh, and Y. Koga, *On the evaluation of the Kirkwood–Buff integrals of aqueous acetone mixtures*, Journal of Chemical Physics **123**, 024503 (2005).
- [52] A. Perera, F. Sokolić, L. Almásy, and Y. Koga, *Kirkwood–Buff integrals of aqueous alcohol binary mixtures*, Journal of Chemical Physics **124**, 124515 (2006).
- [53] L. Almásy, M. Turmine, and A. Perera, *Structure of aqueous solutions of ionic liquid 1-butyl-3-methylimidazolium tetrafluoroborate by small-angle neutron scattering*, Journal of Physical Chemistry B **112** (2008).
- [54] A. A. Galata, S. D. Anogiannakis, and D. N. Theodorou, *Thermodynamic analysis of Lennard-Jones binary mixtures using Kirkwood–Buff theory*, Fluid Phase Equilibria **470**, 25–37 (2018).
- [55] M. J. Mitchell and J. McCammon, *Free energy difference calculations by thermodynamic integration: difficulties in obtaining a precise value*, Journal of Computational Chemistry **12**, 271–275 (1991).
- [56] D. A. Kofke and P. T. Cummings, *Precision and accuracy of staged free-energy perturbation methods for computing the chemical potential by molecular simulation*, Fluid Phase Equilibria **150**, 41–49 (1998).
- [57] B. Widom, *Some topics in the theory of fluids*, Journal of Chemical Physics **39**, 2808–2812 (1963).

- [58] D. Dubbeldam, A. Torres-Knoop, and K. S. Walton, *On the inner workings of Monte Carlo codes*, Molecular Simulation **39**, 1253–1292 (2013).
- [59] W. Shi and E. J. Maginn, *Continuous Fractional Component Monte Carlo: an adaptive biasing method for open system atomistic simulations*, Journal of Chemical Theory and Computation **3**, 1451–1463 (2007).
- [60] A. Poursaeidesfahani, A. Torres-Knoop, D. Dubbeldam, and T. J. H. Vlugt, *Direct free energy calculation in the Continuous Fractional Component Gibbs ensemble*, Journal of Chemical Theory and Computation **12**, 1481–1490 (2016).
- [61] A. Poursaeidesfahani, A. Rahbari, A. Torres-Knoop, D. Dubbeldam, and T. J. H. Vlugt, *Computation of thermodynamic properties in the continuous fractional component Monte Carlo Gibbs ensemble*, Molecular Simulation **43**, 189–195 (2017).
- [62] A. Rahbari, R. Hens, M. Ramdin, O. A. Moulton, D. Dubbeldam, and T. J. H. Vlugt, *Recent advances in the Continuous Fractional Component Monte Carlo methodology*, Molecular Simulation (2021), in press. <https://doi.org/10.1080/08927022.2020.1828585>.
- [63] N. Dawass, P. Krüger, S. K. Schnell, J.-M. Simon, and T. J. H. Vlugt, *Kirkwood-Buff integrals from molecular simulation*, Fluid Phase Equilibria **486**, 21–36 (2019).
- [64] X. Liu, S. K. Schnell, J.-M. Simon, P. Krüger, D. Bedeaux, S. Kjelstrup, A. Bardow, and T. J. H. Vlugt, *Diffusion coefficients from molecular dynamics simulations in binary and ternary mixtures*, International Journal of Thermophysics **34**, 1169–1196 (2013).
- [65] R. Taylor and H. A. Kooijman, *Composition derivatives of activity coefficient models (for the estimation of thermodynamic factors in diffusion)*, Chemical Engineering Communications **102**, 87–106 (1991).
- [66] R. Krishna and J. Wesselingh, *The Maxwell–Stefan approach to mass transfer*, Chemical Engineering Science **52**, 861–911 (1997).
- [67] S. Kjelstrup, S. K. Schnell, T. J. H. Vlugt, J.-M. Simon, A. Bardow, D. Bedeaux, and T. Trinh, *Bridging scales with thermodynamics: from nano to macro*, Advances in Natural Sciences: Nanoscience and Nanotechnology **5**, 023002 (2014).

- [68] R. Taylor and R. Krishna, *Multicomponent mass transfer*, 1st ed., Vol. 2 (John Wiley & Sons, New York, USA, 1993).
- [69] X. Liu, S. K. Schnell, J.-M. Simon, D. Bedeaux, S. Kjelstrup, A. Bardow, and T. J. H. Vlugt, *Fick diffusion coefficients of liquid mixtures directly obtained from equilibrium molecular dynamics*, Journal of Physical Chemistry B **115**, 12921–12929 (2011).
- [70] X. Liu, A. Martín-Calvo, E. McGarrity, S. K. Schnell, S. Calero, J.-M. Simon, D. Bedeaux, S. Kjelstrup, A. Bardow, and T. J. H. Vlugt, *Fick diffusion coefficients in ternary liquid systems from equilibrium molecular dynamics simulations*, Industrial & Engineering Chemistry Research **51**, 10247–10258 (2012).
- [71] R. Fingerhut, G. Herres, and J. Vrabec, *Thermodynamic factor of quaternary mixtures from Kirkwood–Buff integration*, Molecular Physics **118**, e1643046 (2020).
- [72] S. H. Jamali, L. Wolff, T. M. Becker, A. Bardow, T. J. H. Vlugt, and O. A. Moulton, *Finite-size effects of binary mutual diffusion coefficients from molecular dynamics*, Journal of Chemical Theory and Computation **14**, 2667–2677 (2018).
- [73] A. T. Celebi, S. H. Jamali, A. Bardow, T. J. H. Vlugt, and O. A. Moulton, *Finite-size effects of diffusion coefficients computed from molecular dynamics: a review of what we have learned so far*, Molecular Simulation (2020), in press. <https://doi.org/10.1080/08927022.2020.1810685>.
- [74] P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, T. J. H. Vlugt, and J.-M. Simon, *Kirkwood–Buff integrals for finite volumes*, Journal of Physical Chemistry Letters **4**, 235–238 (2013).
- [75] R. Wedberg, J. P. O’Connell, G. H. Peters, and J. Abildskov, *Accurate Kirkwood–Buff integrals from molecular simulations*, Molecular Simulation **36**, 1243–1252 (2010).
- [76] J. W. Nichols, S. G. Moore, and D. R. Wheeler, *Improved implementation of Kirkwood–Buff solution theory in periodic molecular simulations*, Physical Review E **80**, 051203 (2009).
- [77] M. Miyaji, B. Radola, J.-M. Simon, and P. Krüger, *Extension of Kirkwood–Buff theory to solids and its application to the compressibility of fcc argon*, Journal of Chemical Physics **154**, 164506 (2021).

- [78] T. L. Hill, *Thermodynamics of small systems*, Journal of Chemical Physics **36**, 3182–3197 (1962).
- [79] T. L. Hill, *Thermodynamics of small systems*, 1st ed. (Dover, New York, US, 1994).
- [80] N. Dawass, P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, J.-M. Simon, and T. J. H. Vlugt, *Finite-size effects of Kirkwood–Buff integrals from molecular simulations*, Molecular Simulation **44**, 1–14 (2017).
- [81] P. Krüger and T. J. H. Vlugt, *Size and shape dependence of finite-volume Kirkwood–Buff integrals*, Physical Review E **97**, 051301 (2018).
- [82] E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics*, 2nd ed. (Chapman and Hall, Boca Raton, Florida).
- [83] R. Cortes-Huerto, K. Kremer, and R. Potestio, *Communication: Kirkwood–Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations*, Journal of Chemical Physics **145**, 141103 (2016).
- [84] N. Dawass, P. Krüger, J.-M. Simon, and T. J. H. Vlugt, *Kirkwood–Buff integrals of infinite systems: shape effects*, Molecular Physics **116**, 1–8 (2018).
- [85] R. Wedberg, J. P. O’Connell, G. H. Peters, and J. Abildskov, *Total and direct correlation function integrals from molecular simulation of binary systems*, Fluid Phase Equilibria **302**, 32–42 (2011).
- [86] L. Verlet, *Computer "experiments" on classical fluids. II. Equilibrium correlation functions*, Physical Review **165**, 201 (1968).
- [87] D. M. Rogers, *Extension of Kirkwood–Buff theory to the canonical ensemble*, Journal of Chemical Physics **148**, 054102 (2018).
- [88] N. H. March and M. P. Tosi, *Atomic dynamics in liquids*, 1st ed. (Dover, New York, US, 1991).
- [89] S. K. Schnell, R. Skorpa, D. Bedeaux, S. Kjelstrup, T. J. H. Vlugt, and J.-M. Simon, *Partial molar enthalpies and reaction enthalpies from equilibrium molecular dynamics simulation*, Journal of Chemical Physics **141**, 144501 (2014).
- [90] R. Skorpa, J.-M. Simon, D. Bedeaux, and S. Kjelstrup, *The reaction enthalpy of hydrogen dissociation calculated with the small system method from simulation of molecular fluctuations*, Physical Chemistry Chemical Physics **16**, 19681–19693 (2014).

