From Finite-Size Thermodynamics to Grand Canonical Molecular Dynamics

Dissertation

zur Erlangung des Grades
"Doktor der Naturwissenschaften"
am Fachbereich 08: Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität in Mainz

von

Jose Mauricio Sevilla Moreno

geboren in Bogotá, Kolumbien

Advisor: [redacted]
Day of the exam: 06.10.2025

From Finite-Size Thermodynamics to Grand Canonical Molecular Dynamics

Jose Mauricio Sevilla Moreno

2026

Contents

Preamble

Publications

Conclusions

List of Figures

Included Publications

This cumulative PhD thesis is made out of the compilation of several research studies that are presented in the second part of this document in six research articles. All of them have already been published at the moment of submission of this document in several peer-reviewed scientific journals (Chapters: I,II,III,IV,V and VI):

§ Luis A. Baptista, Ravi C Dutta, Mauricio Sevilla, Maziar Heidari, Raffaello Potestio, Kurt Kremer, and Robinson Cortes-Huerto. Density-functional-theory approach to the hamiltonian adaptive resolution simulation method. *Journal of Physics: Condensed Matter*, 33(18):184003, apr 2021

§ Atreyee Banerjee[1], Mauricio Sevilla[1], Joseph F. Rudzinski, and Robinson Cortes-Huerto. Finite-size scaling and thermodynamics of model supercooled liquids: long-range concentration fluctuations and the role of attractive interactions. *Soft Matter*, 18:2373–2382, 2022.

§ Mauricio Sevilla and Robinson Cortes-Huerto. Connecting density fluctuations and Kirkwood–Buff integrals for finite-size systems. *The Journal of Chemical Physics*, 156(4):044502, 01 2022.

§ Mauricio Sevilla, Atreyee Banerjee, and Robinson Cortes-Huerto. Finite-size excess-entropy scaling for simple liquids. *The Journal of Chemical Physics*, 158(20):204502, 05 2023

§ Luis A. Baptista[2], Mauricio Sevilla[2], Manfred Wagner, Kurt Kremer, and Robinson Cortes-Huerto. Chilling alcohol on the computer: isothermal compressibility and the formation of hydrogen-bond clusters in liquid propan-1-ol. *The European Physical Journal E*, 46(117), 11 2023

§ Mauricio Sevilla, Luis A. Baptista, Kurt Kremer, and Robinson Cortes-Huerto. Density Fluctuations, Solvation Thermodynamics and Coexistence Curves in Grand Canonical Molecular Dynamics Simulations. *The Journal of Chemical Physics*, 162 (8): 080901, 02 2025

The individual contributions to each of these articles are presented as published individually in every section.

[1]Both authors contributed equally to the paper
[2]Both authors contributed equally to the paper

Abstract

Molecular Dynamics (MD) simulations play an increasing role in the studies of soft matter and statistical mechanics. The limitations of MD are unavoidable when comparing results to both, experiments and theory as it is practically impossible to reach, computationally, large time and size scales. Nevertheless, the effects of these limitations can be understood and described in a statistical mechanics compatible manner, therefore corrected. These effects, known as *finite-size* effects, come due to finite integration domains, differences in the ensembles and the contributions attributed to periodic boundary conditions (PBCs) used to mimic an infinite system. Overcoming these limitations require either larger simulation boxes, implying the need of more resources, or new techniques and methods to extract information and to extrapolate it to the thermodynamic limit (TL). In this thesis we focused in the latter. We divided this work into different branches keeping as a ground basis, the connection between thermodynamics and computer simulations.

First, we took the case of the Kirkwood-Buff integrals (KBI), as they provide a clean link between thermodynamic properties and microscopic quantities, where the last ones are accessible directly by MD simulations. We identify and corrected a finite-size version of the KBI, bridging the density fluctuations with integrals of the radial distribution functions, identifying and correcting independently all sources of finite-size effects during the process. Along these lines, we extended the analysis to the computation of the finite-size two-body excess-entropy s_2, where despite the fact of having no ensemble dependent finite-size effect it does scale with the inverse of the simulation box, and more importantly, we found that the empirical Dzugutov scaling relation also exhibits a finite-size dependence, where our values in the thermodynamic limit, agree with those of the literature.

Second, going to applications to the computation of thermodynamics, we moved to the understanding of structural properties of liquids by using KBIs, structure factors and chemical potentials. On the one hand, we studied the effect of small differences in the tail of the radial distribution functions (RDFs) on the structure of prototypical supercooled liquids, Kob-Andersen mixtures with and without attractive interactions. Our results suggest that the nucleation of long-range domains is induced by the attractive potentials when decreasing temperature. This was verified using the limit $k \to 0$ of the Bathia-Thornton structure factors and the bulk isothermal compressibility. Moreover, after performing similar studies were performed at higher density we found the two systems indistinguishable regarding structure. Whereas on the other hand, we found similar behaviour of the creation of domains, for propan-1-ol simulations for a range of temperature $200 < T < 300$ K. Where around ~ 220 K a change in the isothermal compressibility points out to a transition from a normal to an anomalous liquid. The mechanism for the creation of domains in this case is rather different, as we found them to nucleate from the creation of well defined networks of molecules connected via Hydrogen bonds (H-bonds). This results were verified by the computation of the average number of H-bonds inside the simulation box, the size of the larges cluster, and more importantly NMR experiments.

Lastly, the development of a Density-functional-theory (DFT) approach to the Hamiltonian Adaptive Resolution Simulation (`H-AdResS`) allowed us to show that the external potential needed to induce a constant density profile coincides with the system's excess chemical potential. Hence, given the one-to-one correspondence between equilibrium densities and external potentials in DFT, it is possible to obtain the excess of chemical potential by imposing a constant density profile. Which brings us to the most important point of this work, and that is to the study of the statistics

of the atomistic region in the (H-AdResS) method. We considered a slab geometry where the atomistic region was taken to be the system and the hybrid + Ideal gas the thermal bath. This implies that the contact system-bath is done purely in one coordination whereas the other two keep their PBCs. The consequences of these considerations, is that it is necessary to have a description of a grand canonical simulation where size effects are still present, more specifically, PBCs contributions. This description was proposed and validated as part of this thesis, were we identified and corrected all sources of size effects in the Kirkwood-Buff integrals and therefore, in the density fluctuations. We computed the local density fluctuations inside the atomistic region, and compare those with the results coming from our method to compute the finite-size Kirkwood-Buff integrals finding that in order to ensure the correct stable grand canonical statistics having a H-AdResS set up is not enough. Nevertheless, a correct description of the grand canonical ensemble in the thermal baths, meaning the ideal gas, actually leads to an excellent agreement between the H-AdResS and the finite-size integral equation. The way to improve the description of the bath, is by introducing a particle insertion/deletion algorithm with the correct statistics which is possible due to the fact that the insertion probabilities for an ideal gas in the grand canonical ensemble can be computed analytically. Summarizing, the `Adaptive Resolution with Particle insertion/deletion Steps` (AdResS+ PI) can be used as a grand canonical simulation method.

Zusammenfassung

Molekulardynamische Simulationen (MD) spielen eine wichtige Rolle bei der Erforschung der weichen Materie und der statistischen Mechanik. Die Grenzen der MD sind unvermeidlich, wenn die Ergebnisse sowohl mit Experimenten als auch der Theorie verglichen werden, da es praktisch unmöglich ist, große Zeit- und Größenskalen rechnerisch zu erreichen. Die Auswirkungen dieser Einschränkungen können jedoch verstanden und in einer mit der statistischen Mechanik kompatiblen Weise beschrieben und somit korrigiert werden. Diese Effekte, die als *Finite-Size*-Effekte bekannt sind, ergeben sich aus endlichen Integrationsbereichen, Unterschiede in den Ensembles und Beiträgen, die periodischen Randbedingungen (Periodic Boundary Conditions, PBCs) zugeschrieben werden, die zur Simulation eines unendlichen Systems verwendet werden. Um diese Einschränkungen zu überwinden, sind entweder größere Simulationsboxen erforderlich, die mehr Ressourcen erfordern, oder neue Techniken und Methoden, um Informationen zu extrahieren und bis zum thermodynamischen Grenzfall (Thermodynamic Limit, TL) zu extrapolieren. In dieser Arbeit haben wir uns auf die letztere Möglichkeit konzentriert. Wir haben diese Arbeit in verschiedene Zweige unterteilt, wobei wir die Verbindung zwischen Thermodynamik und Computersimulation als Grundlage beibehalten haben.

Zuerst haben wir uns mit den Kirkwood-Buff-Integralen (KBI) beschäftigt, da sie eine klare Verbindung zwischen thermodynamischen Eigenschaften und mikroskopischen Größen herstellen, wobei letztere direkt durch MD-Simulationen zugänglich sind. Wir haben eine endliche Version des KBI identifiziert und sie korrigiert, indem wir die Dichtefluktuationen mit Integralen radialer Verteilungsfunktionen überbrückt haben, wobei alle Quellen endlicher Effekte unabhängig voneinander identifiziert und korrigiert wurden. In diesem Sinne haben wir die Analyse auf die Berechnung der Zweikörper-Exzessentropie s_2 in endlicher Größe erweitert, wo sie trotz der Tatsache, dass es keine ensembleabhängigen Finite-Size-Effekte gibt, mit dem Kehrwert der Simulationsbox skaliert. Noch bedeutungsvoller ist, dass wir herausfanden,dass die empirische Dzugutov-Skalierungsrelation ebenfalls eine Finite-Size-Abhängigkeit aufweist, wobei unsere Werte im thermodynamischen Grenzfall mit denen in der Literatur übereinstimmen.

Zweitens sind wir von Anwendungen der Thermodynamik zum Verständnis der strukturellen Eigenschaften von Flüssigkeiten übergegangen, indem wir KBIs, Strukturfaktoren und chemische Potentiale verwendet haben. Einerseits untersuchten wir die Auswirkungen kleiner Unterschiede in den Radialen Verteilungsfunktionen (Radial Distribution Functions, RDFs) auf die Struktur von prototypischen unterkühlten Flüssigkeiten, Kob-Andersen-Mischungen mit und ohne attraktive Wechselwirkungen. Unsere Ergebnisse deuten darauf hin, dass die Nukleation von Langstreckendomänen durch attraktive Potentiale bei sinkender Temperatur induziert wird. Dies wurde durch die $k \to 0$-Grenze der Bathia-Thornton-Strukturfaktoren und die isotherme Kompressibilität der Masse bestätigt. Nachdem wir ähnliche Untersuchungen bei höheren Dichten durchgeführt haben, stellten wir ebenfalls fest, dass sich die beiden Systeme hinsichtlich ihrer Struktur nicht voneinander unterscheiden. Andererseits haben wir ein ähnliches Verhalten bei der Domänenbildung für Propan-1-ol-Simulationen für einen Temperaturbereich von $200 < T < 300$ K gefunden, wobei um ~ 220 K eine Änderung der isothermen Kompressibilität auf einen Übergang von einer normalen zu einer anomalen Flüssigkeit hinweist. Der Mechanismus der Domänenbildung ist in diesem Fall etwas anders, da sie durch die Bildung gut definierter Netzwerke von Molekülen entstehen, die durch Wasserstoffbrücken (H-Brücken) verbunden sind. Diese Ergebnisse wurden durch die Berechnung der durchschnittlichen Anzahl von H-Brücken innerhalb der Simulationsbox, der Größe des größten

Clusters und vor allem durch NMR-Experimente verifiziert.

Schließlich konnten wir durch die Entwicklung eines DFT-Ansatzes für die Simulation der adaptiven Hamilton-Auflösung (Hamiltonian Adaptive Resolution Simulation, H-AdResS) zeigen, dass das externe Potential, das für ein konstantes Dichteprofil erforderlich ist, mit dem chemischen Überschusspotential des Systems übereinstimmt. Aufgrund der Eins-zu-Eins-Korrespondenz zwischen Gleichgewichtsdichten und externen Potentialen in der DFT ist es daher möglich, das chemische Überschusspotential durch Vorgabe eines konstanten Dichteprofils zu erhalten. Damit kommen wir zum wichtigsten Punkt dieser Arbeit: der Untersuchung der Statistik der atomistischen Domäne in der (H-AdResS)-Methode. Wir haben eine Slab-Geometrie betrachtet, in der die atomistische Region als System und das Hybrid- und Idealgas als thermisches Bad betrachtet wurden. Dies bedeutet, dass der System-Bad-Kontakt nur in einer Koordination stattfindet, während die beiden anderen ihre PBCs behalten. Diese Überlegungen führen zu der Notwendigkeit, eine großkanonische Simulation zu beschreiben, in der die Finite-Size-Effekte noch vorhanden sind, genauer gesagt die Beiträge der PBCs. Diese Beschreibung wurde in dieser Arbeit vorgeschlagen und validiert, indem alle Quellen von Finite-Size-Effekten in den Kirkwood-Buff-Integralen und somit in den Dichtefluktuationen identifiziert und korrigiert wurden. Wir berechneten die lokalen Dichtefluktuationen innerhalb der atomistischen Region und verglichen sie mit den Ergebnissen unserer Methode zur Berechnung der Kirkwood-Buff-Integrale endlicher Größe. Wir fanden herausdass es nicht ausreicht, ein H-AdResS aufzustellen, um eine korrekte stabile großkanonische Statistik zu gewährleisten. Dennoch führt eine korrekte Beschreibung des großkanonischen Ensembles in thermischen Bädern, d.h. des idealen Gases, zu einer ausgezeichneten Übereinstimmung zwischen der H-AdResS und der Integralgleichung endlicher Größe. Der Weg zur Verbesserung der Beschreibung des Bades besteht in der Einführung eines Algorithmus zum Einfügen/Löschen von Teilchen mit der korrekten Statistik, was möglich ist, da die Einfügungswahrscheinlichkeiten für ein ideales Gas in einem großkanonischen Ensemble analytisch berechnet werden können. Zusammenfassend kann gesagt werden, dass die `Adaptive Resolution with Particle Insertion/Deletion Steps` (AdResS+ PI) als großkanonische Ensemble Simulationsmethode verwendet werden kann.

To my parents

Preamble

1. Introduction

In the last few decades, computer simulations have become crucial to research activity in several fields of physics, particularly in the theory of liquids. The fast development and rapid increase in software and hardware resources make running longer simulations for increasingly larger systems possible. Nevertheless, the length and time scales accessible to most computational methods are several orders of magnitude below the experimental ones, which hinders a direct simulation-experiment connection. It is thus necessary to develop theoretical frameworks and numerical strategies to attain a sound description of the statistical mechanics of small systems and a reasonable extrapolation to the thermodynamic limit. Only in this way the comparison between finite-size computer simulation results and experiments becomes meaningful[1].

This thesis is devoted to developing, tuning and applying numerical methods and analysis tools to study the statistical mechanics of small systems and their extrapolation to the thermodynamic limit in the context of molecular dynamics simulations. On the one hand, I have developed finite-size versions of integral equations in statistical mechanics (e.g. Ornstein-Zernicke, Kirkwood-Buff and two-body entropy) where ensemble, finite-domain and periodic boundary effects are included explicitly. On the other hand, I have used these techniques to demonstrate that the Hamiltonian Adaptive Resolution Method (H-AdResS), together with a particle insertion/deletion algorithm, enables the simulation of liquid systems in the grand canonical ensemble. This result paves the way for flexible, efficient and theoretically sound open boundary molecular dynamics simulations.

This document is branched into three parts, the first one is an introductory build up for the methods, quantities and computations used. The second part is a collection of seven research papers that are the final product of the research leading to this dissertation. And finally, a third part of conclusions and final comments.

This introductory part of the document is divided in four main sections. First, a brief description of the Adaptive Resolution Simulation Method (AdResS) is presented. This summary covers the computation of excess chemical potentials, the particle insertion/deletion algorithm and the connection to density functional theory. The second part includes a discussion of the types of size effects present in standard molecular dynamics simulations and the need for defining finite-size integral equations. A third section includes a summary of the thermodynamics for simple liquids and binary mixtures, *i.e.*, radial distribution functions, structure factors, density fluctuations, Kirkwood-Buff integrals, excess entropy and their corresponding small system corrections. Finally, these three sections provide the context for discussing the fluctuations of the number of particles in the grand canonical ensemble with periodic boundary conditions (PBCs). This section

[1] Note that this situation can be different in other fields of soft matter where, for instance, in some cases polymer chains can be longer than their experimental counterparts.

serves the thesis's main goal, which is to demonstrate that AdResS combined with particle insertion/deletion steps (AdResS + PI) correctly samples the grand canonical ensemble and paves the way for open-boundary (non-)equilibrium molecular dynamics simulations of soft matter systems.

1.1 Adaptive Resolution Simulation Method

The adaptive resolution simulation (AdResS) is a multi-scale molecular dynamics (MD) method. The central idea is that two different representations of a system coexist in different subdomains of the same simulation box, allowing particles to change identity *on-the-fly* from one region to another. Typically, the two representations are fully atomistic (AT), where a detailed description of the system is considered, and a coarse-grained (CG) version with reduced degrees of freedom or simpler interactions. In the transition domain between the AT and CG representations, the hybrid region (HY) is defined as the subdomain in which the identity of the particles changes continuously, parameterized by a switching field $\lambda(\mathbf{R}_\alpha)$ with $\mathbf{R}_\alpha$ the instantaneous position of a molecule α.

As the interactions are well described in both the AT and the CG representations, it is necessary to define how the transition is done in the HY region. Two possibilities have been proposed to tackle this aspect: (a) a force interpolation between the two representations weighted by the correspondent distances (AdResS) [27, 29, 28]; (b) a Hamiltonian based strategy where a global Hamiltonian is defined in such a manner that the interactions are turned on or off, depending on the direction of the transition (H-AdResS) [26, 13, 4].

As this thesis focuses on using the Hamiltonian alternative of the method, the following section will explore and describe such a framework in more detail. For a recent review on the Adaptive Resolution, see ref [7].

1.1.1 Hamiltonian Adaptive Resolution Simulation (H-AdResS) Method

In the (H-AdResS) method, the two different representations of the system coexist in the same simulation box, completely described by a global Hamiltonian. In the following, let us consider the particular coarse-grained representation to coincide with an ideal gas (IG), meaning the interatomic interactions left are those given by elastic collisions. In practice, however, this representation corresponds to thermalized, non-interacting particles. We have a Hamiltonian as follows [26, 15],

$$H_{[\lambda]}(\mathbf{r},\mathbf{p}) = \mathcal{K} + V^{\text{intra}} + \sum_{\alpha=1}^{N}\{\lambda_\alpha V_\alpha + V^{\text{ext}}(\lambda_\alpha)\}\,, \tag{1.1}$$

with $(\mathbf{r},\mathbf{p})$ the phase-space coordinates. In this Hamiltonian, we can identify,

- $\mathcal{K} = \sum_{i=1}^{N_a} \mathbf{p}_i^2/2m_i$ is the total kinetic energy of the system.
- In the potential term, the Latin indices run over atoms while Greek indices over molecules.
- The term $V^{\text{intra}} = \sum_{\alpha=1}^{N}\sum_{i\neq j\in\alpha} V^{\text{intra}}(r_{ij})$ accounts for intra-molecular interactions, with r_{ij} the distance between atoms i and j belonging to the same molecule α.

§ The intermolecular interactions are included in the term $V_\alpha = \frac{1}{2}\sum_{\beta\neq\alpha}\sum_{i\neq j} V(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|)$ with $\mathbf{r}_{\alpha i}$ the position of the atom i in molecule α.

§ The molecules' identity is determined by the switching field $\lambda_\alpha \equiv \lambda(\mathbf{R}_\alpha)$ where $\mathbf{R}_\alpha$ is the position of the centre of mass of the molecule α[2].[3]

§ V^{ext} is a generic external potential. In this context, we added an explicit dependence of λ to get a more general situation, as both representations undergo different interaction schemes.

To understand better the effect attributed to each term, let us consider the case where $V^{\text{ext}} = 0$ and from the Hamiltonian (1.1), the total force acting on a particular atom i of a molecule α goes as

$$\mathbf{F}_{\alpha i} = \mathbf{F}^{\text{intra}}_{\alpha i} + \sum_{\beta,\beta\neq\alpha}\left\{\frac{\lambda_\alpha+\lambda_\beta}{2}\mathbf{F}^{\text{AT}}_{\alpha i|\beta} + \left(1-\frac{\lambda_\alpha+\lambda_\beta}{2}\right)\mathbf{F}^{\text{IG}}_{\alpha i|\beta}\right\} - \left[V^{\text{AT}}_\alpha - V^{\text{IG}}_\alpha\right]\nabla_{\alpha i}\lambda_\alpha, \tag{1.2}$$

where it is clear that we used the indexes to highlight the coordinates of the specific atom or molecule involved. The term of the intra-molecular interactions, as well as the summation, are those expected from the description of the method, as the interpolation of the interactions happens as a result of the change of resolution. Nevertheless, the last term corresponds to an artificial force dragging particles from the AT to the IG region. Note that the addition of this drifft force violates both Newton's third law and momentum conservation. As the two representations exhibit different interactions, the pressures are different between the two subdomains, creating an imbalance of density governed by such a force. This imbalance has to be compensated in such a way that the density is constant along the whole simulation box. As the AT and IG regions are described by characteristic interactions and the thermodynamics are well defined, the place where a free-energy compensation has to be added is in the HY region, where the dragging force is acting as $\nabla\lambda \neq 0$ [14].

The correction is straightforwardly added by including to the Hamiltonian an extra term compensating the drag force applied on the HY region as

$$H_\Delta = H - \sum_{\alpha=1}^{N}\Delta H(\lambda_\alpha), \tag{1.3}$$

where the extra term has to satisfy that, on average, the drift force is cancelled out and linear momentum is conserved,

$$\left.\frac{d\Delta H(\lambda)}{d\lambda}\right|^{\text{Drift}}_{\lambda=\lambda_\alpha} = \left\langle\left[V^{\text{AT}}_\alpha - V^{\text{IG}}_\alpha\right]\right\rangle_\alpha \equiv \mathcal{V}(\lambda_\alpha) \tag{1.4}$$

Finally, besides the drag force, the effects of the difference in pressure have to be corrected in order to get the correct density profiles. To achieve that, a so-called *thermodynamic force* has to be introduced. This force is obtained after an iterative procedure, updating forces at `n+1` step as,

[2]This can also be done by choosing a representative atom of the molecule and keep track of the position of such an atom as $\mathbf{R}_\alpha$

[3]When $\lambda = 0$ the Hamiltonian describes an homogeneous ideal gas system provided $V^{\text{ext}}(0) = \text{constant}$. For the case of $0 < \lambda \leq 1$, the interactions start taking place, and the Hamiltonian describes an inhomogeneous system in the presence of an external field. In particular, when $\lambda = 1$, the system is completely atomistic.

$$\mathbf{F}^{\text{Thermo}}_{\mathtt{n+1}} = \mathbf{F}^{\text{Thermo}}_{\mathtt{n}} + \frac{c\nabla\rho_{\mathtt{n}}(\mathbf{r})}{\rho^*}, \tag{1.5}$$

where c is a parameter that modulates the strength of the correction in energy units, while ρ^* is a reference density. This updating procedure of the thermodynamic force converges to a uniform density profile throughout the simulation box, making $\nabla\rho = 0$.

These two effects can be corrected via the external potential V^{ext} and from here the importance of including it. With these two corrections, the system is free of any artificial effect due to the change of identity along the direction of interest. To investigate the physical meaning of the corrections included to the Hamiltonian, we can see how the Hamiltonian changed, as the total correction to the Hamiltonian has to obey

$$\left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda_\alpha} = -\mathcal{V}(\lambda_\alpha)\nabla\lambda_\alpha + \mathbf{F}^{\text{Thermo}}_{\alpha}. \tag{1.6}$$

Or, after integrating,

$$\Delta H = -\int_{\mathbf{r}^{\text{AT+HY}}}^{\mathbf{r}^{\text{AT}}} d\mathbf{R}[-\mathcal{V}(\lambda) + \mathbf{F}^{\text{Thermo}}(\mathbf{R}))]. \tag{1.7}$$

From the equation (1.7), it is possible to compute the free energy compensation necessary to flatten the density profile. This compensation coincides with the differences in Gibbs free energy, which in the case of the Grand Canonical ensemble, is the chemical potential,

$$\Delta H(\lambda) \equiv \frac{\Delta G}{N} = \Delta\mu \tag{1.8}$$

where $\Delta G/N$ is the molar Gibbs free energy. Note that this difference means the difference between the AT and IG regions, not the difference of Gibbs energy between two points states of a part of the whole system.

In summary, this means that by simply applying the equilibration of a standard H-AdResS the resulting force can be integrated so we get *for free* the excess chemical potential of the atomistic system. In contrast to traditional methods, the H-AdResS perform better under high-density/concentration conditions, as the populations of the density profiles needed to compute the thermodynamic force can be generated with higher resolution.

Particle insertion

As stated before, considering the CG representation to be an ideal gas (IG), allows to compute the excess chemical potential of the liquid in the AT region, suggesting a μVT, or Grand Canonical, ensemble. Still, a system is said to be in the grand canonical ensemble if it is in thermal and chemical equilibrium with an infinite reservoir of particles and energy, which is not the case in computer simulations. In practice, considering a system of a fixed number of particles N_0 and volume V_0 at a temperature T, a subdomain of volume V and average number of particles $\langle N\rangle$ can be said to be in the grand canonical ensemble if the subvolume V is of the order of 1% of the total volume V_0[21] which in practical terms is, if not impossible, unreachable. In order to address this issue, it is possible to take advantage of the geometrical and statistical construction of the H-AdResS to include a particle insertion/deletion protocol directly on the IG. The procedure here described follows closely Ref [15].

Consider the IG region to be in the grand canonical ensemble. This means that thermodynamic quantities fixed are (μ, V_0, T). The probability distribution of such a system having exactly N particles, follows a Poisson distribution [19]

$$P(N) = \frac{(N^*)^N e^{-N^*}}{N!}, \tag{1.9}$$

where N^* is the mean number of particles inside the volume V_0. In particular, an ideal gas N^* can be written as a function of the chemical potential as,

$$N^* = \frac{V_0 e^{\beta\mu}}{\lambda^3} \tag{1.10}$$

with $\beta = \frac{1}{k_B T}$ and λ being the mean thermal wavelength. Considering the limit $N, N^* \gg 1$ with $|(N - N^*)/N^*| \ll 1$, the Poisson distribution can be approximated to a normal distribution with mean N^* and standard deviation $\sqrt{N^*}$ as,

$$P(N) = \frac{e^{-(N-N^*)^2/(2N)}}{\sqrt{2\pi N^*}}. \tag{1.11}$$

As we are interested in the fluctuations in densities, this result can be rewritten to be a function of densities instead of the number of particles, getting

$$P(\rho) \sim e^{-k_\mu(\rho-\rho^*)^2/2}, \tag{1.12}$$

where we used that $k_\mu = V_0/\rho^*$. In practical terms, this parameter will be used to reach a target density ρ^*.

Having a well-defined probability distribution for the ideal gas region, a Metropolis algorithm can be introduced for the insertion/deletion of particles. The probability of accepting a move that the density ρ changes by an amount ν is given by

$$\text{acc}(\rho \to \rho + \nu) = \min\left[1, \exp\left(-k_\mu \nu(\nu + 2(\rho - \rho^*))\right)\right] \tag{1.13}$$

and correspondingly,

$$\text{acc}(\rho \to \rho - \nu) = \min\left[1, \exp\left(-k_\mu \nu(\nu - 2(\rho - \rho^*))\right)\right]. \tag{1.14}$$

The parameter μ is taken from the normal distribution

$$P(\nu) = \sqrt{\frac{k_\mu}{2\pi}} e^{-\frac{k_\mu \nu}{2}}. \tag{1.15}$$

Note that in the acceptance probabilities, we do not have to explicitly include the chemical potential, as is the case with Grand Canonical Monte Carlo methods. This occurs because we do not need to fix the chemical potential but simply fix the density to a desired target density ρ^*.

In a nutshell, the method guarantees the correct fluctuations on the number of particles of the IG given by the grand canonical ensemble, allowing an exchange of particles at constant chemical potential between a reservoir of ideal gas particles with the AT region.

A schematic representation of the simulation's geometry is presented in Fig. 1.1, where the AT region is connected to the IG, where the particle insertion algorithm is implemented. In all

the cases studied here, a slab geometry, oriented along the x-direction, was taken for the change of resolution, while the other two present periodic boundary conditions (PBC). These considerations are going to exhibit specific size effects when analyzing the finite size fluctuations of the number of particles. More on this particularities in section 1.7

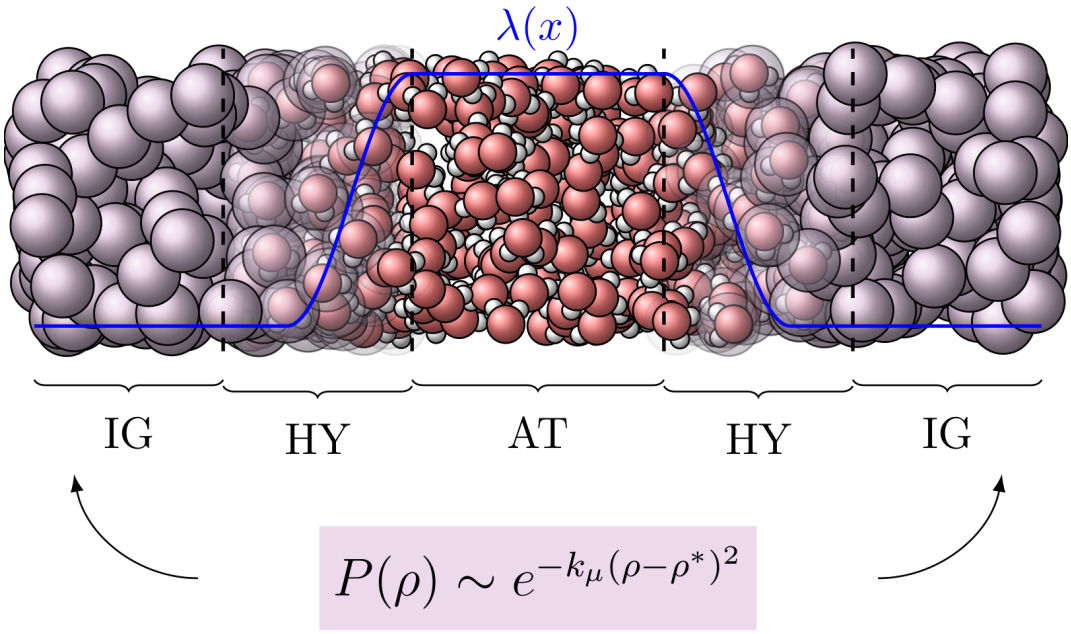

Figure 1.1: Schematic representation of the H-AdResS method with the particle insertion. A slab geometry allows including not one but two ideal gas regions where the PI method is applied in order to enforce the constant temperature-chemical potential transfer of particles along the simulation box.

DFT approach to H-AdResS

Before finishing this section of the introductory chapter, let us work on the further development of the understanding of the H-AdResS Hamiltonian. As we saw before, including the thermodynamic force is necessary in order to eliminate any kind of spurious effect introduced by the switching of interactions, but it is still unclear what the effects are on the statistical mechanics. This section is a part of the paper included in this thesis in chapter I, where it was proven that this is a well-defined Hamiltonian suited to be used for standard statistical mechanics provided the Grand Canonical ensemble is used. This last part is the main contribution of this thesis, where methods and simulations were developed to show that the H-AdResS actually fulfils all the requirements to bring the AT region to the grand canonical ensemble. Those tools and analyses will be provided throughout the whole document, but in particular, in the last article included.

Considering the system with a Hamiltonian (1.1) being in the Grand Canonical ensemble, we define the corresponding grand canonical partition function

$$\mathcal{Z}[\lambda] = \mathrm{Tr}\left\{\exp\left(-\beta(H_{[\lambda]} - \mu(\lambda)N)\right)\right\}, \tag{1.16}$$

with again, $\beta = 1/k_B T$ and k_B being the Boltzmann's constant. Using the partition function, in

order to define the thermal average, we take the equilibrium distribution function as

$$f_0[\lambda] = \frac{1}{\mathcal{Z}[\lambda]} \exp\left(-\beta(H_{[\lambda]} - \mu(\lambda)N)\right), \tag{1.17}$$

in such a way that $\mathrm{Tr}\, f_0[\lambda] = 1$ and the thermal average of a generic operator $\hat{O}$ is given by

$$\left\langle \hat{O} \right\rangle^{[\lambda]} = \mathrm{Tr}\left\{ f_0[\lambda]\hat{O} \right\}. \tag{1.18}$$

So, taking the definition of thermal average to write the equilibrium density profile in the form

$$\rho_0^{[\lambda]}(\mathbf{r}) = \langle \hat{\rho}(\mathbf{r}) \rangle^{[\lambda]}, \tag{1.19}$$

with $\hat{\rho}(\mathbf{r}) = \sum_\alpha \delta(\mathbf{R}_\alpha - \mathbf{r})$.
It has been proposed a functional form for the grand potential using an arbitrary distribution function $f[\lambda]$, satisfying the condition $\mathrm{Tr} f[\lambda] = 1$, that written for our conditions

$$\Omega_{[\lambda]}[f[\lambda]] = \mathrm{Tr}\{f[\lambda](H_{[\lambda]} - \mu(\lambda)N + \beta^{-1} \ln f[\lambda])\}. \tag{1.20}$$

In particular, by replacing the equilibrium distribution, we get

$$\Omega_{[\lambda]}[f_0[\lambda]] = -\beta^{-1} \ln \mathcal{Z}[\lambda] = \Omega^0_{[\lambda]} \tag{1.21}$$

the definition of grand potential in terms of the grand canonical partition function.
Using the density operator $\hat{\rho}$, we can rewrite the external potential as

$$\sum_\alpha^N V^{\mathrm{ext}}(\lambda_\alpha) = \int d\mathbf{r} \sum_{\alpha=1}^N \delta(\mathbf{R}_\alpha - \mathbf{r}) V^{\mathrm{ext}}(\lambda(\mathbf{r})) = \int d\mathbf{r}\, \hat{\rho}(\mathbf{r}) V^{\mathrm{ext}}(\lambda(\mathbf{r})). \tag{1.22}$$

The density distribution $f[\lambda]$ is, in general, a functional of the density $\rho^{[\lambda]}(\mathbf{r})$. Then, by using the Hamiltonian (1.1) we write the grand potential as a functional of the density

$$\begin{aligned} \Omega_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})] &= \mathrm{Tr}\{f[\lambda](H_{[\lambda]} - \mu(\lambda)N + \beta^{-1} \ln f[\lambda])\} \\ = \mathrm{Tr}\left\{ f[\lambda] \left(\mathcal{K} + V^{\mathrm{int}} + \textstyle\sum_{\alpha=1}^N \lambda_\alpha V_\alpha + \beta^{-1} \ln f[\lambda] + \sum_{\alpha=1}^N V^{\mathrm{ext}}(\lambda_\alpha) - \mu(\lambda)N \right) \right\} \\ = F_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})] + \textstyle\int d\mathbf{r}\, \rho^{[\lambda]}(\mathbf{r}) \left(V^{\mathrm{ext}}(\lambda(\mathbf{r})) - \mu(\lambda(\mathbf{r})) \right), \end{aligned} \tag{1.23}$$

in which we use Eq. (I.7), where the total external potential is written as a functional of the density. The functional $F_{[\lambda]}[\rho^{[\lambda]}]$ is the intrinsic Helmholtz free energy corresponding to the Hamiltonian (1.1), which is independent of the external potential:

$$F_{[\lambda]}[\rho^{[\lambda]}] = \mathrm{Tr}\left\{ f[\lambda] \left(\mathcal{K} + V^{\mathrm{int}} + \sum_{\alpha=1}^N \lambda_\alpha V_\alpha + \beta^{-1} \ln f[\lambda] \right) \right\}. \tag{1.24}$$

The last two equations provide us with a direct connection between H-AdResS and density functional theory. The expression (I.8) is a functional Legendre transform relating the Helmholtz free energy and the grand potential [2].