- [91] S. K. Schnell, P. Englebienne, J.-M. Simon, P. Krüger, S. P. Balaji, S. Kjelstrup, D. Bedeaux, A. Bardow, and T. J. H. Vlugt, *How to apply the Kirkwood–Buff theory to individual species in salt solutions*, Chemical Physics Letters **582**, 154–157 (2013).
- [92] X. Liu, *Diffusion in Liquids: equilibrium Molecular Simulations and Predictive Engineering Models*, Ph.D. thesis, Delft University of Technology (2013), <https://doi.org/10.4233/uuid:239b19f7-59d0-47b3-a3e3-dd1aeb19701f>.
- [93] J. M. O. de Zárate and J. V. Sengers, *Frame-invariant fick diffusion matrices of multicomponent fluid mixtures*, Physical Chemistry Chemical Physics **22**, 17597–17604 (2020).
- [94] S. H. Jamali, A. Bardow, T. J. H. Vlugt, and O. A. Mourtos, *Generalized form for finite-size corrections in mutual diffusion coefficients of multicomponent mixtures obtained from equilibrium Molecular Dynamics simulation*, Journal of Chemical Theory and Computation **16**, 3799–3806 (2020).
- [95] B. Dünweg and K. Kremer, *Molecular dynamics simulation of a polymer chain in solution*, Journal of Chemical Physics **99**, 6983–6997 (1993).
- [96] C. J. Leverant, J. A. Harvey, and T. M. Alam, *Machine learning-based upscaling of finite-size Molecular Dynamics diffusion simulations for binary fluids*, The Journal of Physical Chemistry Letters **11**, 10375–10381 (2020).
- [97] I. Yeh and G. Hummer, *System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions*, Journal of Physical Chemistry B **108**, 15873–15879 (2004).
- [98] M. Erdős, M. Frangou, T. J. H. Vlugt, and O. A. Mourtos, *Diffusivity of  $\alpha$ -,  $\beta$ -,  $\gamma$ -cyclodextrin and the inclusion complex of  $\beta$ -cyclodextrin: Ibuprofen in aqueous solutions; A Molecular Dynamics simulation study*, Fluid Phase Equilibria **528**, 112842 (2020).
- [99] V. Pierce, M. Kang, M. Aburi, S. Weerasinghe, and P. E. Smith, *Recent applications of Kirkwood–Buff theory to biological systems*, Cell Biochemistry and Biophysics **50**, 1–22 (2008).
- [100] E. A. Oprzeska-Zingrebe and J. Smiatek, *Preferential binding of urea to single-stranded DNA structures: a molecular dynamics study*, Biophysical Journal **114**, 1551–1562 (2018).

- [101] A. Ben-Naim, *Theoretical aspects of self-assembly of proteins: A Kirkwood-Buff theory approach*, Journal of Chemical Physics **138**, 224906 (2013).
- [102] P. E. Smith, *Cosolvent interactions with biomolecules: relating computer simulation data to experimental thermodynamic data*, Journal of Physical Chemistry B **108**, 18716–18724 (2004).
- [103] S. Chiba, T. Furuta, and S. Shimizu, *Kirkwood–buff integrals for aqueous urea solutions based upon the quantum chemical electrostatic potential and interaction energies*, Journal of Physical Chemistry B **120**, 7714–7723 (2016).
- [104] M. Aburi and P. E. Smith, *A combined simulation and Kirkwood-Buff approach to quantify cosolvent effects on the conformational preferences of peptides in solution*, Journal of Physical Chemistry B **108**, 7382–7388 (2004).
- [105] N. Naleem, N. Bentenitis, and P. E. Smith, *A Kirkwood–Buff derived force field for alkaline earth halide salts*, Journal of Chemical Physics **148**, 222828 (2018).
- [106] S. Weerasinghe and P. E. Smith, *A Kirkwood–Buff derived force field for sodium chloride in water*, Journal of Chemical Physics **119**, 11342–11349 (2003).
- [107] M. Kang and P. E. Smith, *A Kirkwood–Buff derived force field for amides*, Journal of Computational Chemistry **27**, 1477–1485 (2006).
- [108] S. Weerasinghe and P. E. Smith, *A Kirkwood–Buff derived force field for mixtures of urea and water*, Journal of Physical Chemistry B **107**, 3891–3898 (2003).
- [109] S. Weerasinghe and P. E. Smith, *Kirkwood–Buff derived force field for mixtures of acetone and water*, Journal of Chemical Physics **118**, 10663–10670 (2003).
- [110] S. Weerasinghe and P. E. Smith, *A Kirkwood–Buff derived force field for methanol and aqueous methanol solutions*, Journal of Physical Chemistry B **109**, 15080–15086 (2005).
- [111] M. Mijaković, K. D. Polok, B. Kežić, F. Sokolić, A. Perera, and L. Zoranić, *A comparison of force fields for ethanol-water mixtures*, Molecular Simulation **41**, 699–712 (2015).

- [112] J. D. Weeks, D. Chandler, and H. C. Andersen, *Role of repulsive forces in determining the equilibrium structure of simple liquids*, Journal of Chemical Physics **54**, 5237–5247 (1971).
- [113] P. Ganguly and N. F. A. van der Vegt, *Convergence of sampling Kirkwood-Buff integrals of aqueous solutions with molecular dynamics simulations*, Journal of Chemical Theory and Computation **9**, 1347–1355 (2013).
- [114] P. E. Smith, E. Matteoli, and J. P. O'Connell, *Fluctuation theory of solutions: applications in chemistry, chemical engineering, and biophysics* (CRC Press, Boca Raton, Florida, 2013).
- [115] A. Ben-Naim, *Theoretical aspects of self-assembly of proteins: A Kirkwood-Buff-theory approach*, Journal of Chemical Physics **138**, 224906 (2013).
- [116] G. Giambasu, T. Luchko, D. Herschlag, D. York, and D. Case, *Ion counting from explicit-solvent simulations and 3d-rism*, Biophysical Journal **106**, 883 – 894 (2014).
- [117] M. M. Reif, M. Winger, and C. Oostenbrink, *Testing of the GROMOS force-field parameter set 54a8: Structural properties of electrolyte solutions, lipid bilayers, and proteins*, Journal of Chemical Theory and Computation **9**, 1247–1264 (2013).
- [118] T. J. H. Vlugt and N. Dawass, *Kirkwood-Buff integrals of finite systems: geometric functions  $w(x)$* , [https://data.4tu.nl/articles/Kirkwood\\_Buff\\_integrals\\_of\\_finite\\_systems\\_geometric\\_functions\\_w\\_x\\_/12708683](https://data.4tu.nl/articles/Kirkwood_Buff_integrals_of_finite_systems_geometric_functions_w_x_/12708683) (2018).
- [119] G. Torrie and J. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, Journal of Computational Physics **23**, 187 – 199 (1977).
- [120] J. G. Kirkwood and E. M. Boggs, *The radial distribution function in liquids*, Journal of Chemical Physics **10**, 394–402 (1942).
- [121] D. V. Griffiths and I. M. Smith, *Numerical methods for engineers*, 2nd ed. (CRC press, Boca Raton, Florida, 2006).
- [122] B. A. Strøm, J.-M. Simon, S. K. Schnell, S. Kjelstrup, J. He, and D. Bedeaux, *Size and shape effects on the thermodynamic properties of nanoscale volumes of water*, Physical Chemistry Chemical Physics **19**, 9016–9027 (2017).