To understand Eq. (I.8), we underline that the system is inhomogeneous, and the grand potential is a functional of the external potential [30, 2]. That is

$$\frac{\delta\Omega_{[\lambda]}}{\delta(\mu(\lambda(\mathbf{r})) - V^{\text{ext}}(\lambda(\mathbf{r})))} = -\rho^{[\lambda]}(\mathbf{r})\,, \tag{1.25}$$

which is consistent with the thermodynamic identity $\partial\Omega/\partial\mu = -N$.
Written as a functional of the density, $\Omega_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})]$ represents the cost in free energy necessary to find the system at precisely the density $\rho^{[\lambda]}(\mathbf{r})$ [2]. In particular, we find the density field, $\rho_0^{[\lambda]}(\mathbf{r})$, that minimises this cost by evaluating the functional derivative of the grand potential with respect to the density. Hence, the grand potential satisfies

$$\left.\frac{\delta\Omega_{[\lambda]}[\rho^{[\lambda]}]}{\delta\rho^{[\lambda]}}\right|_{\rho^{[\lambda]}=\rho_0^{[\lambda]}} = 0\,, \tag{1.26}$$

thus implying

$$\left.\frac{\delta F_{[\lambda]}[\rho^{[\lambda]}]}{\delta\rho^{[\lambda]}}\right|_{\rho^{[\lambda]}=\rho_0^{[\lambda]}} = -V^{\text{ext}}(\lambda(\mathbf{r})) + \mu(\lambda(\mathbf{r}))\,. \tag{1.27}$$

The exact form of the intrinsic free energy functional $F_{[\lambda]}[\rho^{[\lambda]}]$ is, in general, unknown. However, we can re-write it as

$$F_{[\lambda]}[\rho^{[\lambda]}] = F^{\text{exc}}_{[\lambda]}[\rho^{[\lambda]}] + F_{[0]}[\rho^{[0]}]\,, \tag{1.28}$$

with $F^{\text{exc}}_{[\lambda]}[\rho^{[\lambda]}]$ the excess free energy, calculated with respect to the free energy of the reference system (ideal gas), $F_{[0]}[\rho^{[0]}]$. To ensure thermodynamic consistency in the adaptive resolution description, we require the free energy to be independent of λ [13]. Nevertheless, we anticipate here that this condition could be changed to induce non-equilibrium conditions in the system [15]. In equilibrium, $F^{\text{exc}}_{[\lambda]}[\rho^{[\lambda]}] = 0$, thus,

$$F_{[0]}[\rho^{[0]}(\mathbf{r})] = F^{\text{id}}[\rho(\mathbf{r})] = \beta^{-1}\int d\mathbf{r}\rho(\mathbf{r})\{\ln(\lambda_T^3\rho(\mathbf{r})) - 1\}\,, \tag{1.29}$$

with F^{id} the Helmholtz free energy of the ideal gas (our reference state at $\lambda = 0$), $\rho^{[0]}(\mathbf{r}) = \rho(\mathbf{r})$ and $\lambda_T = (\hbar^2\beta/2\pi m)^{1/2}$ the thermal, de Broglie, wavelength. We rewrite Eq. (I.12) as

$$\frac{\delta F^{\text{id}}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} = -V^{\text{ext}}(\lambda(\mathbf{r})) + \mu^{\text{exc}}(\lambda(\mathbf{r})) + \mu^{\text{id}}\,, \tag{1.30}$$

with $\mu^{\text{exc}}(\lambda(\mathbf{r})) = \mu(\lambda(\mathbf{r})) - \mu^{\text{id}}$, a continuous function of λ, $\mu^{\text{id}} = \beta^{-1}\ln(\lambda_{\text{dB}}^3\rho_0)$ and ρ_0 the reference (ideal gas) density. By replacing μ^{id} into the previous equation, we find

$$\rho(\mathbf{r}) = \rho_0 \exp\left(-\beta\{V^{\text{ext}}(\lambda(\mathbf{r})) - \mu^{\text{exc}}(\lambda(\mathbf{r}))\}\right), \tag{1.31}$$

where it is apparent that this density field does not explicitly depend on the switching field. Indeed, this expression shows that the external potential determines the density. In particular, the condition

$$V^{\text{ext}}(\lambda(\mathbf{r})) = \mu^{\text{exc}}(\lambda(\mathbf{r})) \tag{1.32}$$

guarantees a constant density ρ_0. In other words, the external potential that enforces the reference density for the whole system is equal to the excess chemical potential.
This result provides an interesting parallel with density functional theory. In the latter, we apply an external potential and then use a self-consistent approach to find the corresponding system's equilibrium density. By considering the one-to-one correspondence between external potential and equilibrium density as provided by DFT, the approach generally used in `H-AdResS` naturally follows. We enforce a uniform density throughout the simulation box and then calculate the resulting external potential $V^{\text{ext}}(\lambda(\mathbf{r}))$. This is particularly useful to compute chemical potentials of molecular fluids and liquid mixtures since $V^{\text{ext}}(1) = \mu^{\text{exc}}$ with μ^{exc} the excess chemical potential of the atomistic system over the ideal gas.

In the following sections, we will devote our efforts to developing methods to compute thermodynamics for small systems and, in particular, to bring those results to the thermodynamic limit where all the side effects are corrected. In this respect, the next section will focus on the possible size effects present in computer simulations.

1.2 Finite-size effects

In standard MD simulations, the systems considered have a fixed number of particles N_0, and depending on the integration scheme, different ensembles can be achieved. By default, in solving Newton's equations (or equivalently Hamilton's equations) of motion, the energy E is conserved[4] as well as the volume V_0 where usually the system is taken to be connected to itself in a torus-like geometry; these kind of boundaries are called periodic boundary conditions (PBC). In this context, those considerations show that there are situations where the simulations might differ from the theoretical and experimental counterparts in two aspects. First, the ensemble considered in most of the vast theoretical microscopical descriptions of liquids, as well as the experimental conditions, are different to the ones described by simulations. Second, the addition of PBCs will add correlations artificially given by the positions of the copies of every particle. These two problems can be enclosed in what we call *finite-size effects*. The size effects given by the different ensembles considered are refereed to be *Explicit* size effects [34, 32], while those contributions given by the PBCs are called *Implicit* size effects[33].

In this thesis, we used and developed tools to identify and correct the finite-size effects present in computer simulations. This enables us to extrapolate the results of, typically, finite-size simulations to the thermodynamic limit. In the rest of this section, we present a brief discussion of the sources and consequences of implicit and explicit size effects.

1.2.1 Explicit Size-Effects

Besides simulations in the microcanonical ensemble, MD simulations can be performed straightforwardly by adding thermostats or barostats to fix the temperature or pressure by changing the integration schemes. Both options require an update of velocities, as the temperature depends on the kinetic energy of the particles, or volume after averaging over time to get to a target value. For

[4]From here the need of symplectic numerical integrators not to include changes in the energy produced purely from the discrete integration of the correspondent equations of motion

instance, the temperature can be fixed by integrating using a Langevin equation of motion. This is by introducing random white noise and a friction force that, via the Fluctuation-Dissipation theorem, enforces a certain temperature on average. A similar situation can happen with the pressure that can be obtained by a virial expansion in terms of ensemble averages.

In the case of a Grand Canonical (GC) ensemble, where most of the theory has been developed, the situation changes, making it more challenging as an infinite bath of both particles and energy is required to get the correct statistics described by fixing the chemical potential μ. More on this particular case will be discussed in the following sections, as it is of particular interest for us to identify how the GC statistics can be recovered to sample correctly the ensemble.

Nevertheless, when the systems are big enough, all these effects get minimized, and all results converge to the *thermodynamic limit* as the sizes of both bath and system grow as $V_0, N_0 \to \infty$ keeping a constant ratio $N_0/V_0 < \infty$. Hence, in order to be able to compare our simulations with theoretical and experimental results, it is necessary to find a way to measure and extrapolate the MD simulations to the thermodynamic limit.

1.2.2 Implicit Size-Effects

The implicit size effects arise as a consequence of the inclusion of periodic boundary conditions, where a virtual copy of the particles results from replicating the simulation box in all directions. This creates a limitation on the maximum size that can be used for the computation of thermodynamic quantities, as taking into account two or more copies should be avoided. Therefore, up to $L_0/2$ linear sizes can be taken, with L_0 the linear size of the box[5].

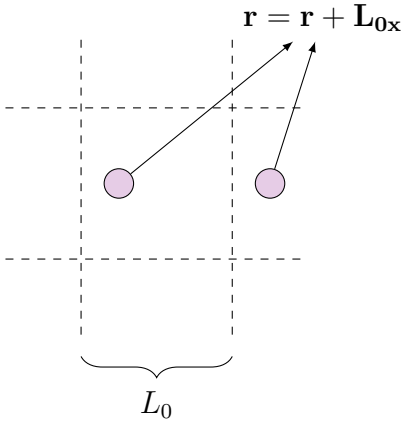

Figure 1.2: Representation of the correlations given by the PBCs

In principle, it is expected that including PBCs should lead to the system approaching bulk conditions, so no border effects arise. This means that the system represents an infinite system with artificial correlations given by the copies. These correlations can actually be measured and corrected, as we will see in the following sections.

[5]This occurs by taking a cubic box.

1.3 Integral Equations and thermodynamics

Thermodynamics is a set of emergent properties that result from a collective of particles interacting among themselves. In many cases, the interactions can be reduced simply to pair-like interactions, making it possible to study effects on the statistics of populations of particles up to the second order.

In this section, we will focus on the explanation of the necessary concepts and definitions to compute thermodynamics in the context of computer simulations focusing on quantities that account for the correlations up to the second order.

1.3.1 Radial Distribution Function

The radial distribution function (RDF) plays a key role in the thermodynamics of molecular systems [9, 11]. This occurs because, in practice, it has the information of the local structure of a system up to pair correlations. It allows us to distinguish the structure of systems depending on the positions, shapes and intensities of the characteristic peaks. Such peaks represent the coordination shells around a reference particle by measuring the average density at a certain distance as represented in Fig. 1.3.

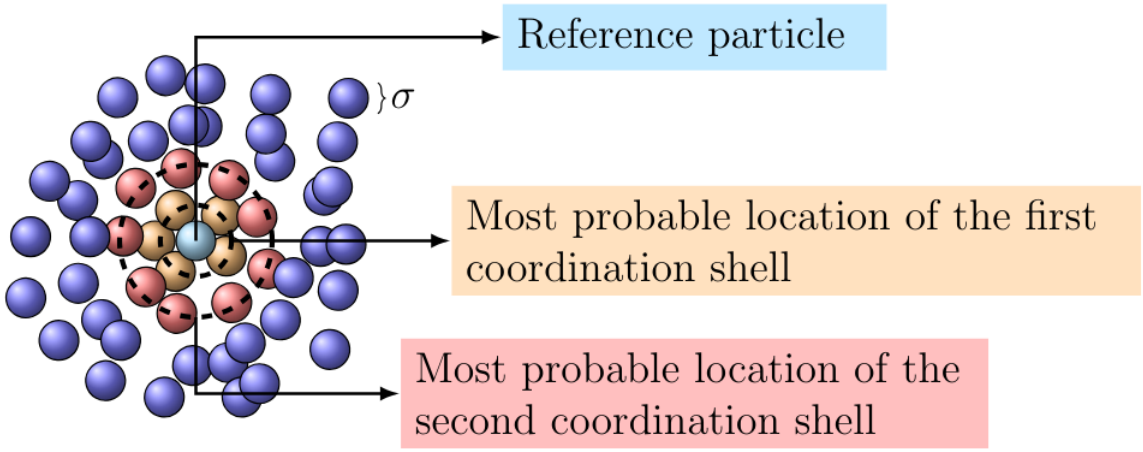

Figure 1.3: Schematic representation of the pair correlation function. The colored particles represent the reference (light blue) first coordination shell (orange), second coordination shell (red) and the rest were not highlighted (dark blue) but they keep having information about the liquid's structure.

To define the RDF in mathematical terms, let us start with a single-component system of N particles. We can take the density distribution as δ-functions located in the positions of the particles as

$$\rho_N^{(1)}(\mathbf{r}) = \left\langle \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}^i) \right\rangle . \tag{1.33}$$

Note that the superscript (1) denotes that this is the first moment of a more complex density distribution function. The definition of the average density presented here is not the only one, but this particular one is useful in the context of computer simulations, among other applications. The

particle distribution function measures how much the structure of a fluid deviates from uniform randomness, which corresponds to an ideal gas in the limit of large distances.

In the same spirit, we can consider the case of a second-order density distribution, where now the question is how likely it is to find a particle in a position $\mathbf{r}^i$ given that another reference particle is already in $\mathbf{r}^j$

$$\rho_N^{(2)}(\mathbf{r}) = \left\langle \sum_{i=1}^{N} \sum_{j=i+1}^{N}{}' \delta(\mathbf{r} - \mathbf{r}^i)\delta(\mathbf{r} - \mathbf{r}^j) \right\rangle . \tag{1.34}$$

This density pair correlation function can be rewritten and normalized to the ideal gas limit, defining the Radial Distribution Function (RDF) as

$$g_{\alpha\beta}(\mathbf{r}, \mathbf{r}') = \frac{1}{\rho_\alpha \rho_\beta} \left\langle \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_\alpha^i) \sum_{j=1}^{M} \delta(\mathbf{r}' - \mathbf{r}_\beta^j) \right\rangle . \tag{1.35}$$

This function can be used to make a direct comparison between computer simulations and experiments as it can be measured via scattering experiments. Computationally, this quantity also presents several challenges, as the information is limited to the size of the simulation box, as well as the time simulated in order to build the averages. A discretization of the space has to be done, and the number of particles in a shell between r and $r + dr$ has to be averaged. In this sense, the RDF is built as a histogram, making the resolution achievable limited.

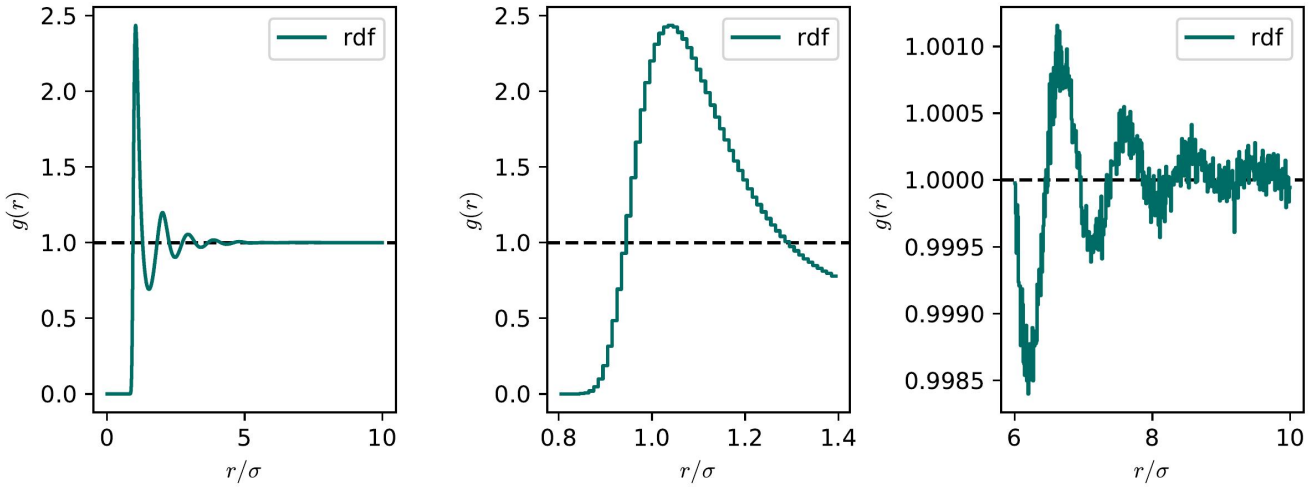

Figure 1.4: Radial Distribution Function for a Lennard-Jones system. (left) RDF computed from a simulation (middle) Discretization of the RDF in steps of dr (right) size effects in the tail.