- [123] S. K. Schnell, T. J. H. Vlugt, J.-M. Simon, D. Bedeaux, and S. Kjelstrup, *Thermodynamics of small systems embedded in a reservoir: a detailed analysis of finite size effects*, Molecular Physics **110**, 1069–1079 (2012).
- [124] M. Deserno, *How to calculate a three-dimensional g(r) under periodic boundary conditions*, [www.cmu.edu/biophys/deserno/pdf/gr\\_periodic.pdf](http://www.cmu.edu/biophys/deserno/pdf/gr_periodic.pdf) (2004).
- [125] J. J. Salacuse, A. R. Denton, and P. A. Egelstaff, *Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. i. Theoretical method*, Physical Review E **53**, 2382–2389 (1996).
- [126] J. D. Hoffman and S. Frankel, *Numerical methods for engineers and scientists*, 3rd ed. (CRC press, Boca Raton, Florida, 2001).
- [127] S. Plimpton, *Fast parallel algorithms for short-range molecular dynamics*, Journal of Computational Physics **117**, 1–19 (1995).
- [128] N. Dawass, P. Krüger, S. K. Schnell, O. A. Moulton, I. G. Economou, T. J. H. Vlugt, and J. M. Simon, *Kirkwood–Buff integrals using Molecular Simulation: Estimation of surface effects*, Nanomaterials **10**, 771 (2020).
- [129] A. Santos, *Finite-size estimates of Kirkwood–Buff and similar integrals*, Physical Review E **98**, 063302 (2018).
- [130] J. E. Lennard-Jones, *On the determination of molecular fields. II. from the equation of state of gas*, Proc. Roy. Soc. A **106**, 463–477 (1924).
- [131] A. T. Celebi, N. Dawass, O. A. Moulton, and T. J. H. Vlugt, *How sensitive are physical properties of choline chloride–urea mixtures to composition changes: Molecular Dynamics simulations and Kirkwood Buff theory*, Journal of Chemical Physics **154**, 184502 (2021).
- [132] E. L. Smith, A. P. Abbott, and K. S. Ryder, *Deep eutectic solvents (DESs) and their applications*, Chemical Reviews **114**, 11060–11082 (2014).
- [133] Q. Zhang, K. D. O. Vigier, S. Royer, and F. Jerome, *Deep eutectic solvents: syntheses, properties and applications*, Chemical Society Reviews **41**, 7108–7146 (2012).
- [134] R. D. Rogers and K. R. Seddon, *Ionic liquids–solvents of the future?* Science **302**, 792–793 (2003).
- [135] K. R. Seddon, *Ionic liquids for clean technology*, Journal of Chemical Technology & Biotechnology **68**, 351–356 (1997).

- [136] T. Welton, *Ionic liquids in catalysis*, Coordination Chemistry Reviews **248**, 2459–2477 (2004).
- [137] N. V. Plechkova and K. R. Seddon, *Applications of ionic liquids in the chemical industry*, Chemical Society Reviews **37**, 123–150 (2008).
- [138] S. A. Forsyth, J. M. Pringle, and D. R. MacFarlane, *Ionic liquids—an overview*, Australian Journal of Chemistry **57**, 113–119 (2004).
- [139] Y. Liu, J. B. Friesen, J. B. McAlpine, D. C. Lankin, S.-N. Chen, and G. F. Pauli, *Natural deep eutectic solvents: properties, applications, and perspectives*, Journal of Natural Products **81**, 679–690 (2018).
- [140] A. Shishov, A. Bulatov, M. Locatelli, S. Carradori, and V. Andruch, *Application of deep eutectic solvents in analytical chemistry. a review*, Microchemical Journal **135**, 33–38 (2017).
- [141] D. V. Wagle, H. Zhao, and G. A. Baker, *Deep eutectic solvents: sustainable media for nanoscale and functional materials*, Accounts of Chemical Research **47**, 2299–2308 (2014).
- [142] E. Ali, M. K. Hadj-Kali, S. Mulyono, I. Alnashef, A. Fakieha, F. Mjalli, and A. Hayyan, *Solubility of CO<sub>2</sub> in deep eutectic solvents: experiments and modelling using the Peng–Robinson equation of state*, Chemical Engineering Research and Design **92**, 1898–1906 (2014).
- [143] H. S. Salehi, R. Hens, O. A. Moultsos, and T. J. H. Vlugt, *Computation of gas solubilities in choline chloride urea and choline chloride ethylene glycol Deep Eutectic Solvents using Monte Carlo simulations*, Journal of Molecular Liquids **316**, 113729 (2020).
- [144] A. T. Celebi, T. J. H. Vlugt, and O. A. Moultsos, *Structural, thermodynamic, and transport properties of aqueous reline and ethaline solutions from molecular dynamics simulations*, Journal of Physical Chemistry B **123**, 11014–11025 (2019).
- [145] X.-D. Hou, A.-L. Li, K.-P. Lin, Y.-Y. Wang, Z.-Y. Kuang, and S.-L. Cao, *Insight into the structure-function relationships of deep eutectic solvents during rice straw pretreatment*, Bioresource Technology **249**, 261–267 (2018).
- [146] D. J. G. P. van Osch, C. H. J. T. Dietz, J. van Spronsen, M. C. Kroon, F. Gallicci, M. van Sint Annaland, and R. Tuinier, *A search for natural hydrophobic deep eutectic solvents based on natural components*, ACS Sustainable Chemistry & Engineering **7**, 2933–2942 (2019).

- [147] S. L. Perkins, P. Painter, and C. M. Colina, *Molecular dynamic simulations and vibrational analysis of an ionic liquid analogue*, Journal of Physical Chemistry B **117**, 10250–10260 (2013).
- [148] W. Shi and E. J. Maginn, *Molecular simulation and regular solution theory modeling of pure and mixed gas absorption in the ionic liquid 1-n-hexyl-3-methylimidazolium bis (trifluoromethylsulfonyl) amide ([hmim][Tf<sub>2</sub>N])*, Journal of Physical Chemistry B **112**, 16710–16720 (2008).
- [149] M. Ramdin, S. P. Balaji, A. Torres-Knoop, D. Dubbeldam, T. W. de Loos, and T. J. H. Vlugt, *Solubility of natural gas species in ionic liquids and commercial solvents: experiments and Monte Carlo simulations*, Journal of Chemical & Engineering Data **60**, 3039–3045 (2015).
- [150] J. E. S. J. Reid, R. J. Gammons, J. M. Slattery, A. J. Walker, and S. Shimizu, *Interactions in water-ionic liquid mixtures: comparing protic and aprotic systems*, Journal of Physical Chemistry B **121**, 599–609 (2017).
- [151] S. Heydarian, M. Almasi, and Z. Saadati, *Calculation of Kirkwood-Buff integrals for binary mixtures of 1-butyl-3-methylimidazolium nitrate ionic liquid and alcohols at 298.15 k*, Journal of Molecular Liquids **275**, 122–125 (2019).
- [152] R. Krishna and J. M. Van Baten, *Onsager coefficients for binary mixture diffusion in nanopores*, Chemical Engineering Science **63**, 3120–3140 (2008).
- [153] R. Krishna and J. M. Van Baten, *The Darken relation for multicomponent diffusion in liquid mixtures of linear alkanes: an investigation using Molecular Dynamics (MD) simulations*, Industrial & Engineering Chemistry Research **44**, 6939–6947 (2005).
- [154] S. H. Jamali, L. Wolff, T. M. Becker, M. de Groen, M. Ramdin, R. Hartkamp, A. Bardow, T. J. H. Vlugt, and O. A. Moulton, *OCTP: A tool for on-the-fly calculation of transport properties of fluids with the order-n algorithm in lammps*, Journal of Chemical Information and Modelling **59**, 1290–1294 (2019).
- [155] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *Development and testing of a general amber force field*, Journal of Computational Chemistry **25**, 1157–1174 (2004).

- [156] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model*, *The Journal of Physical Chemistry* **97**, 10269–10280 (1993).
- [157] S. Mainberger, M. Kindlein, F. Bezold, E. Elts, M. Minceva, and H. Briesen, *Deep eutectic solvent formation: A structural view using Molecular Dynamics simulations with classical force fields*, *Molecular Physics* **115**, 1309–1321 (2017).
- [158] S. L. Perkins, P. Painter, and C. M. Colina, *Experimental and computational studies of choline chloride-based deep eutectic solvents*, *Journal of Chemical & Engineering Data* **59**, 3652–3662 (2014).
- [159] H. Liu, E. J. Maginn, A. E. Visser, N. J. Bridges, and E. B. Fox, *Thermal and transport properties of six ionic liquids: An experimental and Molecular Dynamics study*, *Industrial & Engineering Chemistry Research* **51**, 7242–7254 (2012).
- [160] D. Shah and F. S. Mjalli, *Effect of water on the thermo-physical properties of Reline: An experimental and molecular simulation based approach*, *Physical Chemistry Chemical Physics* **16**, 23900–23907 (2014).
- [161] V. V. Chaban, I. V. Voroshilova, and O. N. Kalugin, *A new force field model for the simulation of transport properties of imidazolium-based ionic liquids*, *Phys. Chem. Chem. Phys.* **13**, 7910–7920 (2011).
- [162] S. Blazquez, I. M. Zeron, M. M. Conde, J. L. F. Abascal, and C. Vega, *Scaled charges at work: Salting out and interfacial tension of methane with electrolyte solutions from computer simulations*, *Fluid Phase Equilibria* **513**, 112548 (2020).
- [163] J. Baz, C. Held, J. Pleiss, and N. Hansen, *Thermophysical properties of glycylene–water mixtures investigated by molecular modelling*, *Physical Chemistry Chemical Physics* **21**, 6467–6476 (2019).
- [164] H. S. Salehi, M. Ramdin, O. A. Mourtos, and T. J. H. Vlugt, *Computing solubility parameters of deep eutectic solvents from Molecular Dynamics simulations*, *Fluid Phase Equilibria* **497**, 10–18 (2019).
- [165] L. Martínez, R. Andrade, E. G. Birgin, and J. Martínez, *PACKMOL: a package for building initial configurations for molecular dynamics simulations*, *Journal of Computational Chemistry* **30**, 2157–2164 (2009).

- [166] D. Dubbeldam, D. C. Ford, D. E. Ellis, and R. Q. Snurr, *A new perspective on the order-n algorithm for computing correlation functions*, Molecular Simulation **35**, 1084–1097 (2009).
- [167] P. Ganguly and N. F. A. van der Vegt, *Convergence of sampling Kirkwood–Buff integrals of aqueous solutions with molecular dynamics simulations*, Journal of Chemical Theory and Computation **9**, 1347–1355 (2013).
- [168] H. Sun, Y. Li, X. Wu, and G. Li, *Theoretical study on the structures and properties of mixtures of urea and choline chloride*, Journal of molecular modeling **19**, 2433–2441 (2013).
- [169] A. Yadav and S. Pandey, *Densities and viscosities of (choline chloride+urea) deep eutectic solvent and its aqueous mixtures in the temperature range 293.15 K to 363.15 K*, Journal of Chemical & Engineering Data **59**, 2221–2229 (2014).
- [170] M. T. Humbert, Y. Zhang, and E. J. Maginn, *PyLAT: Python LAMMPS Analysis Tools*, Journal of Chemical Information and Modelling **59**, 1301–1305 (2019).
- [171] N. Dawass, R. R. Wanderley, M. Ramdin, O. A. Moulton, H. K. Knuutila, and T. J. H. Vlugt, *Solubility of carbon dioxide, hydrogen sulfide, methane, and nitrogen in monoethylene glycol; experiments and molecular simulation*, Journal of Chemical & Engineering Data **66**, 524–534 (2020), <https://doi.org/10.1021/acs.jced.0c00771>.
- [172] *Ethylene glycol*, <https://www.chemicalsafetyfacts.org/ethylene-glycol/>. Accessed on 11/08/2020.
- [173] *Monoethylene glycol (MEG) market by major drivers 2020 - global industry insights by global players, regional segmentation, growth, applications, value and forecasts till 2026*, <https://www-marketwatch-com.tudelft.idm.oclc.org/press-release/monoethylene-glycol-meg-market-by-major-drivers-2020—global-industry-insights-by-global-players-regional-segmentation-growth-applications-value-and-forecasts-till-2026-2020-08-03>. Accessed on 11/08/2020.
- [174] H. Yue, Y. Zhao, X. Ma, and J. Gong, *Ethylene glycol: properties, synthesis, and applications*, Chemical Society Reviews **41**, 4218–4244 (2012).

- [175] M. Cha, K. Shin, J. Kim, D. Chang, Y. Seo, H. Lee, and S. Kang, *Thermodynamic and kinetic hydrate inhibition performance of aqueous ethylene glycol solutions for natural gas*, Chemical Engineering Science **99**, 184–190 (2013).
- [176] W. Afzal, M. P. Breil, P. Théveneau, A. H. Mohammadi, G. M. Kontogeorgis, and D. Richon, *Phase equilibria of mixtures containing glycol and n-alkane: experimental study of infinite dilution activity coefficients and modeling using the cubic-plus-association equation of state*, Industrial & Engineering Chemistry Research **48**, 11202–11210 (2009).
- [177] D. Zheng, W. Ma, R. Wei, and T. Guo, *Solubility study of methane, carbon dioxide and nitrogen in ethylene glycol at elevated temperatures and pressures*, Fluid Phase Equilibria **155**, 277–286 (1999).
- [178] W. Afzal, M. P. Breil, I. Tsivintzelis, A. H. Mohammadi, G. M. Kontogeorgis, and D. Richon, *Experimental study and phase equilibrium modeling of systems containing acid gas and glycol*, Fluid Phase Equilibria **318**, 40–50 (2012).
- [179] A. Henni, P. Tontiwachwuthikul, and A. Chakma, *Solubilities of carbon dioxide in polyethylene glycol ethers*, Canadian Journal of Chemical Engineering **83**, 358–361 (2005).
- [180] H. Xu, C. Zhang, and Z. Zheng, *Solubility of hydrogen sulfide and carbon dioxide in a solution of methyldiethanolamine mixed with ethylene glycol*, Industrial & Engineering Chemistry Research **41**, 6175–6180 (2002).
- [181] C. Lin, R. B. Leron, A. R. Caparanga, and M. Li, *Henry's constant of carbon dioxide-aqueous deep eutectic solvent (choline chloride/ethylene glycol, choline chloride/glycerol, choline chloride/malonic acid) systems*, Journal of Chemical Thermodynamics **68**, 216–220 (2014).
- [182] R. B. Leron and M. Li, *Solubility of carbon dioxide in a choline chloride-ethylene glycol based deep eutectic solvent*, Thermochimica Acta **551**, 14–19 (2013).
- [183] R. Nagumo, Y. Muraki, S. Iwata, H. Mori, H. Takaba, and H. Yamada, *Molecular dynamics simulation study on CO<sub>2</sub> physical absorption mechanisms for ethylene-glycol-based solvents using free energy calculations*, Industrial & Engineering Chemistry Research **55**, 8200–8206 (2016).

- [184] E. Skylogianni, R. R. Wanderley, S. S. Austad, and H. K. Knuutila, *Density and viscosity of the nonaqueous and aqueous mixtures of methyldiethanolamine and monoethylene glycol at temperatures from 283.15 to 353.15 K*, Journal of Chemical & Engineering Data **64**, 5415–5431 (2019).
- [185] E. Skylogianni, C. Perinu, B. Y. C. Gamarosa, and H. K. Knuutila, *Carbon dioxide solubility in mixtures of methyldiethanolamine with monoethylene glycol, monoethylene glycol–water, water and triethylene glycol*, Journal of Chemical Thermodynamics **151**, 106176 (2020).
- [186] F. Jou, R. D. Deshmukh, F. D. Otto, and A. E. Mather, *Vapor–liquid equilibria of  $H_2S$  and  $CO_2$  and ethylene glycol at elevated pressures*, Chemical Engineering Communications **87**, 223–231 (1990).
- [187] F. S. Serpa, R. S. Vidal, J. H. B. A. Filho, J. F. d. Nascimento, J. R. P. Ciambelli, C. M. S. Figueiredo, G. R. Salazar-Banda, A. F. Santos, M. Fortuny, E. Franceschi, *et al.*, *Solubility of carbon dioxide in ethane-1, 2-diol–water mixtures*, Journal of Chemical & Engineering Data **58**, 3464–3469 (2013).
- [188] P. Ungerer, C. Nieto-Draghi, B. Rousseau, G. Ahunbay, and V. Lachet, *Molecular simulation of the thermophysical properties of fluids: From understanding toward quantitative predictions*, Journal of Molecular Liquids **134**, 71–89 (2007).
- [189] H. Yarveicy, M. M. Ghiasi, and A. H. Mohammadi, *Performance evaluation of the machine learning approaches in modeling of  $CO_2$  equilibrium absorption in Piperazine aqueous solution*, Journal of Molecular Liquids **255**, 375–383 (2018).
- [190] H. Yarveicy, H. Saghafi, M. M. Ghiasi, and A. H. Mohammadi, *Decision tree-based modeling of  $CO_2$  equilibrium absorption in different aqueous solutions of absorbents*, Environmental Progress & Sustainable Energy **38**, S441–S448 (2019).
- [191] L. Zhang and J. Siepmann, *Direct calculation of Henry's law constants from Gibbs ensemble Monte Carlo simulations: nitrogen, oxygen, carbon dioxide and methane in ethanol*, Theoretical Chemical Accounts **115**, 391–397 (2006).