In Fig. 1.4, three regions of the same plot are presented. First, on the left-hand side, an RDF computed from a simulation of a prototypical Lennard-Jonnes fluid in order to illustrate the limitations. Second, in the middle, the first peak is zoomed in, where now it becomes clear that the construction of the RDF was done by creating a histogram over different shells, and lastly, the right-hand side shows the tail of the RDF. The RDF approaches 1 as $r \to \infty$, but the statistics available for longer values of r drop dramatically, making it impossible to get an RDF without artificial effects given by simply the finite nature of the simulations. From here, all quantities that

depend on the RDF might inherit such effects if the contributions of the tail are relevant hence finite-size corrections are needed.

1.3.2 Structure Factor

The structure factor is closely related to the RDF, as it can be computed as the Fourier transform of the difference of $g(r)$ with respect to the ideal gas case.

$$S_{\alpha\beta}(\mathbf{k}) - \delta_{\alpha\beta} = \rho \int_V d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} (g_{\alpha\beta}(\mathbf{r}) - 1). \tag{1.36}$$

The advantage of considering the structure factor instead of the RDF is that it is possible to establish connections between the structure of the liquid and long-range effects as the focus of the scales changes. Nevertheless, it is important to mention that the structure factor written in Eq. (1.36) has the limitations given by $g(\mathbf{r})$ in addition to those given by the discrete Fourier transform, as the information can be achieved in steps of $k_{\min} = 2\pi/L_0$ where the Fourier space is well defined.

If a spherical symmetry is considered, in addition to taking the liquid to be homogeneous, the structure factor can be integrated in spherical coordinates as [35],

$$S_{\alpha\beta}(k) - \delta_{\alpha\beta} = 4\pi\rho \int dr \frac{r \sin(rk)}{k} (g_{\alpha\beta}(r) - 1), \tag{1.37}$$

which is a common practice in the literature. Still, Eqs. (1.36) and (1.37) present big disadvantages especially at low values of k, hence, an alternative definition can be used because it presents a clear path to numeric implementations since the structure factor can be deduced from scattering theory as [18, 36, 11],

$$S_{\alpha\beta}(\mathbf{k}) = \frac{1}{N} \left\langle \sum_{i\in\alpha}^{N_\alpha} \sum_{j\in\beta}^{N_\beta} \exp\left(-i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)\right) \right\rangle. \tag{1.38}$$

More importantly, this expression can be rewritten into a sum of purely real components that is numerically more stable, once again, primarily at low values of k [25],

$$S_{\alpha\beta}(\mathbf{k}) = \left\langle \frac{1}{N} \left[\sum_{i\in\alpha}^{N_\alpha} \sum_{j\in\beta}^{N_\beta} \sin\left(\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)\right) \right]^2 \right\rangle + \left\langle \frac{1}{N} \left[\sum_{i\in\alpha}^{N_\alpha} \sum_{j\in\beta}^{N_\beta} \cos\left(\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)\right) \right]^2 \right\rangle. \tag{1.39}$$

It is worth mentioning here that another difference between the Eqs. (1.39) and (1.37) is that in the latter, we do not consider the vector $\mathbf{k}$ but simply the norm k. This also means that we will be able to sample many more points on $S(k)$ than those accessible with the RDF. The reason for such an affirmation is basically because we will have a grid distanced $dk = 2\pi/L_0$.

Fig. 1.5 shows a 2D representation of the Fourier space. In this case, we can see in blue those points with length k an integer multiple of dk, but we also have available points at different distances given by the diagonals. This will increase the possible points where we can compute the structure factor compared to the one given by the Fourier transform of the RDF.

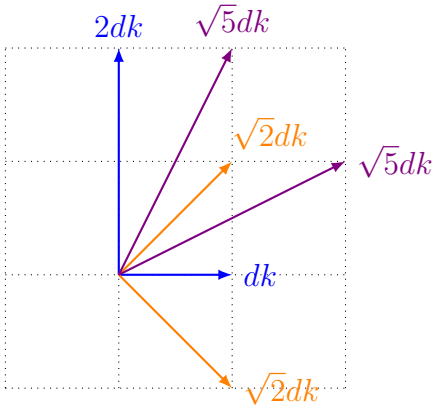

Figure 1.5: Representation of the Fourier space and **k** vectors. Note that now the possible values of k sample increased as the spacing is not regular when taking the norm of **k**

The structure factor is a great tool for the understanding of molecular liquids. First, it can be measured by scattering experiments and can also be used to know how the system works at large scales, for instance, when large clusters are being created or when the system is going through a critical point as the shape of the peaks and the limit $k \to 0$ changes drastically.

1.3.3 limit $S(k \to 0)$

After the discussion about the structure factor, the question of how to approach the limit of $k \to 0$ still remains as the discrete Fourier transform does not reach such a limit, or in practice, we do not have infinite systems in computer simulations. In order to complete or complement the results obtained from Eq. (1.39), the limit has to be computed in a different fashion. To achieve this, we focus our attention on the connection between the structure factor and the isothermal compressibility, which, for a single component system, reads [11],

$$S(k \to 0) = \rho k_B T \kappa_T, \tag{1.40}$$

with k_B being the Boltzmann's constant, T the temperature and ρ the density of the liquid. Therefore, the whole range structure factor can be computed provided the bulk isothermal compressibility κ_T can be extrapolated from finite-size simulations. This connection is obtained via the compressibility equation and extends naturally for mixtures via the Kirkwood-Buff integrals, as will be discussed in the following section.

1.3.4 Size-Effects: Notation

In the sections to come, we will focus on the specific contributions of each source of size effects, hence, it is worth considering a certain notation to identify where such effects are present when writing certain equations down. To do so, we will use a color code as follows

§ Finite Integration Domains Computer simulations are limited to the sizes of the systems we can consider. Even further, to avoid artificial correlations given by the PBCs, the maximum distances reachable are typically smaller than the linear simulation box sizes.

§ Periodic Boundary Conditions In the previous point, the PBCs play already a role in the limitations. In this case, we want to emphasize and make explicit the role of the correlations the PBCs include.

§ Ensemble Effects Lastly, we will show how the differences on ensemble between the definitions an simulations, affect the results we compute.

1.4 Kirkwood-Buff Integrals

The Kirkwood-Buff integrals are of great interest as they connect the microscopic structure with the macroscopic thermodynamics of a liquid [17]. They are integral equations, defined in the GC ensemble, of the RDF. It is also possible to relate them with the total fluctuations of the number of particles, being both approaches theoretically equivalent [23],

$$G_{\alpha\beta} = V\left(\frac{\langle N_\alpha N_\beta\rangle - \langle N_\alpha\rangle\langle N_\beta\rangle}{\langle N_\alpha\rangle\langle N_\beta\rangle} - \frac{\delta_{\alpha\beta}}{\langle N_\alpha\rangle}\right) = \frac{1}{V}\iint d\mathbf{r}\; d\mathbf{r}'\left[g^{\mathrm{o}}_{\alpha\beta}(\mathbf{r},\mathbf{r}') - 1\right], \tag{1.41}$$

where N_α is the number of particles of the α-species and the bracket $\langle\cdots\rangle$ denotes a thermal average, $\delta_{\alpha\beta}$ is the Kronecker delta and $g^{\mathrm{o}}_{\alpha\beta}$ is the pair correlation function defined in the CG ensemble with $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. Note that we refer to the GC ensemble as an open system.

These integrals are full of information regarding the thermodynamics of the system, yet given the limitations of computer simulations, explicit and implicit size effects might appear. In order to visualize this, let us consider a liquid to be spherically isotropic and homogeneous, the KBI can be reduced taking the right-hand-side of Eq. (1.41). The double integral simplifies to a single integral that performed in spherical coordinates goes as,

$$G^{\mathrm{Single}\,\infty}_{\alpha\beta} = 4\pi\int_0^\infty dr\left[g^{\mathrm{o}}_{\alpha\beta}(r) - 1\right]r^2. \tag{1.42}$$

Nevertheless, this expression presents some fundamental problems[6]. Firstly, it is not possible to integrate such a quantity up to ∞ but maximum to $L_0/2$ in order to prevent the PBCs from playing a role as discussed before. Secondly, it is then necessary to know the *open system's* RDF. Hence, a single integral coming from computer simulations should look like [6]

$$G^{\mathrm{Single}}_{\alpha\beta} = 4\pi\int_0^{L_0/2} dr\left[g^{\mathrm{c}}_{\alpha\beta}(r) - 1\right]r^2, \tag{1.43}$$

where now $g^{\mathrm{c}}_{\alpha\beta}(r)$ represents the closed system RDF. These subtle differences will play important roles in the computation of thermodynamics.

To illustrate this point, in Fig. 1.6, a set of different size simulations of the same system are presented. A prototypical Lennard Jones liquid at fixed temperature and density $T = 2.0\,\epsilon/kB$ and $\rho = 0.864\,\sigma^{-3}$ respectively. All simulations were performed in cube boxes such that the volume was determined by the linear size presented in the label.

[6] Note that in this case we highlighted the places where the finite-size effects will take place to ease the transition, but in principle, they are not present in here.

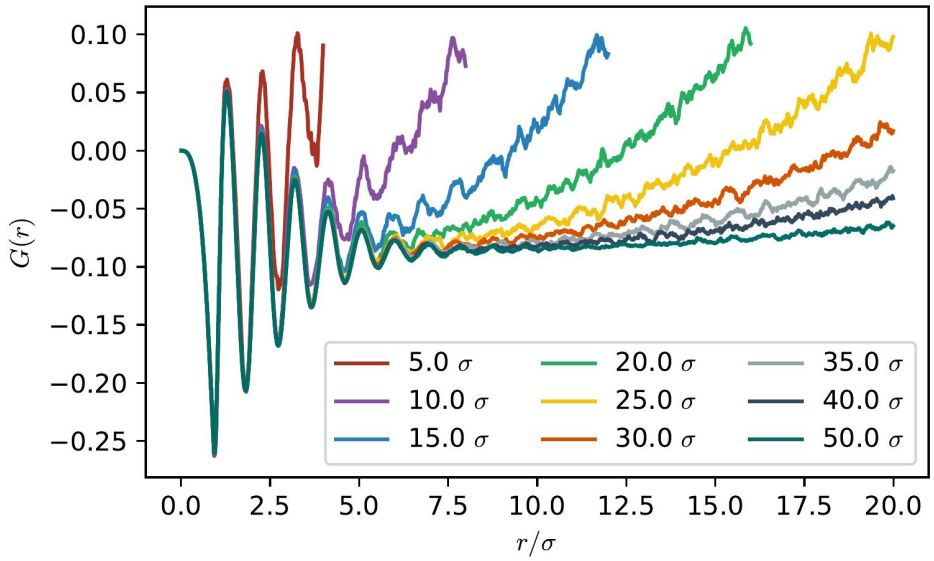

Figure 1.6: The truncated integral G^{Single} for different Lennard-Jones system sizes at the same density ($\rho = 0.864$) and temperature ($T = 2$) in reduced units. The legend shows the size of the linear size L_0 for cubic simulation boxes. Note that the integrals here presented go beyond half of the linear size of the simulation box. These RDFs were computed using the statistics in the cornes of the simulation box, making it possible to extend our results.

In this case, the integral is not converging to a constant value, mainly because $G(r)$ inherits the size effects from $g(r)$. The contribution given to the tail of the RDF comes from the limit $r \to \infty$ [6, 16, 35, 34, 20, 39, 31]

$$g^{\text{c}}(r \to \infty) = g^{\text{o}}(r \to \infty) - \frac{\text{const}}{N_0}, \tag{1.44}$$

where in the case of $N_0 \gg 1$, the contribution of the difference of ensemble decreases as it is clear from the Fig. 1.6 as the single integral goes flatter for larger simulation boxes. From here, it would be possible to estimate how big the system has to be in order to reduce the artificial effects as the oscillations in Fig. 1.6 coincide with those of the RDF. It is necessary to guarantee that those correlations are suppressed and N_0 is large enough that it makes the integration volume term vanish or that it is feasible to neglect it.

Taking all these considerations into account, it is worth defining a finite size version of the Eq. (1.41) and, from it, reach the thermodynamic limit. To do so, we define the finite-size KBI as

$$\begin{aligned} G_{\alpha\beta}(V; V_0) =& V \left(\frac{\langle N_\alpha N_\beta \rangle' - \langle N_\alpha \rangle' \langle N_\beta \rangle'}{\langle N_\alpha \rangle' \langle N_\beta \rangle'} - \frac{\delta_{\alpha\beta}}{\langle N_\alpha \rangle'} \right) \\ =& \frac{1}{V} \iint d\mathbf{r}_1 d\mathbf{r}_2 \; R(\mathbf{r}_1) \; R(\mathbf{r}_2) \; [g_{\alpha\beta}(\mathbf{r}; V_0) - 1] \end{aligned} \tag{1.45}$$

where the average $\langle \cdots \rangle' \equiv \langle \cdots \rangle_{V,V_0}$ now explicitly depends on the subdomain and total volumes, V and V_0, respectively, and we made explicit the dependence of the closed system RDF

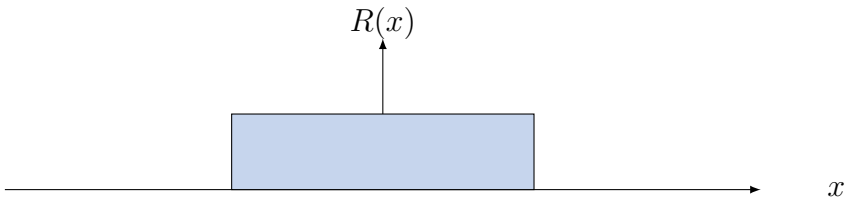

Figure 1.7: One-dimensional box function $R(x)$ used to change the integration limits over the whole space.

with the volume V_0 as $g^{c}_{\alpha\beta}(\mathbf{r}) \equiv g_{\alpha\beta}(\mathbf{r};\, V_0\,) \equiv g_{\alpha\beta}(\mathbf{r};\, N_0\,)$. The function $R(r)$ is a box function delimiting the integration region to inside the simulation box as shown in Fig. 1.7

The following two sections intend the computation of Eq. (1.45). First, by using the Spatial Block Analysis (SBA) method to compute the local fluctuation on the number of particles, and from them, the extrapolation to the thermodynamic limit and therefore to the equation of the isothermal compressibility, or in other words the limit $S(k \to 0)$. Second, using these results, we show how the second half of the finite-size version of the KBI (1.45) can be used not as a method to compute $G_{\alpha\beta}$ in the TL limit, but to study the nature and consequence of each one of the size effects. More importantly, to modify, artificially, the size and shape of the simulation box and to correct each size effect individually from the others.

1.4.1 Spatial Block Analysis

The Spatial Block Analysis (SBA) method has been proven to work to extrapolate the thermodynamic quantities from finite-size computer simulations to the thermodynamic limit in an efficient manner, this result and the rest of this section, follow closely the theoretical developent in ref. [16]. Using the correction to the RDF for closed systems (1.44) suggested in [21, 6]

$$g^{c}_{\alpha\beta}(r \to \infty) = g^{o}_{\alpha\beta}(r \to \infty) - \frac{1}{V_0}\left(\frac{\delta_{\alpha\beta}}{\rho_\alpha} + G^{\infty}_{\alpha\beta}\right). \tag{1.46}$$

And replacing (1.46) into the integral equation (1.41), the integral can be split into the individual contributions of each term, in such a way that up to terms of order $\mathcal{O}(V_0^{-4/3})$

$$\nu G_{\alpha\beta}(\lambda) = \nu G^{\infty}_{\alpha\beta}(1-\nu^3) - \nu^4 \frac{\delta_{\alpha\beta}}{\rho_\alpha} + \frac{\alpha_{\alpha\beta}}{L_0} \tag{1.47}$$

where $\nu \equiv L/L_0$ and α_{ij} is an intensive parameter in length units.

If we take the limit $\nu \ll 1$ (or equivalently $L_0 \to \infty$), the equation (1.47) reduces to a linear function

$$G_{\alpha\beta}(\nu) = G^{\infty}_{\alpha\beta} + \frac{\alpha_{\alpha\beta}}{L} \tag{1.48}$$

which is the case for the KBI in the GC limit. Now, with this expression to fit the results given by the finite size local fluctuations, we can focus on the fluctuation equation to compute the KBI and their corresponding limit to the TL,

$$G_{\alpha\beta}(V;V_0) = V\left(\frac{\langle N_\alpha N_\beta\rangle' - \langle N_\alpha\rangle'\langle N_\beta\rangle'}{\langle N_\alpha\rangle'\langle N_\beta\rangle'} - \frac{\delta_{\alpha\beta}}{\langle N_\alpha\rangle'}\right). \tag{1.49}$$

By measuring the local density fluctuations for different sizes of subvolumes (ν), averaging over the same simulation box and over time as shown in Fig. 1.8, we build Eq. (1.49) as a function of ν as represented in Fig. 1.8.

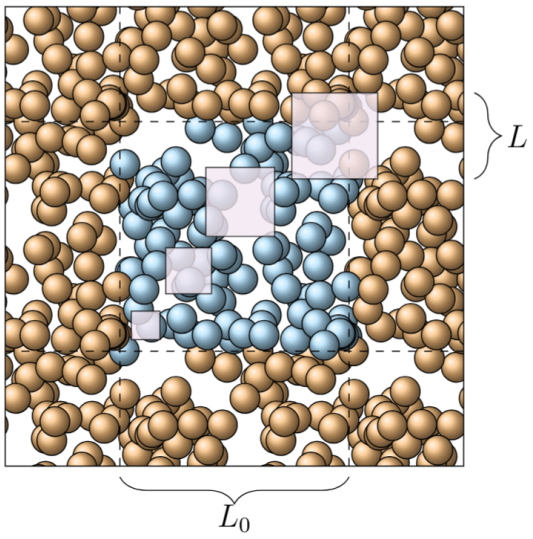

Figure 1.8: Schematic representation of the spatial block analysis method. The N_0 blue particles represent the system with linear size L_0, and the red particles represent the periodic images. The purple box is a subvolume of linear size $L < L_0$ defined to compute fluctuations of the number of particles.

The Eqs. (1.47) and (1.48) can be easily reduced to a single component case. In this case, using the Ornstein-Zernicke equation for finite systems [34], the normalized finite-size isothermal compressibility

$$\chi_T(\nu) = \frac{\Delta^2(N;V,V_0)}{\langle N\rangle'} = \frac{\langle N^2\rangle' - \langle N\rangle'^2}{\langle N\rangle'}. \tag{1.50}$$

The equivalent to Eq. (1.47) and (1.48) read,

$$\begin{aligned}\nu\chi_T^{\mathrm{SBA}}(\nu) &= \nu\chi_T^{\infty}(1-\nu^3) + \frac{\alpha}{L_0} \\ &\approx \nu\chi_T^{\infty} + \frac{\alpha}{L_0} \ \text{(for the case } \nu \text{ small)}.\end{aligned} \tag{1.51}$$

To present a typical example of how this method works, we present here a simulation of a Lennard Jones liquid at $T = 2k_B/\epsilon$, $\rho = 0.864\sigma^{-3}$ and a box size of $L_0 = 10\sigma$, from it we took 1000 frames after 10^8 steps with a $dt = 0.001\tau$ and a Langevin thermostat.

First, we computed the fluctuations of the number of particles and plotted the normalized finite-size isothermal compressibility from Eq. (1.50). Then, from it, we took the regime $0.1 < \nu < 0.3$ and used the linear version of (1.51) to fit the parameters α and χ_T^{∞}. Finally, to evaluate how

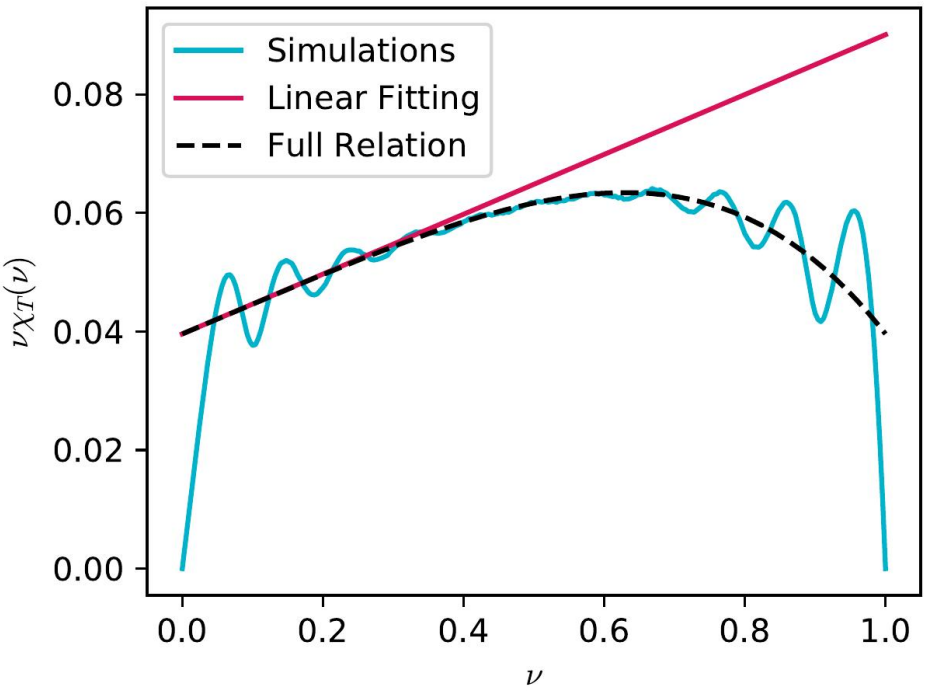

Figure 1.9: Normalized finite-size compressibility χ_T from fluctuations of the number of particles for an LJ system. The straight line represents the linear fitting, while the dashed back line was done using the same parameters but including the ensemble size effect contribution

well these parameters reproduce the profile for the whole range of ν, we plotted in black using the obtained parameters, the first line of (1.51).

This plot presents several points to be highlighted:

§ The oscillations at low values of ν coincide in position and shape with those of the RDF. This comes from the fact that both quantities are actually measuring similar characteristics. On the one hand, the RDF is taking into account the local structure by averaging the number of particles at a certain distance range from a reference particle, and on the other hand, we measure how the fluctuation of the number of particles is characterised by a certain size.

§ The bending comes as a result of the difference in ensemble. This is from the fact that from Eq. (1.49), the linear part comes from the GC ensemble limit, while non-linear term ν^3 is the correction given by the ensemble considerations.

§ The oscillations at ν values approaching 1 present two important points. The first one is that exactly at $\nu = 1$, the normalized compressibility vanishes ($\nu\chi_T|_{\nu=1} = 0$) as in such a limit, the number of particles is fixed, and therefore there are no fluctuations. The second point to discuss is about the similarities of the fluctuations with respect to those at the beginning. The reason for this is actually that these peaks come as a result of the periodic boundary conditions. This will be shown in more detail with explicit theoretical expressions in the next section.

§ If the fluctuations coincide with the non-linear function (1.49), it is a sign that the simulation has reached a correct equilibration and that the phase space point is far from any phase transition.

Summarizing, using this approach, the correct limit of the Kirkwood-Buff integrals $G_{\alpha\beta}$ can be computed completely free of size effects.

1.5 Integral Equations for finite systems

In the previous section, we focused on the computation of the Kirkwood-Buff integral (1.41) from the fluctuations point of view, as the considerations done to reduce the double integral to a single integral are not enough to get a satisfactory finite-size effect free result.

In this section, we will show how the double integration can be computed, not as a method to find the value of the TL Kirkwood-Buff, but as a method to explore the contributions given by the size effects independently instead. An advantage of this approach is that it is possible to build the fluctuations profile for different sizes and shapes independently, on top of the addition or removal of the contributions of both implicit and explicit size effects. Moreover, the Kirkwood-Buff integrals are not the only thermodynamic quantities of the same nature, a double integral over functions of the RDF. In this context, we will propose and validate an expression for the finite-size two-body excess-entropy s_2 in the next section.

1.5.1 Finite-Size Kirkwood-Buff Integrals

First, let us start from the integral version of the KBI (1.41)

$$G_{\alpha\beta}(V;V_0) = \frac{1}{V}\iint d\mathbf{r}_1 d\mathbf{r}_2 \; R(\mathbf{r}_1) \; R(\mathbf{r}_2) \; [g_{\alpha\beta}(\mathbf{r};\, V_0\,) - 1] \tag{1.52}$$

The first correction needed is the one given by (1.46), so that defining $h_{\alpha\beta}(\mathbf{r};\, V_0\,) = g_{\alpha\beta}(\mathbf{r};\, V_0\,) - 1$, the ensemble corrected version [6]

$$h_{\alpha\beta}(\mathbf{r};\, V_0\,) = h_{\alpha\beta}(\mathbf{r}) - \frac{1}{V_0}\left(\frac{\delta_{\alpha\beta}}{\rho_\alpha} + G^{\infty}_{\alpha\beta}\right), \tag{1.53}$$

so that, the integral (1.52) becomes

$$G_{\alpha\beta}(V;V_0) = \frac{1}{V}\iint d\mathbf{r}_1 d\mathbf{r}_2 \; R(\mathbf{r}_1) \; R(\mathbf{r}_2) \; h_{\alpha\beta}(\mathbf{r}) - \frac{1}{V_0}\left(\frac{\delta_{\alpha\beta}}{\rho_\alpha} + G^{\infty}_{\alpha\beta}\right) \tag{1.54}$$

Hence, using the fact that limiting the integration domain with the functions $R(\mathbf{r})$, it is possible to transform to the Fourier space, where the two integrals reduce to a single integral in the full reciprocal space,

$$G_{\alpha\beta}(V;V_0) = \frac{1}{2\pi}\int d\mathbf{k} \; \tilde{R}(\mathbf{k}) \; \tilde{R}(-\mathbf{k}) \; \tilde{h}_{\alpha\beta}(\mathbf{k}) - \frac{1}{V_0}\left(\frac{\delta_{\alpha\beta}}{\rho_\alpha} + G^{\infty}_{\alpha\beta}\right) \tag{1.55}$$

where $\tilde{f}_{\alpha\beta}(\mathbf{k})$ is the Fourier transform of a function $f_{\alpha\beta}(\mathbf{r})$.

Besides the fact that the integral is now reduced to a single space integral, a very important feature of the Fourier space is that adding the contributions given by the PBCs is natural by simply adding translations of the quantity of interest [33],

$$\tilde{h}^{\mathrm{PBC}}_{\alpha\beta}(\mathbf{k}) = \sum_{n_x,n_y,n_z} e^{-k\cdot\mathbf{s}_{n_x,n_y,n_z}} \tilde{h}_{\alpha\beta}(\mathbf{k}) \tag{1.56}$$

with $\mathbf{s}_{n_x,n_y,n_z} = (n_x L_{0x}, n_y L_{0y}, n_z L_{0z})$ a vector specifying the system's periodic images characterized by the integers n_{xyz}. Note that from here, adding the copies at different distances will result in different shapes of the simulation box in a completely artificial fashion. Hence, the total expression for the computation of the finite-size KBI in the Fourier space reads

$$G_{\alpha\beta}(V;V_0) = \frac{1}{2\pi}\int d\mathbf{k}\, \tilde{R}(\mathbf{k})\, \tilde{R}(-\mathbf{k})\, \tilde{h}^{\mathrm{PBC}}_{\alpha\beta}(\mathbf{k}) - \frac{1}{V_0}\left(\frac{\delta_{\alpha\beta}}{\rho_\alpha} + G^{\infty}_{\alpha\beta}\right). \tag{1.57}$$

Coming back to previous sections where we discussed the partial structure factors and their relationship with the RDF, or better with $h_{ij}(\mathbf{r})$ [2, 12], we recall

$$S_{\alpha\beta}(\mathbf{k}) = \delta_{\alpha\beta} + \tilde{h}_{\alpha\beta}(\mathbf{k}). \tag{1.58}$$

This relation can be efficiently computed from simulations using the expression (1.39) where the limit of $k \to 0$ is not yet obtained but is still needed for the computation of the integral. Nevertheless, this point can actually be computed with the SBA method as we discussed before using the KBI (1.49) and from (1.40), we have

$$\lim_{k\to 0} S_{\alpha\beta}(k) = \delta_{\alpha\beta} + \rho_\alpha G^{\infty}_{\alpha\beta}. \tag{1.59}$$

These complementing methods allow us to remove all ensemble size effects present in the structure factor or in the RDF itself.

Fig. 1.10 presents many physical insights into the liquid of interest and the size effects present in the simulations. Let us go through the main points to get a better understanding of the plot.

§ The ensemble effects can be included (excluded) from the integration by including (not including) the correction (1.53). The results presented show that the open system, or GC ensemble, goes as a straight line just as we described in the limit of large volumes, or $\nu \ll 1$, in Eqs. (1.47) and (1.48), or their single component version, the compressibility Eq. (1.51). Instead, in the case of the canonical ensemble, there is a bending result of the lack of the exchange of particles with an external bath.

§ Including PBCs adds oscillations in the limit of large subvolumes $\nu \approx 1$. These oscillations are of a similar nature as those of small volumes $\nu \approx 0$ as both correspond with the typical sizes and positions of the oscillations of the RDF.

§ Correcting the limit $S(k \to 0)$ makes the system artificially infinite. This, in practice, allows us to change the size of the system simply by including the copies of the PBCs in different places. More importantly, it is also possible to change the sizes of each direction independently, as the copies are added by translating by a vector in the reciprocate space $\mathbf{s}_{n_x,n_y,n_z} = (n_x L_{0x}, n_y L_{0y}, n_z L_{0z})$.

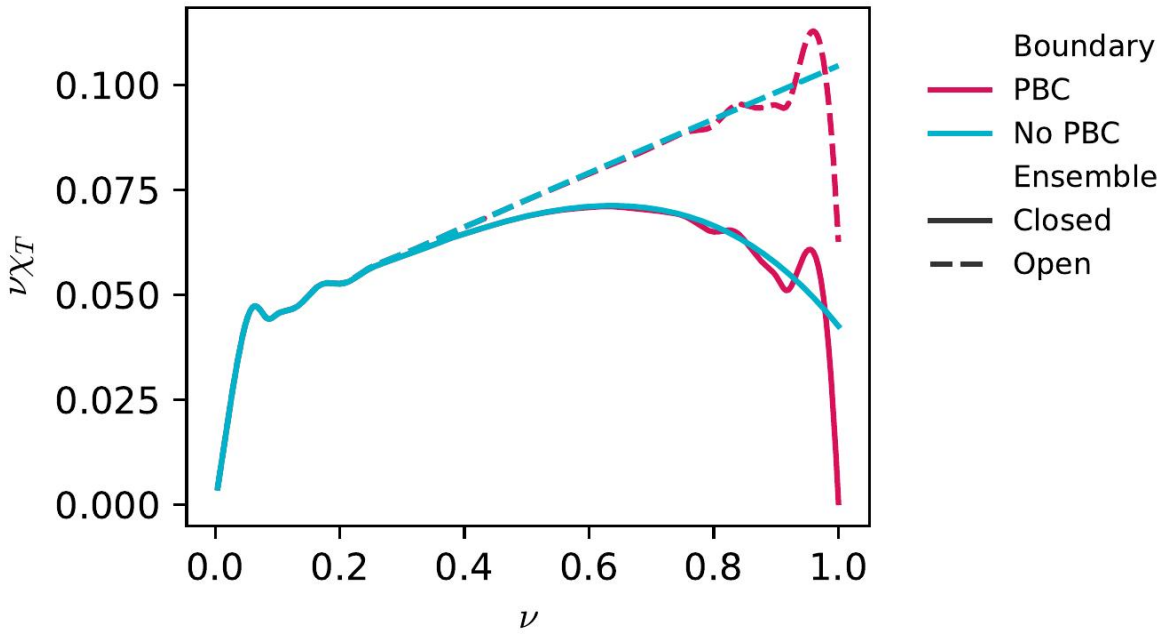

Figure 1.10: Normalized finite-size compressibility χ_T from from the double integral for a LJ system. All contributions are considered, including PBCs and ensemble corrections as well as not including them

In summary, by computing the structure factor of $k > 0$ from the simulation trajectories using Eq. (1.39) and the limit $k \to 0$ from the SBA (1.48) and (1.59), it is possible to compute the whole range of fluctuations of the number of particles, allowing to add or delete the contributions by the different finite-size effects integrating Eq. (1.57).

1.6 Two-Body Finite-Size Excess-Entropy s_2

In the same spirit, we can build other finite-size integral equations and solve them in the Fourier space, identifying the contributions of explicit and implicit size effects.

Continuing the line of correlations up to the second order, in this section, we define a finite-size integral for the calculation of two-body excess entropy as it has been proved to bring most of the contributions to the total entropy of many simple liquids. This quantity has been less studied than the Kirkwood-Buff integrals in the context of the extrapolation to the thermodynamic limit and the convergence for finite sizes. Therefore, we will start from the definition of the entropy and a brief description of the traditional method to compute it to propose not just a finite-size scaling as $1/L$ but also going through the fact that this quantity is free of any kind of ensemble corrections, making the PBCs the only extra contribution we needed to compute it for finite sizes. This section follows closely part of the theoretical part of [37]

First, let us consider the definition of excess entropy for a system with a fixed number of particles N with respect to the ideal gas

$$s_{\text{exc}} = \frac{S - S_{\text{IG}}}{N k_{\text{B}}} = \frac{S_2 + S_3 + \cdots}{N k_{\text{B}}} \,. \tag{1.60}$$

It has been shown that for simple liquids, the contributions to the total excess entropy of the second order terms is around 80–90% of the total value.[11, 7]

In particular, the GC ensemble we have that [45, 42]

$$s_2 = -\frac{\rho}{2V}\int_V\int_V d\mathbf{r}_1\, d\mathbf{r}_2\, \left[g(\mathbf{r})\ln g(\mathbf{r}) - (g(\mathbf{r})-1)\right] , \tag{1.61}$$

where $s_2 = S_2/Nk_\mathrm{B}$ is the two-body excess entropy per particle and $g(\mathbf{r})$ the radial distribution function (RDF) with $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$.