- [192] S. H. Jamali, M. Ramdin, T. M. Becker, A. Torres-Knoop, D. Dubbeldam, W. Buijs, and T. J. H. Vlugt, *Solubility of sulfur compounds in commercial physical solvents and an ionic liquid from Monte Carlo simulations*, Fluid Phase Equilibria **433**, 50–55 (2017).
- [193] M. S. Shah, M. Tsapatsis, and J. I. Siepmann, *Monte Carlo simulations probing the adsorptive separation of hydrogen sulfide/methane mixtures using all-silica zeolites*, Langmuir **31**, 12268–12278 (2015).
- [194] U. Kapoor and J. K. Shah, *Monte Carlo simulations of pure and mixed gas solubilities of CO<sub>2</sub> and CH<sub>4</sub> in nonideal ionic liquid–ionic liquid mixtures*, Industrial & Engineering Chemistry Research **58**, 22569–22578 (2019).
- [195] W. Shi and E. J. Maginn, *Improvement in molecule exchange efficiency in Gibbs ensemble Monte Carlo: development and implementation of the Continuous Fractional Component move*, Journal of Computational Chemistry **29**, 2520–2530 (2008).
- [196] R. Hens, A. Rahbari, S. Caro-Ortiz, N. Dawass, M. Erdős, A. Poursaeidesfahani, H. Salehi, A. Celebi, M. Ramdin, O. A. Mourtos, D. Dubbeldam, and T. J. H. Vlugt, *Brick–CFCMC: Open source software for Monte Carlo simulations of phase and reaction equilibria using the Continuous Fractional Component method*, Journal of Chemical Information and Modelling **60**, 2678–2682 (2020).
- [197] A. Rahbari, R. Hens, I. K. Nikolaidis, A. Poursaeidesfahani, M. Ramdin, I. G. Economou, O. A. Mourtos, D. Dubbeldam, and T. J. H. Vlugt, *Computation of partial molar properties using Continuous Fractional Component Monte Carlo*, Molecular Physics **116**, 3331–3344 (2018).
- [198] M. G. Martin and J. I. Siepmann, *Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes*, Journal of Physical Chemistry B **102**, 2569–2577 (1998).
- [199] J. M. Stubbs, J. J. Potoff, and J. I. Siepmann, *Transferable potentials for phase equilibria. 6. United-atom description for ethers, glycols, ketones, and aldehydes*, Journal of Physical Chemistry B **108**, 17596–17605 (2004).
- [200] Q. Yang and C. Zhong, *Molecular simulation of carbon dioxide/methane/hydrogen mixture adsorption in metal–organic frameworks*, Journal of Physical Chemistry B **110**, 17776–17783 (2006).

- [201] G. Kamath, M. Ketko, G. A. Baker, and J. J. Potoff, *Monte Carlo predictions of phase equilibria and structure for dimethyl ether+ sulfur dioxide and dimethyl ether + carbon dioxide*, Journal of Chemical Physics **136**, 044514 (2012).
- [202] V. K. Michalis, I. N. Tsimpanogiannis, A. K. Stubos, and I. G. Economou, *Direct phase coexistence molecular dynamics study of the phase equilibria of the ternary methane–carbon dioxide–water hydrate system*, Physical Chemistry Chemical Physics **18**, 23538–23548 (2016).
- [203] S. Wang, L. Lu, D. Wu, X. Lu, W. Cao, T. Yang, and Y. Zhu, *Molecular simulation study of the adsorption and diffusion of a mixture of CO<sub>2</sub>/CH<sub>4</sub> in activated carbon: effect of textural properties and surface chemistry*, Journal of Chemical & Engineering Data **61**, 4139–4147 (2016).
- [204] O. A. Mourtos, I. N. Tsimpanogiannis, A. Z. Panagiotopoulos, and I. G. Economou, *Atomistic molecular dynamics simulations of CO<sub>2</sub> diffusivity in H<sub>2</sub>O for a wide range of temperatures and pressures*, Journal of Physical Chemistry B **118**, 5532–5541 (2014).
- [205] I. K. Nikolaidis, A. Poursaeidesfahani, Z. Csaszar, M. Ramdin, T. J. H. Vlugt, I. G. Economou, and O. A. Mourtos, *Modeling the phase equilibria of asymmetric hydrocarbon mixtures using molecular simulation and equations of state*, AIChE Journal **65**, 792–803 (2019).
- [206] J. Cardona, R. Fartaria, M. B. Sweatman, and L. Lue, *Molecular dynamics simulations for the prediction of the dielectric spectra of alcohols, glycols and monoethanolamine*, Molecular Simulation **42**, 370–390 (2016).
- [207] T. Kristóf and J. Liszi, *Effective intermolecular potential for fluid hydrogen sulfide*, Journal of Physical Chemistry B **101**, 5480–5483 (1997).
- [208] J. Vorholz, B. Rumpf, and G. Maurer, *Prediction of the vapor–liquid phase equilibrium of hydrogen sulfide and the binary system water–hydrogen sulfide by molecular simulation*, Physical Chemistry Chemical Physics **4**, 4449–4457 (2002).
- [209] F. Faure, B. Rousseau, V. Lachet, and P. Ungerer, *Molecular Simulation of the solubility and diffusion of carbon dioxide and hydrogen sulfide in polyethylene melts*, Fluid Phase Equilibria **261**, 168–175 (2007).
- [210] D.-Y. Peng and D. B. Robinson, *A new two-constant equation of state*, Industrial & Engineering Chemistry Fundamentals **15**, 59–64 (1976).

- [211] C. D. Wick, M. G. Martin, and J. I. Siepmann, *Transferable potentials for phase equilibria. 4. United-atom description of linear and branched alkenes and alkylbenzenes*, Journal of Physical Chemistry B **104**, 8008–8016 (2000).
- [212] J. J. Potoff and J. I. Siepmann, *Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen*, AIChE Journal **47**, 1676–1682 (2001).
- [213] M. S. Shah, M. Tsapatsis, and J. I. Siepmann, *Development of the transferable potentials for phase equilibria model for hydrogen sulfide*, Journal of Physical Chemistry B **119**, 7041–7052 (2015).
- [214] J. Gross and G. Sadowski, *Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules*, Industrial & Engineering Chemistry Research **40**, 1244–1260 (2001).
- [215] N. I. Diamantonis, G. C. Boulogouris, E. Mansoor, D. M. Tsangaris, and I. G. Economou, *Evaluation of cubic, SAFT, and PC-SAFT equations of state for the vapor–liquid equilibrium modeling of CO<sub>2</sub> mixtures with other gases*, Industrial & Engineering Chemistry Research **52**, 3933–3942 (2013).
- [216] W. Shi and E. J. Maginn, *Atomistic simulation of the absorption of carbon dioxide and water in the ionic liquid 1-n-Hexyl-3-methylimidazolium Bis(trifluoromethylsulfonyl) imide ([hmim][Tf<sub>2</sub>N])*, Journal of Physical Chemistry B **112**, 2045–2055 (2008).
- [217] A. Torres-Knoop, S. P. Balaji, T. J. H. Vlugt, and D. Dubbeldam, *A comparison of advanced Monte Carlo methods for open systems: CFCMC vs CBMC*, Journal of Chemical Theory and Computation **10**, 942–952 (2014).
- [218] T. T. Pham and M. R. Shirts, *Identifying low variance pathways for free energy calculations of molecular transformations in solution phase*, Journal of Chemical Physics **135**, 034114 (2011).
- [219] A. Rahbari, R. Hens, S. H. Jamali, M. Ramdin, D. Dubbeldam, and T. J. H. Vlugt, *Effect of truncating electrostatic interactions on predicting thermodynamic properties of water–methanol systems*, Molecular Simulation **45**, 336–350 (2019).
- [220] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, *Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins*, Journal of Chemical Physics **119**, 5740–5761 (2003).

- [221] M. R. Shirts and V. S. Pande, *Solvation free energies of amino acid side chain analogs for common molecular mechanics water models*, Journal of Chemical Physics **122**, 134508 (2005).
- [222] S. G. Lias, J. E. Bartmess, J. F. Liebman, J. L. Holmes, R. D. Levin, and W. G. Mallard, *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, National Institute of Standards and Technology, Gaithersburg MD **20899** (2018).
- [223] D. Dubbeldam, S. Calero, and T. J. H. Vlugt, *iRASPA: GPU-accelerated visualization software for materials scientists*, Molecular Simulation **44**, 653–676 (2018).
- [224] R. R. Wanderley, D. D. D. Pinto, and H. K. Knuutila, *Investigating opportunities for water-lean solvents in CO<sub>2</sub> capture: VLE and heat of absorption in water-lean solvents containing MEA*, Separation and Purification Technology **231**, 115883 (2020).
- [225] A. C. Galvão and A. Z. Francesconi, *Solubility of methane and carbon dioxide in ethylene glycol at pressures up to 14 MPa and temperatures ranging from (303 to 423) K*, Journal of Chemical Thermodynamics **42**, 684–688 (2010).
- [226] J. M. Prausnitz, R. N. Lichtenthaler, and E. G. De Azevedo, *Molecular thermodynamics of fluid-phase equilibria*, 3rd ed. (Pearson Education, Upper Saddle River, N.J, 1998).
- [227] K. S. Shing, K. E. Gubbins, and K. Lucas, *Henry constants in non-ideal fluid mixtures: computer simulation and theory*, Molecular Physics **65**, 1235–1252 (1988).
- [228] M. Ramdin, S. P. Balaji, J. M. Vicent-Luna, J. J. Gutiérrez-Sevillano, S. Calero, T. W. de Loos, and T. J. H. Vlugt, *Solubility of the precombustion gases CO<sub>2</sub>, CH<sub>4</sub>, CO, H<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>S in the ionic liquid [bmim][Tf<sub>2</sub>N] from Monte Carlo simulations*, Journal of Physical Chemistry C **118**, 23599–23604 (2014).
- [229] R. R. Wanderley, S. Evjen, D. D. D. Pinto, and H. K. Knuutila, *The salting-out effect in some physical absorbents for CO<sub>2</sub> capture*, Chemical Engineering Transactions **69** (2018).

- [230] M. Ramdin, Q. Chen, S. P. Balaji, J. M. Vicent-Luna, A. Torres-Knoop, D. Dubbeldam, S. Calero, T. W. de Loos, and T. J. H. Vlugt, *Solubilities of CO<sub>2</sub>, CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, and SO<sub>2</sub> in ionic liquids and selexol from Monte Carlo simulations*, Journal of Computational Science **15**, 74–80 (2016).



# SUMMARY

The Kirkwood–Buff (KB) theory is one of the most rigorous solution theories that connects molecular structure to macroscopic behaviour. The key quantity, the so-called KB Kirkwood–Buff Integrals (KBIs), are defined either in terms of fluctuations in the number of molecules or integrals over radial distribution functions over open subvolumes. In the grand-canonical ensemble, KBIs of infinitely large and open systems are directly related to thermodynamic properties such as partial derivatives of chemical potentials and partial molar volumes. Using molecular simulations, it is only possible to study small systems with a finite number of molecules, and therefore finite-size effects should be considered. In chapter 1, a literature review of studies of KBIs was conducted. This review focused on: (1) inversion of the KB theory, where KBIs are estimated from macroscopic properties; (2) available methods to compute KBIs from molecular simulations, and (3) applications of KBIs to molecular systems. Generally, three levels of estimations for KBIs were used in literature: (1) the easiest, and most common approach is to simply truncate KBIs of open and infinite systems to half the size of the simulation box; (2) a number of mathematical involved methods were developed that are not easily extended to complex molecules; (3) The approach of Krüger and co-workers provides an approach that is more accurate than truncating KBIs and with an intermediate difficulty. To compute KBIs from molecular simulations, Krüger and co-workers derived an expression for KBIs of finite and open subvolumes embedded in larger reservoirs. According to thermodynamics of small systems (nanothermodynamics), thermodynamic properties of small systems scale linearly with the inverse size of the subvolume. Extrapolating KBIs to the thermodynamic limit yields KBIs of open and infinite systems. In this thesis, various aspects related to estimating KBIs from molecular simulations of finite and closed systems were investigated. As a result, an improved framework to compute KBIs accurately and conveniently is developed. The methodology was used to compute KBIs of model systems, and realistic solutions.

In chapter 2, shape effects of KBIs from molecular simulations are investigated. The dependence of KBIs on the shape and dimensionality of the subvolume is characterised by a weightfunction  $w(x)$ . A method to numerically compute the weightfunction  $w(x)$  for any arbitrary convex subvolume was developed. We computed KBIs of an analytic Radial Distribution Function (RDF) for

the following shapes of the subvolume: sphere, cube, and cuboids and spheroids with different aspect ratios. It turns out that in the thermodynamic limit, KBIs are independent of the shape of the used subvolume.

In chapter 3, finite-size effects related to computing KBIs from molecular simulations were investigated. Finite-size effects originating from the size of the system and the size of the used subvolumes were studied using an analytic RDF. From KBIs computed for different sizes of the subvolume, we found that the size of the subvolume should not exceed the size of the simulation box. In this chapter, finite-size effects related to computing RDFs from closed systems were also studied for a Weeks-Chandler-Andersen (WCA) fluid. RDFs computed from Molecular Dynamics (MD) simulations of closed and finite systems were corrected to estimate RDFs of open systems, which are required by the KB theory. The performance of three correction methods was assessed: (1) a  $1/N$  correlation [74], (2) a correction by Ganguly and van der Vegt [113], and (3) an expression by Cortes-Huerto and co-workers [83]. The correction by Ganguly and van der Vegt was found to provide the most accurate KBIs. KBIs in the thermodynamic limit were obtained from the linear part of the scaling of KBIs of small subvolumes ( $G_{\alpha\beta}^V$ ) with the inverse size of the system  $1/L$ . Identifying a linear regime was not straightforward, and some guidelines were provided.

In chapter 4, KBIs of Lennard-Jones (LJ) and WCA fluids at various densities were computed. To obtain KBIs in the thermodynamic limit from RDFs computed using MD simulations of finite systems, three extrapolation methods were considered. All extrapolation methods resulted in similar estimations of KBIs, however, the scaling of  $LG_{\alpha\beta}^V$  with  $L$  was found the easiest to use. Additionally, surface effects of KBIs of LJ and WCA fluids were quantified. The results demonstrated that for LJ and WCA systems, surface terms can be of the same order of magnitude of KBIs.

In chapter 5, KBIs of mixtures of urea and choline-chloride (ChCl) were computed using MD simulations. The system was studied at  $T = 343.15$  K, atmospheric pressure, and molar ratios of ChCl to urea ranging from 2:1 to 1:5. RDFs were corrected using the Ganguly and van der Vegt correction, and the scaling of  $LG_{\alpha\beta}^V$  with  $L$  was used to obtain KBIs in the thermodynamic limit. KBIs were used to examine the affinity between the various components at different molar ratios. From the KBIs, thermodynamic factors and partial molar volumes were directly computed. Also, from MD simulations, a number of transport properties were studied including MS diffusivities, that are connected to Fick diffusion coefficients via the thermodynamic factors. It was shown that the KB theory is advantageous for simultaneously studying molecular interactions and computing thermodynamic properties of salt solutions.

The KB approach proves to be a useful tool for studying interactions in multi-component fluids. As KBIs are linked to fluctuations in the grand-canonical ensemble, thermodynamic properties can be computed while avoiding insertion and deletion of molecules. However, when computing some thermodynamic quantities, molecular simulations in open ensembles are commonly used. Phase equilibria and solubilities are typically computed from Monte Carlo (MC) simulations in open ensembles. In chapter 6, the solubilities of CO<sub>2</sub>, CH<sub>4</sub>, H<sub>2</sub>S, and N<sub>2</sub> in monoethylene glycol (MEG) were computed using MC simulations in the osmotic ensemble. The Continuous Fractional Component Monte Carlo (CFCMC) method was used to enhance the efficiency of MC simulations. Classical force fields were used without any reparameterisation. Predictions of CFCMC calculations were compared to experimental data. Good agreement between simulations and experiments was observed when comparing solubilities at low pressures. At high pressures, force field modifications may be required to improve predicted solubilities of small gas molecules in MEG.