In the same way, as we described before for the KBI, it is possible to reduce this integral to a single integral by considering the system to be homogeneous and isotropic,

$$s_2^\infty = -2\pi\rho\int_0^\infty dr\, r^2\left[g(r)\ln g(r) - (g(r)-1)\right] . \tag{1.62}$$

Similarly, as discussed in Eq. (1.43), the domain of the integral is limited by the size of the simulation box. For such a reason, computationally, the integral can be only computed up to $L_0/2$ and the RDF of a closed system

$$s_2^R = -2\pi\rho\int_0^R dr\, r^2\left[g(r;\,N_0\,)\ln g(r;\,N_0\,) - (g(r;\,N_0\,)-1)\right] . \tag{1.63}$$

Inspired by the previous calculations of the compressibility equation and Kirkwood Buff integrals and starting from (1.62), let us define a finite-size two-body excess-entropy for a subvolume V of a system of a fixed number of particles N_0 and total volume V_0

$$s_2(V;N_0) = -\frac{\rho}{2V}\int_V\int_V d\mathbf{r}_1\, d\mathbf{r}_2\, \left[g(\mathbf{r};\,N_0\,)\ln g(\mathbf{r};\,N_0\,) - (g(\mathbf{r};\,N_0\,)-1)\right] . \tag{1.64}$$

Making a parallel with the finite-size KBI, and in order to study the consequence of the ensemble size effects, we take the correction to the asymptotic limit of the RDF as

$$g(\mathbf{r};\,N_0\,) = g(\mathbf{r}) - \frac{\chi_T^\infty}{N_0} , \tag{1.65}$$

with $\chi_T^\infty = \rho k_\mathrm{B}T\kappa_T$, and κ_T being the bulk isothermal compressibility. We write the integrand in Eq. (1.64) as

$$\begin{aligned} g(\mathbf{r};\,N_0\,)\ln g(\mathbf{r};\,N_0\,) &\approx g(\mathbf{r})\ln g(\mathbf{r}) - \frac{\chi_T^\infty}{N_0}(1+\ln g(\mathbf{r})) , \\ g(\mathbf{r};\,N_0\,) - 1 &= g(\mathbf{r}) - 1 - \frac{\chi_T^\infty}{N_0} , \end{aligned} \tag{1.66}$$

where in the first line in the previous expression, keeping terms up to the order $O\left(1/N_0^2\right)$. The contributions χ_T^∞/N_0 of the two terms of the excess entropy cancel out exactly. The contribution $\chi_T^\infty \ln g(\mathbf{r})/N_0$ can be neglected by assuming a large number of particles (there is no subvolume

V/V_0 contribution, only $1/V_0$. Hence, we can neglect it). This indicates that the two-body excess entropy is ensemble invariant.We thus rewrite Eq. (1.64) as

$$s_2(V) = -\frac{\rho}{2V} \int_V \int_V d\mathbf{r}_1 \; d\mathbf{r}_2 \; [g(\mathbf{r}) \ln g(\mathbf{r}) - (g(\mathbf{r}) - 1)] \; . \tag{1.67}$$

The volume V is finite and embedded into the volume V_0.

Before going to the following section, where we introduce a method to compute $s_2(L)$ for a finite-size system, let us explore the consequences of the ensemble size effects cancelling out perfectly. First, we can make an identification of the terms in the integrand of Eq. (1.67). The second half, $(g(\mathbf{r}) - 1)$ is exactly the same as the Kirkwood-Buff integrals, while the first part resembles the expression of the entropy from information theory taking $g(\mathbf{r})$ as the corresponding probability distribution. Using these labels for the two contributions, and using (1.63), we plotted the two terms independently in Fig. 1.11 for a system Lennard Jones liquid of size $L/\sigma_{\mathrm{LJ}} = 35$ at $k_{\mathrm{B}}T = 2.0\epsilon$. Both integrals diverge similarly for large values of R but in different directions,

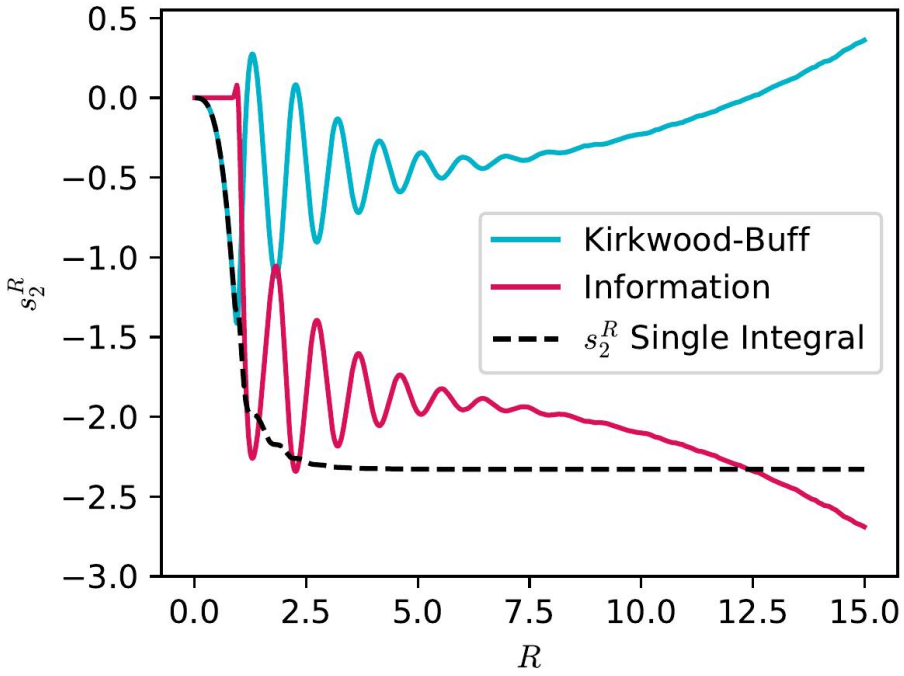

Figure 1.11: Plot of the two contributions, *Kirkwood-Buff* $(g(r; N_0) - 1)$ and *Information* $(g(r; N_0) \ln g(r; N_0))$, to the truncated integral s_2^R for a system of linear size $L/\sigma_{\mathrm{LJ}} = 30$ at $k_{\mathrm{B}}T = 2.0\epsilon$. It is apparent that the two terms oscillate out-of-phase for small values of R, and their sum converges to s_2^∞ when $R \to \infty$. The black plot is the sum of the two terms highlighting the s_2 convergence.

Kirkwood-Buff to infinity and Information to minus infinity, which shows that both terms exhibit strong ensemble finite-size effects. However, these two finite-size contributions balance each other, and the sum of the two integrals converges to s_2^∞ for large values of R as shown in the dashed black line. Due to this error cancellation, the truncation Eq. (1.63) gives a reasonable approximation to s_2^∞, providing a large enough value of R.

1.6.1 Finite-size excess entropy

Continuing with the comparison to the previous section, as the integrals of the excess entropy s_2 and the Kirkwood-Buff integrals G_{ij} for finite systems belong to the same kind, we define a finite-volume two-body excess entropy as

$$s_2(V) = -\frac{\rho}{2V} \int \int d\mathbf{r}_1 \, d\mathbf{r}_2 \; R(\mathbf{r}_1) \; R(\mathbf{r}_2) \; h(\mathbf{r}) \,, \tag{1.68}$$

with $R(\mathbf{r})$ the same step function that defines the finite integration subdomain shown in Fig. 1.7. The new function $h(\mathbf{r})$ (comparing with (1.53)) is defined as

$$h(\mathbf{r}) = g(\mathbf{r}) \ln g(\mathbf{r}) - (g(\mathbf{r}) - 1) \,. \tag{1.69}$$

We write the double integral of $s_2(V)$ in Fourier space and include the periodicity of the simulation of the box in $h(\mathbf{r})$ explicitly. Thus

$$s_2(V) = -\frac{\rho}{2(2\pi)^3 V} \int d\mathbf{k} \; \tilde{R}(\mathbf{k}) \; \tilde{R}(-\mathbf{k}) \; \tilde{h}^{\mathrm{PBC}}(\mathbf{k}) \,, \tag{1.70}$$

where we added the contributions given by the periodic boundary conditions in the same way as before in Eq. (1.56)

$$\tilde{h}^{\mathrm{PBC}}(\mathbf{k}) = \sum_{n_x, n_y, n_z} e^{-\mathbf{k} \cdot \mathbf{s}_{n_x, n_y, n_z}} \, \tilde{h}(\mathbf{k}) \,, \tag{1.71}$$

with $\tilde{h}(\mathbf{k})$ the Fourier transform of $h(\mathbf{r})$ and $\mathbf{s}_{n_x,n_y,n_z} = (n_x \, L_x, n_y \, L_y, n_z \, L_x)$ a vector specifying the system's periodic images such that $n_{x,y,z}$ takes integer values.

Let us finish this section by making some notes on this Fig. 1.12

- § As the two-body excess entropy ended up being an ensemble invariant size effect, there is no need to include such a contribution as the two terms cancel each other out perfectly.
- § The PBCs are actually bringing the system, artificially, to the thermodynamic limit. This is clear as the blue curve goes fast to the reference point given by the single integral.
- § Taking the last point of this plot, we can build a profile of the entropy as a function of the linear size of the box. Here, we also have to mention that this relationship follows a trend like $1/L$, as will be shown explicitly in the following chapters.

This definition of the two-body finite-size excess-entropy allows a clear connection to other size dependent thermodynamic quantities. In the following chapters, we will explore the connection of s_2 with the reduced self diffusion coefficient D^* proposed by Dzugutov[8] and how the size of the simulation box actually affects the values of the parameters in such a relation.

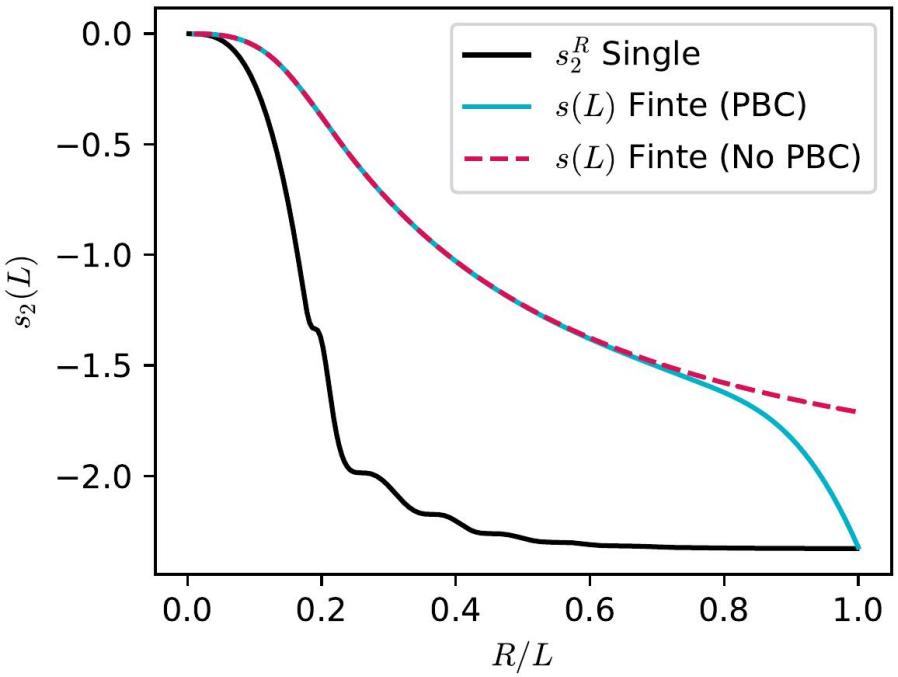

Figure 1.12: The Finite-size integral s_2. In black we present the reference computed with the single integral (1.63) and with the method here proposed: blue, including the contributions given by the PBCs and pink without them.

1.7 Size effects in the Grand Canonical ensemble

In the previous sections, we focused our attention on the correct computation of thermodynamic quantities coming from computer simulations, i.e., RDFs, s_2, $S(k)$ among others. During these calculations, we made emphasis on the nature and consequences of the different kinds of finite-size effects. With these tools, we can now collect results to propose how to test whether or not simulations are performed in the GC ensemble.

In this section, we will propose a setup of MD simulations using the Hamiltonian adaptive resolution simulation method `AdResS` to perform simulations in the GC ensemble. One of the most important factors when claiming the GCE ensemble sampling is the correct fluctuations of the number of particles as this comes as a consequence of having a constant chemical potential. The theoretical calculations as well as the computational method were developed for this thesis in chapter III where the different kind of size effects present in the Kirkwood-Buff integrals for simple liquids, were identified and corrected independently. This work allowed us to have a benchmark for the GC simulations, as we have full control over the statistics and sizes, independently, granting us the opportunity of a fair comparison with the GC with finite size effects in them as will be the case to be studied. This strategy will be finally tested and verified in the last chapter of this thesis VI, where all the conditions needed to claim the correct sampling of the GC ensemble achieved.

1.7.1 Grand Canonical H-AdResS

Recalling from the first section of this chapter, two representations of the system coexist within the same simulation box. Let us start from the setup description of the H-AdResS schematically represented in (1.13). In this case, we have an atomistic representation with PBCs in the y and z directions while a contact with two thermal baths in the x direction

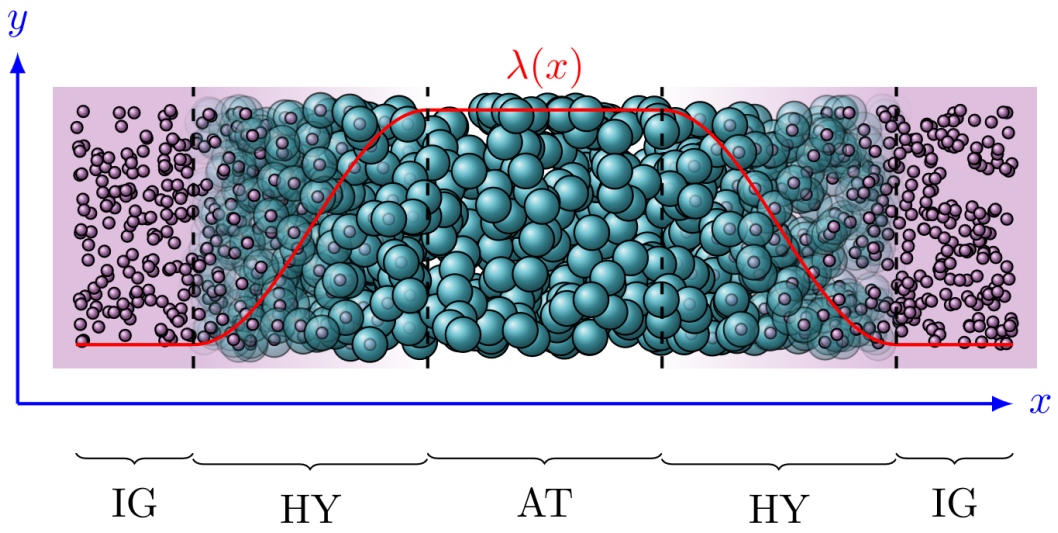

Figure 1.13: Representation of the slab geometry for H-AdResS setup. An atomistic representation in the middle whereas the outside regions are taken as ideal gas representations. in the middle hybrid regions where particles change *on-the-fly* from one representation to the other.

Note that in our case, we will keep the two external baths connected among them via PBCs, but as a reference point for future extensions of the works here presented, the two reservoirs can be completely decoupled opening the door for a correct non-equillibrium description. Focusing our attention on the atomistic representation, we can reduce the described set up to the one presented in Fig. 1.14, where now, countrary to conventional simulations, we will have *open boundaries* in the x direction.

If we take this fraction of the box as the whole system and compute the fluctuations of the number of particles with the SBA method and the Eq. (1.50) as described earlier, a GC simulation should go as exhibited in Fig. 1.10. For simpler visualization, let us replot the fluctuations of the number of particles, for a single component system, as a function of the cubic subvolume of fraction ν[7].

In Fig. 1.15 two cases are presented, both in the GC ensemble. The difference is the presence/absense of PBCs contributions. In our proposal, we are building a setup where the size effects given by the ensemble are reduced but the PBCs ones are still present.