## SAMENVATTING

De Kirkwood-Buff theorie is een van de meest rigoureuze vloeistoftheorieën die moleculaire structuur verbindt met macroscopisch gedrag. De belangrijkste grootheid, de zogenaamde Kirkwood-Buff Integraal, wordt gedefinieerd in termen van fluctuaties in het aantal moleculen in een bepaald volume, of door de integraal van de radiale distributiefunctie over een open volume. In het groot-canonieke ensemble zijn Kirkwood-Buff integralen van oneindig grote en open systemen direct gerelateerd aan thermodynamische eigenschappen zoals partiële afgeleiden van chemische potentialen en partiële molaire volumes. Met behulp van moleculaire simulaties is het alleen mogelijk om relatief kleine systemen met een eindig aantal moleculen te bestuderen, en daarom moet er rekening worden gehouden met zogenaamde finite-size effecten. In hoofdstuk 1 van dit proefschrift is een literatuuroverzicht van studies over Kirkwood-Buff integralen uitgevoerd. Deze studie was gericht op: (1) de inversie van de Kirkwood-Buff theorie, waarbij Kirkwood-Buff integralen worden geschat op basis van macroscopische eigenschappen; (2) methoden om Kirkwood-Buff integralen te berekenen uit moleculaire simulaties, en (3) toepassingen van Kirkwood-Buff integralen op moleculaire systemen. In het algemeen worden in de literatuur drie methodes voor het berekenen van Kirkwood-Buff integralen gebruikt: (1) de gemakkelijkste en meest gebruikelijke benadering is om de uitdrukking van Kirkwood-Buff integralen voor open en oneindige systemen af te kappen tot de helft van de grootte van de simulatiedoos; (2) in de literatuur zijn een aantal complexe methoden ontwikkeld die niet gemakkelijk kunnen worden uitgebreid tot complexe moleculen; (3) de niet al te complexe methode van Krüger en collega's levert een benadering op die nauwkeuriger is dan afgekapte Kirkwood-Buff integralen. Om Kirkwood-Buff integralen te berekenen uit moleculaire simulaties hebben Krüger en collega's een uitdrukking afgeleid voor Kirkwood-Buff integralen van eindige en open subvolumes ingebed in grotere reservoirs. Volgens de thermodynamica van kleine systemen (nanothermodynamica) schalen de thermodynamische eigenschappen van kleine systemen lineair met de inverse grootte van het subvolume. Extrapolatie van Kirkwood-Buff integralen naar de thermodynamische limiet levert Kirkwood-Buff integralen van open en oneindige systemen op. In dit proefschrift zijn verschillende aspecten onderzocht die verband houden met het afschatten van Kirkwood-Buff integralen uit moleculaire simulaties van eindige en gesloten systemen. Er is een verbeterd raamwerk ontwikkeld om

Kirkwood-Buff integralen nauwkeurig en gemakkelijk te berekenen. Deze methodologie werd gebruikt om Kirkwood-Buff integralen van modelsystemen en realistische oplossingen te berekenen.

In hoofdstuk 2 worden vormeffecten van Kirkwood-Buff integralen berekend uit moleculaire simulaties onderzocht. De afhankelijkheid van Kirkwood-Buff integralen van de vorm en dimensionaleit van het subvolume wordt gekenmerkt door een gewichtsfunctie  $w(x)$ . Er werd een methode ontwikkeld om deze gewichtsfunctie  $w(x)$  numeriek te berekenen voor een willekeurig convex subvolume. We hebben Kirkwood-Buff integralen van een analytische radiale distributiefunctie berekend voor de volgende vormen van het subvolume: bol, kubus en kubussen en sferoiden met verschillende aspectverhoudingen. Het blijkt dat Kirkwood-Buff integralen in de thermodynamische limiet onafhankelijk zijn van de vorm van het subvolume.

In hoofdstuk 3 worden finite-size effecten gerelateerd aan het berekenen van Kirkwood-Buff integralen uit moleculaire simulaties onderzocht. Finite-size effecten die voortkomen uit de grootte van het systeem en de grootte van de gebruikte subvolumes werden bestudeerd met behulp van een analytische radiale distributiefunctie. Uit Kirkwood-Buff integralen berekend voor verschillende groottes van het subvolume werd gevonden dat de grootte van het subvolume de grootte van de simulatiedoos niet mag overschrijden. In dit hoofdstuk werden ook finite-size effecten gerelateerd aan het berekenen van radiale distributiefuncties voor gesloten systemen bestudeerd voor een Weeks-Chandler-Andersen (WCA) vloeistof. Radiale distributiefuncties berekend met Moleculaire Dynamica (MD) simulaties van gesloten en eindige systemen werden gecorrigeerd om radiale distributiefuncties van open systemen te schatten, die vereist zijn in de Kirkwood-Buff theorie. De prestaties van drie correctiemethoden werd onderzoek: (1) een  $1/N$  correlatie [74], (2) een correctie door Ganguly en van der Vegt [113], en (3) een methode ontwikkeld door Cortes-Huerto en collega's [83]. De methode van Ganguly en van der Vegt bleek de meest nauwkeurige Kirkwood-Buff integralen te leveren. Kirkwood-Buff integralen in de thermodynamische limiet werden verkregen uit het lineaire deel van de schaling van Kirkwood-Buff integralen van kleine subvolumes ( $G_{\alpha\beta}^V$ ) met de inverse grootte van het systeem  $1/L$ . Identificatie van een lineair regime was niet altijd eenvoudig en hiervoor werden enkele richtlijnen gegeven.

In hoofdstuk 4 werden Kirkwood-Buff integralen van Lennard-Jones (LJ) en WCA-vloeistoffen met verschillende dichthesen onderzocht. Om Kirkwood-Buff integralen in de thermodynamische limiet te verkrijgen uit radiale distributiefuncties berekend met behulp van MD simulaties van eindige systemen, werden drie extrapolatiemethoden onderzocht. Alle extrapolatiemethoden resulteerden in vergelijkbare schattingen van Kirkwood-Buff integralen. De schaling van het

product  $LG_{\alpha\beta}^V$  met  $L$  bleek het gemakkelijkst te gebruiken. Bovendien werden oppervlakte-effecten van Kirkwood-Buff integralen van LJ- en WCA-vloeistoffen gekwantificeerd. De resultaten toonden aan dat voor LJ- en WCA-systemen oppervlakte-termen van dezelfde orde van grootte kunnen zijn als de Kirkwood-Buff integralen zelf.

In hoofdstuk 5 werden Kirkwood-Buff integralen van mengsels van ureum en choline-chloride ( $\text{ChCl}$ ) berekend met behulp van MD simulaties. Het systeem werd bestudeerd bij  $T=343,15$  K, atmosferische druk en molverhoudingen  $\text{ChCl} / \text{ureum}$  variërend van 2:1 tot 1:5. Radiale distributiefuncties werden gecorigeerd met behulp van de Ganguly en van der Vegt-correctie en de schaling van het product  $LG_{\alpha\beta}^V$  met  $L$  werd gebruikt om Kirkwood-Buff integralen in de thermodynamische limiet te verkrijgen. Kirkwood-Buff integralen werden gebruikt om de affiniteit tussen de verschillende componenten bij verschillende molverhoudingen te onderzoeken. Uit de Kirkwood-Buff integralen werden thermodynamische factoren en partiële molaire volumes direct berekend. Ook werden met behulp van MD simulaties een aantal transporteigenschappen bestudeerd, waaronder Maxwell-Stefan diffusiecoëfficiënten, die via de matrix van thermodynamische factoren verbonden zijn met Fick diffusiecoëfficiënten. Er werd aangetoond dat de Kirkwood-Buff theorie handig is voor het gelijktijdig bestuderen van moleculaire interacties en het berekenen van thermodynamische eigenschappen van systemen bestaande uit ionen.