Finally, the only condition we are missing is providing the correct description of the reservoir. In principle, the GC ensemble occurs when a system is set in thermal and chemical equilibrium with an infinite reservoir of particles and energy. Hence to guarantee such conditions, we might work on the statistics of the ideal gas, or the reservoir. The last part of this chapter is devoted to review

[7]Here we refeer to the GC as the previous *Open system*

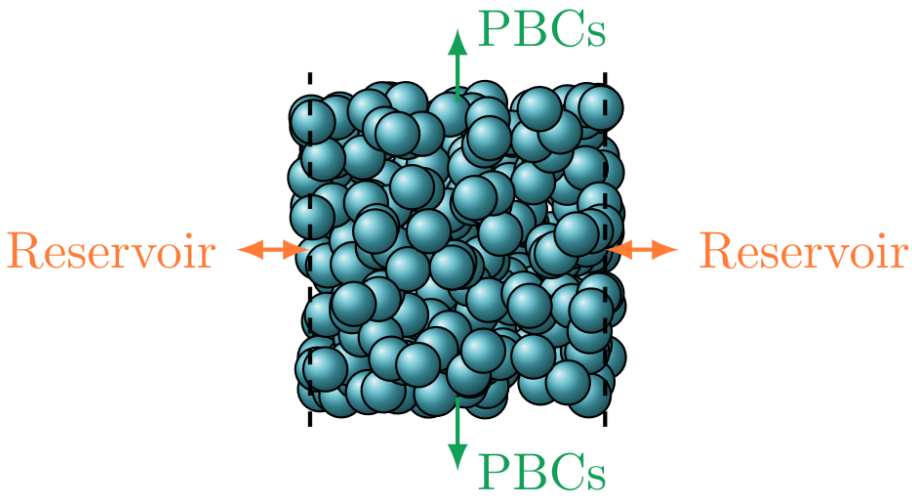

Figure 1.14: Schematic atomistic region within the H-AdResS setup

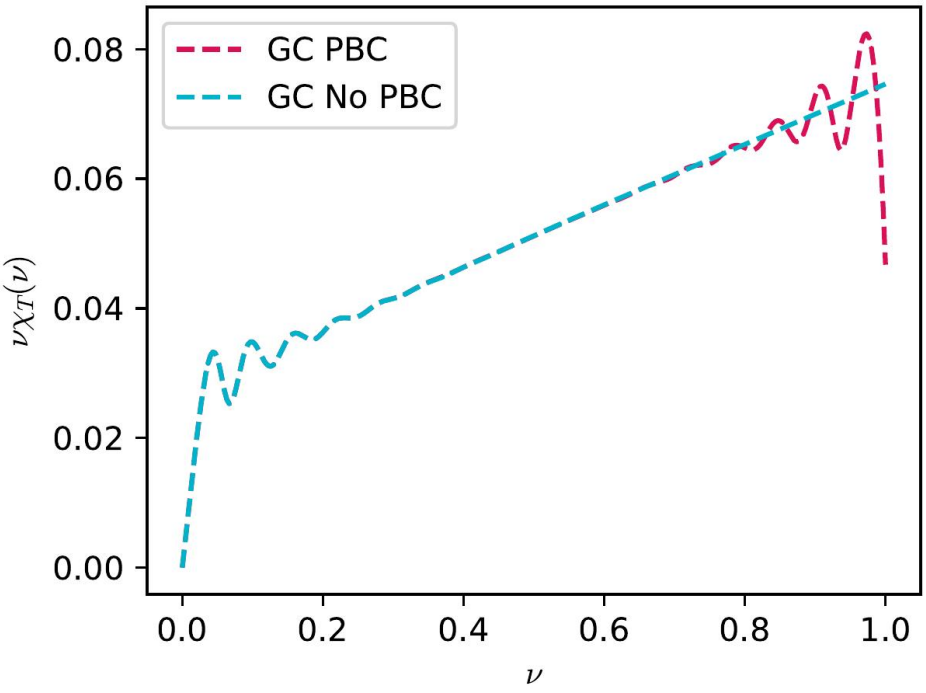

Figure 1.15: Fluctuations of the number of particles for subvolumes of linear length ν for a GC simulation.

the statistics on the reservoir, such that the correct description of the interaction system/reservoir can be assured.

Particle Insertion

In this last section, we will recall some calculations done earlier in order to justify the `Adaptive REsolution with Particle insertion/deletion Steps` (`AdResS + PI`) as a GC simulation method.

The main idea is that we need to enforce the correct statistics on the reservoirs that are in contact to our system. For that, let us study briefly the nature of our reservoirs where a particle insertion algorithm is added as represented in Fig. 1.16

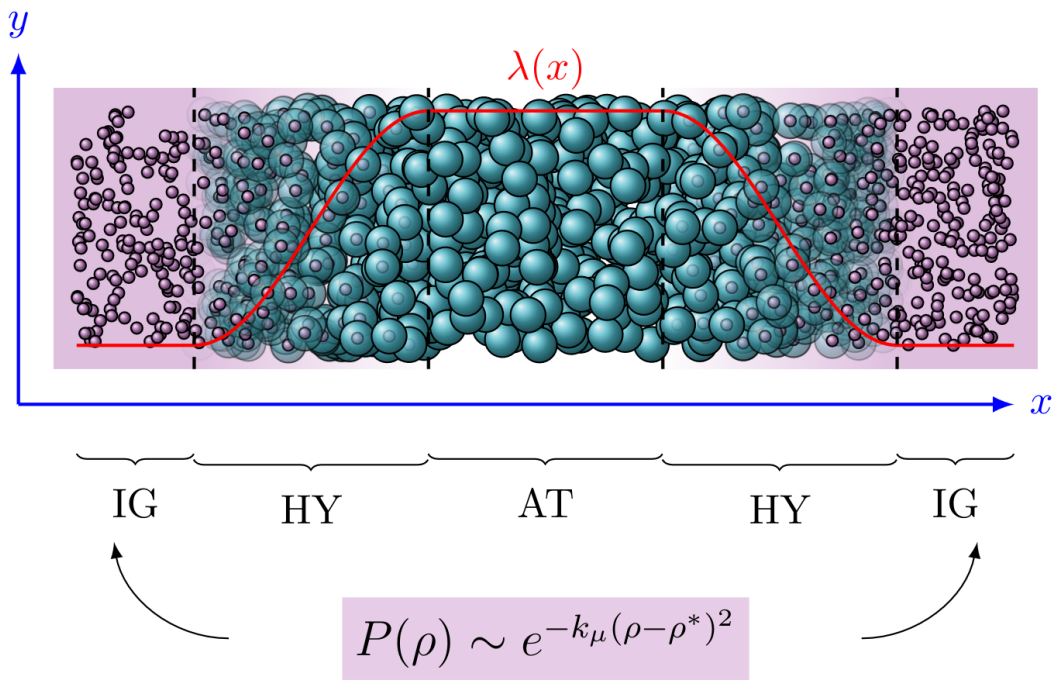

Figure 1.16: Representation of the slab geometry for `H-AdResS` + `PI` setup.

First we will have two different regions, the hybrid (`HY`) where the thermodynamic force reassures the constant chemical potential among the whole simulation box as described in the sec 1.1.1. Second, we have the ideal gas (`HY`) region where the corrections to the reservoirs should be implemented. As reported, again, in ref [15], assuming the ideal gas being in the GC ensemble, and using the fact that we know the analytically the ideal gas statistical mechanics, it leads to a well known probability distribution for the number of particles, or equivalently, density. This transformation simply means that there is an immediate link to an insertion/deletion probabilities in order to match a target density ρ^*.

Let us summarize this section with some comments about these last points

§ In a previous work [38] we established the way to compute the Kirkwood-Buff integrals, and therefore the fluctuations of the number of particles, removing size effects independently among them.

§ This computations allow us to have a benchmark case to two possible scenarios where the GC ensemble is present: With and without PBCs.

§ As in the simulation scheme here proposed, where the GC ensemble is claimed, PBCs are still present in two coordinates, it is necessary to have the description of the number fluctuations in such a particular combination of size effects: PBCs + GC.

§ The way to guarantee the correct nature of the reservoirs, is taking advantage of the well known ideal gas statistics and implementing an insertion/deletion arrangement keeping the chemical potential μ along the simulation box constant.

Bibliography

[1] Nathan Argaman and Guy Makov. Density functional theory: An introduction. *American Journal of Physics*, 68(1):69–79, 2000.

[2] NW Ashcroft and David C Langreth. Structure of binary liquid mixtures. i. *Physical Review*, 156(3):685, 1967.

[3] Atreyee Banerjee, Shiladitya Sengupta, Srikanth Sastry, and Sarika Maitra Bhattacharyya. Role of structure and entropy in determining differences in dynamics for glass formers with different interaction potentials. *Physical review letters*, 113(22):225701, 2014.

[4] L A Baptista, R C Dutta, M Sevilla, M Heidari, R Potestio, K Kremer, and R Cortes-Huerto. Density-functional-theory approach to the hamiltonian adaptive resolution simulation method. *Journal of Physics: Condensed Matter*, 33(18):184003, apr 2021.

[5] István Borzsák and András Baranyai. On the convergence of green's entropy expansion. *Chemical physics*, 165(2-3):227–230, 1992.

[6] R. Cortes-Huerto, K. Kremer, and R. Potestio. Communication: Kirkwood-Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *The Journal of Chemical Physics*, 145(14):141103, 10 2016.

[7] Robinson Cortes-Huerto, Matej Praprotnik, Kurt Kremer, and Luigi Delle Site. From adaptive resolution to molecular dynamics of open systems. *The European Physical Journal B*, 94(9):189, September 2021.

[8] Mikhail Dzugutov. A universal scaling law for atomic diffusion in condensed matter. *Nature*, 381(6578):137–139, May 1996.

[9] P.A. Egelstaff. *An Introduction to the Liquid State*. Oxford science publications. Clarendon Press, 1992.

[10] P. Español, R. Delgado-Buscalioni, R. Everaers, R. Potestio, D. Donadio, and K. Kremer. Statistical mechanics of hamiltonian adaptive resolution simulations. *The Journal of Chemical Physics*, 142(6):064115, 2015.

[11] Jean-Pierre Hansen and Ian Ranald McDonald. *Theory of simple liquids: with applications to soft matter*. Academic press, 2013.

[12] W Härtl, C Segschneider, H Versmold, and P Linse. On the structure factor of liquid-like ordered binary mixtures of colloidal suspensions. *Molecular Physics*, 73(3):541–552, 1991.

[13] Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Spatially resolved thermodynamic integration: An efficient method to compute chemical potentials of dense fluids. *Journal of Chemical Theory and Computation*, 14(7):3409–3417, 2018. PMID: 29874069.

[14] Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Spatially resolved thermodynamic integration: An efficient method to compute chemical potentials of dense fluids. *Journal of Chemical Theory and Computation*, 14(7):3409–3417, 2018. PMID: 29874069.

[15] Maziar Heidari, Kurt Kremer, Ramin Golestanian, Raffaello Potestio, and Robinson Cortes-Huerto. Open-boundary Hamiltonian adaptive resolution. From grand canonical to non-equilibrium molecular dynamics simulations. *The Journal of Chemical Physics*, 152(19):194104, 05 2020.

[16] Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto. Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: Revisiting the spatial block analysis method. *Entropy*, 20(4), 2018.

[17] John G. Kirkwood and Frank P. Buff. The Statistical Mechanical Theory of Solutions. I. *The Journal of Chemical Physics*, 19(6):774–777, 06 1951.

[18] P. Kumari, V. V. S. Pillai, D. Gobbo, P. Ballone, and A. Benedetto. The transition from salt-in-water to water-in-salt nanostructures in water solutions of organic ionic liquids relevant for biological applications. *Phys. Chem. Chem. Phys.*, 23:944–959, 2021.

[19] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Statistical Physics: Volume 5*, volume 5. Elsevier, 2013.

[20] Joel Louis Lebowitz and Jerome Kenneth Percus. Long-range correlations in a closed system with applications to nonuniform fluids. *Physical Review*, 122(6):1675, 1961.

[21] R. Potestio M. Heidari, K. Kremer and R. Cortes-Huerto. Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations. *Molecular Physics*, 116(21-22):3301–3310, 2018.

[22] Raymond D. Mountain and Harold J. Raveché. Entropy and molecular correlation functions in open systems. ii two- and three-body correlations. *The Journal of Chemical Physics*, 55(5):2250–2255, 1971.

[23] Kenneth E. Newman. Kirkwood–buff solution theory: derivation and applications. *Chem. Soc. Rev.*, 23:31–40, 1994.

[24] M. Oettel. Classical density functional theory : Exact density distribution of hard rods between walls in 1 d. 2018.

[25] S. D. Overduin and G. N. Patey. Understanding the structure factor and isothermal compressibility of ambient water in terms of local structural environments. *The Journal of Physical Chemistry B*, 116(39):12014–12020, 2012. PMID: 22963671.

[26] Raffaello Potestio, Sebastian Fritsch, Pep Español, Rafael Delgado-Buscalioni, Kurt Kremer, Ralf Everaers, and Davide Donadio. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.*, 110:108301, Mar 2013.

[27] Matej Praprotnik, Luigi Delle Site, and Kurt Kremer. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *The Journal of Chemical Physics*, 123(22):224106, 12 2005.

[28] Matej Praprotnik, Luigi Delle Site, and Kurt Kremer. Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids. *Phys. Rev. E*, 73:066701, Jun 2006.

[29] Matej Praprotnik, Luigi Delle Site, and Kurt Kremer. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annual Review of Physical Chemistry*, 59(Volume 59, 2008):545–571, 2008.

[30] Harold J. Raveché. Entropy and molecular correlation functions in open systems. i. derivation. *The Journal of Chemical Physics*, 55(5):2242–2250, 1971.

[31] FL Román, A González, JA White, and S Velasco. Fluctuations in the number of particles of the ideal gas: A simple example of explicit finite-size effects. *American Journal of Physics*, 67(12):1149–1151, 1999.

[32] F.L. Román, J.A. White, A. González, and S. Velasco. *Ensemble Effects in Small Systems*, pages 343–381. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[33] F. L. Román, J. A. White, A. González, and S. Velasco. Fluctuations in a small hard-disk system: Implicit finite size effects. *The Journal of Chemical Physics*, 110(20):9821–9824, 05 1999.

[34] F. L. Román, J. A. White, and S. Velasco. Fluctuations in an equilibrium hard-disk fluid: Explicit size effects. *The Journal of Chemical Physics*, 107(12):4635–4641, 09 1997.

[35] 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. *Phys. Rev. E*, 53:2382–2389, Mar 1996.

[36] Felix Sedlmeier, Dominik Horinek, and Roland R Netz. Spatial correlations of density and structural fluctuations in liquid water: A comparative simulation study. *Journal of the American Chemical Society*, 133(5):1391–1398, 2011.

[37] Mauricio Sevilla, Atreyee Banerjee, and Robinson Cortes-Huerto. Finite-size excess-entropy scaling for simple liquids. *The Journal of Chemical Physics*, 158(20):204502, 05 2023.

[38] Mauricio Sevilla and Robinson Cortes-Huerto. Connecting density fluctuations and Kirkwood–Buff integrals for finite-size systems. *The Journal of Chemical Physics*, 156(4):044502, 01 2022.

[39] Dario Villamaina and Emmanuel Trizac. Thinking outside the box: fluctuations and finite size effects. *European Journal of Physics*, 35(3):035011, 2014.

Publications

I. Density-Functional-Theory Approach to the Hamiltonian Adaptive Resolution Simulation Method

Outline

Bibliographic Information

Luis A. Baptista, Ravi C. Dutta, Mauricio Sevilla, Maziar Heidari, Raffaello Potestio, Kurt Kremer, and Robinson Cortes-Huerto. Density-functional-theory approach to the hamiltonian adaptive resolution simulation method. *Journal of Physics: Condensed Matter*, 33(18):184003, apr 2021

Abstract

In the Hamiltonian adaptive resolution simulation method (H-AdResS) it is possible to simulate coexisting atomistic and ideal gas representations of a physical system that belong to different subdomains within the simulation box. The Hamiltonian includes a field that bridges both models by smoothly switching on (off) the intermolecular potential as particles enter (leave) the atomistic region. In practice, external one-body forces are calculated and applied to enforce a reference density throughout the simulation box, and the resulting external potential adds up to the Hamiltonian. This procedure suggests an apparent dependence of the final Hamiltonian on the system's thermodynamic state that challenges the method's statistical mechanics consistency. In this paper, we explicitly include an external potential that depends on the switching function. Hence, we build a grand canonical potential for this inhomogeneous system to find the equivalence between H-AdResS and density functional theory (DFT). We thus verify that the external potential inducing a constant density profile is equal to the system's excess chemical potential. Given DFT's one-to-one correspondence between external potential and equilibrium density, we find that a Hamiltonian description of the system is compatible with the numerical implementation based on enforcing the reference density across the simulation box.
In the second part of the manuscript, we focus on assessing our approach's convergence and computing efficiency concerning various model parameters, including sample size and solute concentrations. To this aim, we compute the excess chemical potential of water, aqueous urea solutions and Lennard-Jones mixtures. The results' convergence and accuracy are convincing in all cases, thus emphasising the method's robustness and capabilities.

I.1 Introduction

In the adaptive resolution method [44, 45, 46, 47, 16] (AdResS) and its Hamiltonian variant (H-AdResS) [41, 40] it is possible to simulate atomistic and ideal gas representations of a physical system coexisting within the simulation box [25, 22]. A hybrid region connects both representations via a field that switches on (off) intermolecular interactions as molecules enter (leave) the atomistic region (Figure I.1). External one-body forces are computed and applied to ensure a constant density profile, and once integrated, the corresponding potential energy adds up to the Hamiltonian. This external potential has been identified with the excess chemical potential between the regions of the system [41, 40, 22]. Nevertheless, this technical procedure raises the question of whether the modified Hamiltonian becomes dependent upon the specific thermodynamic state. In consequence, such a dependence would question the use of the H-AdResS Hamiltonian in the context of classical statistical mechanics.
In this paper, we use the statistical mechanics formalism of H-AdResS, developed by Español and collaborators [13], where the external potential directly enters the Hamiltonian as a functional of the switching field. We use this Hamiltonian to build a grand canonical potential that, for such an inhomogeneous system, is a functional of the external potential [2]. Since the subsystem of interest is in contact with an ideal gas, sampling the grand canonical potential does not involve any major difficulty, as we have recently demonstrated [22]. As a matter of fact, the adaptive resolution framework has been established as a method to perform simulations in the grand canonical ensemble [9, 17, 28, 43, 1, 11, 22]. With this grand potential [39, 30], we find an equivalent classical density functional theory (DFT) approach to H-AdResS. Once the equivalence between

DFT and H-AdResS is established, we show that the external potential that enforces the reference, ideal gas, density throughout the system is precisely the excess chemical potential. Given DFT's one-to-one correspondence between the density and the external potential, we validate the standard H-AdResS strategy based on enforcing a uniform density profile to compute the Hamiltonian's external potential and, therefore, the excess (over the ideal gas) chemical potential [22]. The calculation of excess chemical potentials within the H-AdResS formalism closely resembles the thermodynamic integration method [23]. For such a reason we have called it spatially-resolved thermodynamic integration (`SPARTIAN`) [22].