De Kirkwood-Buff methode blijkt een nuttig instrument om interacties in vloeistoffen bestaande uit meerdere componenten te bestuderen. Aangezien Kirkwood-Buff integralen gekoppeld zijn aan fluctuaties in het groot-canonieke ensemble, kunnen macroscopische thermodynamische eigenschappen worden berekend terwijl het inserteren en verwijderen van moleculen (zoals vereist in het groot-canonieke ensemble) wordt vermeden. Bij het berekenen van sommige thermodynamische grootheden worden echter vaak moleculaire simulaties in open ensembles gebruikt. Fase-evenwichten en oplosbaarheden worden typisch berekend uit Monte Carlo (MC) simulaties in open ensembles. In hoofdstuk 6 werden de oplosbaarheden van  $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{H}_2\text{S}$ , and  $\text{N}_2$  in monoethyleenglycol (MEG) berekend met MC simulaties in het osmotische ensemble. De zogenaamde Continuous Fractional Component Monte Carlo (CFCMC) methode werd gebruikt om de efficiëntie van deze MC simulaties te verbeteren. Klassieke krachtvelden werden gebruikt zonder enige additionele parametrisering. Voor-spellingen van CFCMC-berekeningen werden vergeleken met experimentele gegevens. Een goede overeenkomst tussen simulaties en experimenten werd waargenomen voor oplosbaarheden bij lage druk. Bij hoge druk kunnen modificaties van het krachtveld nodig zijn om de voorspelde oplosbaarheid van kleine gasmoleculen in MEG te verbeteren.



# CURRICULUM VITAE

Noura Dawass was born on 30 September 1992 in Doha, Qatar. In 2014, she received her Bachelor of Science in Chemical Engineering from Qatar University. In the same year, she joined the Chemical Engineering program at Texas A&M University at Qatar to pursue a Master of Science degree. As a Masters student, she worked under the supervision of Prof. Marcelo Castier and Prof. Ioannis Economou on the development of an equation of state for fluids in porous media. After completing her Masters degree in 2016, Noura moved to the Netherlands to join the Engineering Thermodynamic group at the Process and Energy department at Delft University of Technology. As a Ph.D. researcher, she was supervised by Prof. dr. ir. T. J. H. Vlugt and Dr. O. A. Moulton. Her research work focused on the development of methods for computing Kirkwood-Buff Integrals from molecular simulations. She worked on this topic in collaboration with Prof. Peter Krüger (Chiba University, Japan), Dr. Jean-Marc Simon (University of Burgundy, France), and Dr. Sondre Schnell (Norwegian University of Science and Technology, Norway).



# LIST OF PUBLICATIONS

Publications included in this thesis:

1. N. Dawass, P Krüger, S K Schnell, D Bedeaux, S Kjelstrup, J M Simon, and T J H Vlugt, *Finite-size effects of Kirkwood–Buff integrals from molecular simulations*, Molecular Simulation, **44**, 599–612 (2017).
2. N. Dawass, P Krüger, J M Simon, and T J H Vlugt, *Kirkwood–Buff integrals of finite systems: Shape effects*, Molecular Physics, **116**, 1573–1580 (2018).
3. N. Dawass, P Krüger, S K Schnell, J M Simon, and T J H Vlugt, *Kirkwood–Buff integrals from molecular simulation*, Fluid Phase Equilibria, **486**, 21–36 (2019).
4. N. Dawass, P Krüger, S K Schnell, O. A. Mourtos, I. G. Economou, T J H Vlugt, and J M Simon, *Kirkwood–Buff Integrals Using Molecular Simulation: Estimation of Surface Effects*, Nanomaterials, **10**, 771 (2020).
5. N. Dawass, R. R. Wanderley, M. Ramdin, O. A. Mourtos, H. K. Knuutila and T.J.H. Vlugt, *Solubility of Carbon Dioxide, Hydrogen Sulfide, Methane, and Nitrogen in Monoethylene Glycol; Experiments and Molecular Simulation*, Journal of Chemical & Engineering Data, **66**, 524–534 (2020).
6. A. T. Celebi, N. Dawass, O. A. Mourtos, and T.J.H. Vlugt, *How sensitive are physical properties of choline chloride–urea mixtures to composition changes: Molecular Dynamics simulations and Kirkwood Buff theory*, Journal of Chemical Physics, **154**, 184502 (2021).

Publications not included in this thesis:

1. R. Hens, A. Rahbari, S. Caro–Ortiz, N. Dawass, M. Erdős, A. Poursaeidesfahani, H. Salehi, A. Celebi, M. Ramdin, O. A. Mourtos, D. Dubbeldam, and T. J. H. Vlugt, *Brick–CFCMC: Open source software for Monte Carlo simulations of phase and reaction equilibria using the Continuous Fractional Component method*, Journal of Chemical Information and Modelling, **60**, 2678–2682 (2020).



## ACKNOWLEDGMENTS

The time I spent in Delft as a PhD candidate was special. I am extremely grateful to have so many people to thank for the guidance, support, and good memories.

I would like to start with thanking my promotor and supervisor, Thijs Vlugt. I feel privileged to have known you and worked with you. I am deeply indebted to your continuous feedback, scientific insights, and support throughout my PhD. I learned a lot from you and I have been inspired by your devotion to your work. I would like to extend my appreciation to my co-promotor and supervisor Othon (Otto) Moultsos. I was extremely lucky to have you by my side during my PhD, not only as a scientist and a supervisor, but also as a friend. Thank you for the mood lifting conversations and the sincere advise. Most of all, I am grateful to you and to your wife, Sophia, for making me feel like I have a family in the Netherlands. To complete this work, I had the privilege to work in collaboration with researchers from different scientific backgrounds and from different universities. Peter Krüger, Jean-Marc Simon, Sondre Schnell, Signe Kjelstrup and Dick Bedaux. It was a delight to collaborate with all of you on the topic of Kirkwood-Buff integral. I appreciated the valuable feedback, interesting questions, and discussions as they have enriched this research project and broadened my knowledge. Thanks for Hanna Kuutila and Ricardo Wanderley from NTNU for the collaboration and supporting this work with their experimental measurements. I would also like to thank Ioannis Economou, my former masters supervisor, for hosting me as a visiting PhD in TAMUQ. It is always a pleasure to collaborate and work together.

As a member of the Engineering Thermodynamics group, I was lucky to be surrounded by helpful and friendly colleagues. We had shared many experiences together: group meetings, trips, and workshops. I would like to specially thank Ali, Reza and Seyed. Not only for the unlimited support with technical issues, but more importantly for the welcoming and warm attitude, especially at the beginning of my PhD. Thanks Mahinder for collaborating with me on the solubilities in MEG project, your input was appreciated. Thank you Remco, for your exceptional work on Brick and for always being kind and patient. I would like to also thank Alper. I truly enjoyed collaborating with you and although we mostly met online due to the pandemic, working with you was always easy and delightful. To my colleagues Sebastian, Tim, Hirad, Mate, and Max, thanks for the good times.

During my time at P&E, I had the opportunity to meet many lovely people. Many of them will always be my friends. I would especially like to thank Vilborg for the good times we spent together and for always taking the initiative to arrange activities and outings that involve as many people as possible, inside and outside the department. Thanks Karsten, Carla, Samira, Vincent, Mohamed, Hakan, Mengmeng and Elyas for joining those activities and for the nice conversations during working hours. Thanks Gustavo and Nikos for creating a lively atmosphere in our office, for introducing me to your lovely wives, and for the constant moral support. Thanks Rishabh, Simone, Stephan and Manas for the fun times. Our hangouts with Fatma made the pandemic much more bearable. Last but not least, thank you Fatma for being sweet and supportive, and for our beautiful friendship. I will particularly miss drinking coffee with you by the canals.

I would like to also thank all of my friends for the nice times that helped me recharge and escape. Thanks to the friends who I met in the Netherlands, Zahra, Sarah, Ezgi, and Joyce. I am also grateful for my close friend Nouf. Thanks for the lovely times. Meeting with you always felt like home. Thanks for my friends back in Qatar, Dana, Rym, and Michelle. Warmest thanks to my best friend, Tamara, who I felt her love and support even when I was so far away.

Finally, I would like to thank my family. Thanks to my parents for believing in me and supporting me and working hard for our family. Thank you mama for always giving out what you can and for your sacrifices. Thanks to my siblings for teaching me so much. Especially, I would like to thank my sister Dania. You were always there for me. I couldn't have done this without you.