In the second part of the paper, we test the current numerical implementation of the `SPARTIAN` method [22]. In particular, we compute the excess chemical potential of pure water, aqueous urea solutions and Lennard-Jones (LJ) mixtures. Additionally to the comparison with chemical potential values reported in the literature, when available, we tested the convergence of the results with model parameters such as size of the hybrid region, system size and composition. Our results confirm the method's robustness and efficiency and establish it as an alternative approach to compute free energy differences. We also assess the method's computing efficiency concerning fully atomistic simulations. These results highlight the necessity to either fine-tune the domain decomposition conditions in the ideal gas region or to reduce its size to a minimum and use our recently developed particle-insertion method [22] to include an infinite reservoir effectively.

The paper is organised as follows: In Section I.2, we present the H-AdResS method in terms of DFT. In Section I.3, we report the calculations of the test systems' excess chemical potential. Finally, we discuss our results and conclude in Section I.4.

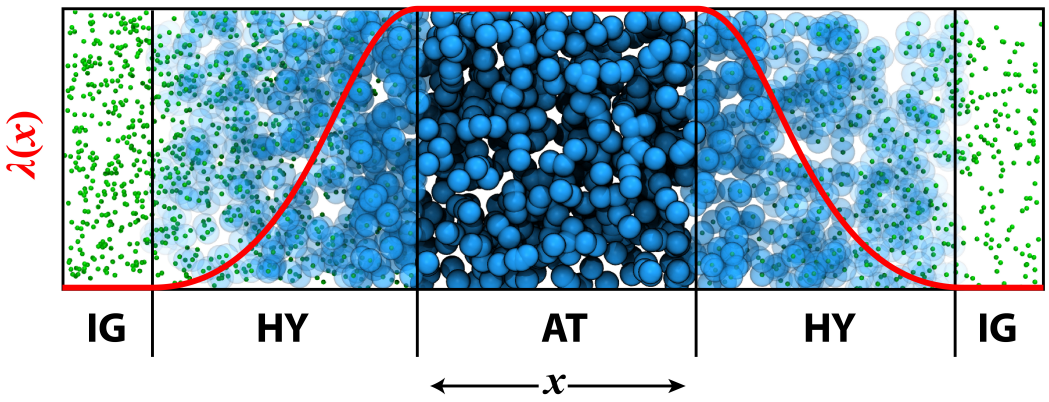

Figure I.1: Schematic representation of a typical adaptive resolution slab. The central, fully atomistic (AT), region is embedded into a reservoir of ideal gas (IG) particles. An interfacial, hybrid (HY), region connects AT and IG subsystems via a switching field λ that smoothly switches on (off) intermolecular interactions as molecules leave (enter) the IG region. An external field, not shown, counterbalances free energy barriers such that molecules freely diffuse between regions.

I.2 H-AdResS as an inhomogeneous system: connection to DFT

We write the adaptive resolution Hamiltonian [41, 40, 22] for a molecular fluid composed of N_a atoms, distributed among N molecules, as

$$H_{[\lambda]}(r,p) = \mathcal{K} + V^{\text{intra}} + \sum_{\alpha=1}^{N} \{\lambda_\alpha V_\alpha + V^{\text{ext}}(\lambda_\alpha)\}\,, \tag{I.1}$$

with (r,p) positions and momenta and $\mathcal{K} = \sum_{i=1}^{N_a} \mathbf{p}_i^2/2m_i$ is the total kinetic energy of the system. Latin indices run over atoms and greek indices over molecules. The term $V^{\text{intra}} = \sum_{\alpha=1}^{N}\sum_{i\neq j\in\alpha} V^{\text{intra}}(r_{ij})$ accounts for intra-molecular interactions, with r_{ij} the separation between atoms i and j belonging to the same molecule α. The intermolecular interactions are included in the term $V_\alpha = \frac{1}{2}\sum_{\beta\neq\alpha}\sum_{i\neq j} V(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|)$ with $\mathbf{r}_{\alpha i}$ the position of the atom i in molecule α. The switching field determines the molecules' identity, with $\lambda_\alpha \equiv \lambda(\mathbf{R}_\alpha)$ and $\mathbf{R}_\alpha$ the position of the centre of mass of the molecule α. When $\lambda = 0$ the Hamiltonian describes an homogeneous ideal gas system provided $V^{\text{ext}}(0) =$ constant. In particular, we set $V^{\text{ext}}(0) = 0$. For $0 < \lambda \leq 1$, the Hamiltonian describes an inhomogeneous system, namely, an interacting system in the presence of an external field (Figure I.1). We assume that the system (atomistic and ideal gas representations) is embedded in a reservoir of infinite size at temperature T and chemical potential μ^{id}, the chemical potential of the ideal gas. In practice, by using the Hamiltonian in Equation (VI.4), we have recently demonstrated that the numerical implementation of these conditions is straightforward [22]. Here, we calculate the grand canonical partition function corresponding to the adaptive resolution Hamiltonian (VI.4)

$$\mathcal{Z}[\lambda] = \text{Tr}\{\exp(-\beta(H_{[\lambda]} - \mu(\lambda)N))\}\,, \tag{I.2}$$

with $\beta = 1/k_{\text{B}}T$ and k_{B} being the Boltzmann's constant. We use the classical trace notation [39, 30] to avoid writing explicitly the integral over all particle momenta and positions and the sum over all possible system sizes

$$\text{Tr} = \sum_{N_a=0}^{\infty} \frac{1}{h^{3N_a} N_a!} \int d^{3N_a} r \int d^{3N_a} p\,,$$

with h the Planck's constant.
By using the partition function (I.2), we define the equilibrium distribution function as [39, 30]

$$f_0[\lambda] = \frac{1}{\mathcal{Z}[\lambda]} \exp(-\beta(H_{[\lambda]} - \mu(\lambda)N)\,, \tag{I.3}$$

which satisfies that $\text{Tr} f_0[\lambda] = 1$. The thermal average of an operator $\hat{O}$ is given by $\langle \hat{O} \rangle^{[\lambda]} = \text{Tr}\{f_0[\lambda]\hat{O}\}$. In particular, the external potential makes the system's density inhomogeneous. We use the previous definition of thermal average to write the equilibrium density profile in the form

$$\rho_0^{[\lambda]}(\mathbf{r}) = \langle \hat{\rho}(\mathbf{r}) \rangle^{[\lambda]}\,, \tag{I.4}$$

with $\hat{\rho}(\mathbf{r}) = \sum_\alpha \delta(\mathbf{R}_\alpha - \mathbf{r})$.
Using an arbitrary distribution function $f[\lambda]$, that satisfies the condition $\text{Tr} f[\lambda] = 1$, Mermin, in the

context of the inhomogeneous electron gas, proposed a functional form for the grand potential [26] that we write for our particular case as

$$\Omega_{[\lambda]}[f[\lambda]] = \mathrm{Tr}\{f[\lambda](H_{[\lambda]} - \mu(\lambda)N + \beta^{-1}\ln f[\lambda])\}\,. \tag{I.5}$$

By replacing the equilibrium distribution, we obtain

$$\Omega_{[\lambda]}[f_0[\lambda]] = -\beta^{-1}\ln \mathcal{Z}[\lambda] = \Omega^0_{[\lambda]} \tag{I.6}$$

the definition of grand potential in terms of the grand canonical partition function.
Using the density operator $\hat{\rho}$, we can rewrite the external potential as

$$\sum_{\alpha}^{N} V^{\mathrm{ext}}(\lambda_\alpha) = \int d\mathbf{r} \sum_{\alpha=1}^{N} \delta(\mathbf{R}_\alpha - \mathbf{r}) V^{\mathrm{ext}}(\lambda(\mathbf{r})) = \int d\mathbf{r}\, \hat{\rho}(\mathbf{r}) V^{\mathrm{ext}}(\lambda(\mathbf{r}))\,. \tag{I.7}$$

The density distribution $f[\lambda]$ is, in general, a functional of the density $\rho^{[\lambda]}(\mathbf{r})$. Therefore [39, 30, 2], by using the Hamiltonian (VI.4) we write the grand potential as a functional of the density

$$\begin{aligned}
\Omega_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})] &= \mathrm{Tr}\{f[\lambda](H_{[\lambda]} - \mu(\lambda)N + \beta^{-1}\ln f[\lambda])\} \\
&= \mathrm{Tr}\left\{f[\lambda]\left(\mathcal{K} + V^{\mathrm{int}} + \textstyle\sum_{\alpha=1}^{N}\lambda_\alpha V_\alpha + \beta^{-1}\ln f[\lambda] + \sum_{\alpha=1}^{N} V^{\mathrm{ext}}(\lambda_\alpha) - \mu(\lambda)N\right)\right\} \\
&= F_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})] + \textstyle\int d\mathbf{r}\, \rho^{[\lambda]}(\mathbf{r})\left(V^{\mathrm{ext}}(\lambda(\mathbf{r})) - \mu(\lambda(\mathbf{r}))\right),
\end{aligned} \tag{I.8}$$

in which we make use of Equation (I.7) where the total external potential is written as a functional of the density. The functional $F_{[\lambda]}[\rho^{[\lambda]}]$ is the intrinsic Helmholtz free energy corresponding to the Hamiltonian (VI.4), which is independent of the external potential:

$$F_{[\lambda]}[\rho^{[\lambda]}] = \mathrm{Tr}\left\{f[\lambda]\left(\mathcal{K} + V^{\mathrm{int}} + \sum_{\alpha=1}^{N}\lambda_\alpha V_\alpha + \beta^{-1}\ln f[\lambda]\right)\right\}\,. \tag{I.9}$$

The last two equations provide us with a direct connection between H-AdResS and density functional theory. The expression (I.8) is a functional Legendre transform relating the Helmholtz free energy and the grand potential [2].
To understand Equation (I.8), we underline that the system is inhomogeneous, and the grand potential is a functional of the external potential [30, 2]. That is

$$\frac{\delta\Omega_{[\lambda]}}{\delta(\mu(\lambda(\mathbf{r})) - V^{\mathrm{ext}}(\lambda(\mathbf{r})))} = -\rho^{[\lambda]}(\mathbf{r})\,, \tag{I.10}$$

which is consistent with the thermodynamic identity $\partial\Omega/\partial\mu = -N$.
Written as a functional of the density, $\Omega_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})]$ represents the cost in free energy necessary to find the system at precisely the density $\rho^{[\lambda]}(\mathbf{r})$ [2]. In particular, we find the density field, $\rho_0^{[\lambda]}(\mathbf{r})$, that minimises this cost by evaluating the functional derivative of the grand potential with respect to the density. Hence, the grand potential satisfies

$$\left.\frac{\delta\Omega_{[\lambda]}[\rho^{[\lambda]}]}{\delta\rho^{[\lambda]}}\right|_{\rho^{[\lambda]}=\rho_0^{[\lambda]}} = 0\,, \tag{I.11}$$

thus implying

$$\left.\frac{\delta F_{[\lambda]}[\rho^{[\lambda]}]}{\delta\rho^{[\lambda]}}\right|_{\rho^{[\lambda]}=\rho_0^{[\lambda]}} = -V^{\text{ext}}(\lambda(\mathbf{r})) + \mu(\lambda(\mathbf{r}))\,. \tag{I.12}$$

The exact form of the intrinsic free energy functional $F_{[\lambda]}[\rho^{[\lambda]}]$ is, in general, unknown. However, we can re-write it as

$$F_{[\lambda]}[\rho^{[\lambda]}] = F_{[\lambda]}^{\text{exc}}[\rho^{[\lambda]}] + F_{[0]}[\rho^{[0]}]\,, \tag{I.13}$$

with $F_{[\lambda]}^{\text{exc}}[\rho^{[\lambda]}]$ the excess free energy, calculated with respect to the free energy of the reference system (ideal gas), $F_{[0]}[\rho^{[0]}]$. To ensure thermodynamic consistency in the adaptive resolution description, we require the free energy being independent of λ [13]. Nevertheless, we anticipate here that this condition could be changed to induce non-equilibrium conditions in the system [22]. In equilibrium, $F_{[\lambda]}^{\text{exc}}[\rho^{[\lambda]}] = 0$, thus,

$$F_{[0]}[\rho^{[0]}(\mathbf{r})] = F^{\text{id}}[\rho(\mathbf{r})] = \beta^{-1}\int d\mathbf{r}\rho(\mathbf{r})\{\ln(\lambda_T^3\rho(\mathbf{r})) - 1\}\,, \tag{I.14}$$

with F^{id} the Helmholtz free energy of the ideal gas (our reference state at $\lambda = 0$), $\rho^{[0]}(\mathbf{r}) = \rho(\mathbf{r})$ and $\lambda_T = (\hbar^2\beta/2\pi m)^{1/2}$ the thermal, de Broglie, wavelength. We rewrite Equation (I.12) as

$$\frac{\delta F^{\text{id}}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} = -V^{\text{ext}}(\lambda(\mathbf{r})) + \mu^{\text{exc}}(\lambda(\mathbf{r})) + \mu^{\text{id}}\,, \tag{I.15}$$

with $\mu^{\text{exc}}(\lambda(\mathbf{r})) = \mu(\lambda(\mathbf{r})) - \mu^{\text{id}}$, a continuous function of λ, $\mu^{\text{id}} = \beta^{-1}\ln(\lambda_{\text{dB}}^3\rho_0)$ and ρ_0 the reference (ideal gas) density. By replacing μ^{id} into the previous equation, we find

$$\rho(\mathbf{r}) = \rho_0 \exp\left(-\beta\{V^{\text{ext}}(\lambda(\mathbf{r})) - \mu^{\text{exc}}(\lambda(\mathbf{r}))\}\right), \tag{I.16}$$

where it is apparent that this density field does not explicitly depend on the switching field. Indeed, this expression shows that the external potential determines the density. In particular, the condition

$$V^{\text{ext}}(\lambda(\mathbf{r})) = \mu^{\text{exc}}(\lambda(\mathbf{r})) \tag{I.17}$$

guarantees a constant density ρ_0. In other words, the external potential that enforces the reference density for the whole system is equal to the excess chemical potential.

This result provides an interesting parallel with density functional theory. In the latter, we apply an external potential and then use a self-consistent approach to find the corresponding system's equilibrium density. By considering the one-to-one correspondence between external potential and equilibrium density as provided by DFT, the approach generally used in H-AdResS naturally follows. We enforce a uniform density throughout the simulation box and then calculate the resulting external potential $V^{\text{ext}}(\lambda(\mathbf{r}))$. This is particularly useful to compute chemical potentials of molecular fluids and liquid mixtures, since $V^{\text{ext}}(1) = \mu^{\text{exc}}$ with μ^{exc} the excess chemical potential of the atomistic system over the ideal gas.

To use the Hamiltonian (VI.4) to perform molecular dynamics simulations, we need to compute the forces. The force, calculated as minus the gradient of the Hamiltonian, acting on a molecule α is given by:

$$\mathbf{F}_\alpha = \mathbf{F}_\alpha^{\text{int}} + \sum_{\beta\neq\alpha}\left\{\frac{\lambda_\alpha + \lambda_\beta}{2}\mathbf{F}_{\alpha|\beta}\right\} - \nabla_\alpha\lambda_\alpha\left[V_\alpha + V'^{\text{ext}}(\lambda)\big|_{\lambda=\lambda_\alpha}\right], \tag{I.18}$$

with $\mathbf{F}_\alpha^{\text{int}}$ and $\mathbf{F}_{\alpha|\beta}$ the forces due to the intra- and intermolecular potentials, respectively, and $V'^{\text{ext}} = dV^{\text{ext}}/d\lambda$. The force proportional to the gradient of the switching field weakly violates Newton's third law and momentum conservation, and introduces spurious density and pressure inhomogenenities into the system. Since we do not know *a priori* the external potential, we cannot directly compute its derivative with respect to λ. However, we can obtain a first approximation to $V'^{\text{ext}}(\lambda)$ by taking a closer look at the invariance of the Helmholtz free energy with respect to the switching field [13]

$$\frac{\delta F_{[\lambda]}(\rho(\mathbf{r}))}{\delta\lambda(\mathbf{r})} = \frac{\delta\Omega_{[\lambda]}(\rho(\mathbf{r}))}{\delta\lambda(\mathbf{r})} - \frac{\delta}{\delta\lambda(\mathbf{r})}\int d\mathbf{r}\,\rho(\mathbf{r})\left(V^{\text{ext}}(\lambda(\mathbf{r})) - \mu(\lambda(\mathbf{r}))\right) = 0\,, \tag{I.19}$$

where we use the inverse of Equation (I.8) for the density field $\rho(\mathbf{r})$. Following the condition given by Equation (I.17), the functional derivative of the integral vanishes. Using Equation (I.6), we find

$$\frac{\delta\Omega_{[\lambda]}(\rho(\mathbf{r}))}{\delta\lambda(\mathbf{r})} = \text{Tr}\left\{f[\lambda]\frac{\delta H_{[\lambda]}}{\delta\lambda(\mathbf{r})}\right\} - \text{Tr}\left\{N\frac{\exp(-\beta H_{[\lambda]})}{\mathcal{Z}[\lambda]}\frac{\delta z[\lambda]}{\delta\lambda(\mathbf{r})}\right\} = 0\,, \tag{I.20}$$

with $z[\lambda] = \exp(\beta\mu(\lambda))$ the system's fugacity. In general, the two terms on the r.h.s of the previous equation cancel each other out to guarantee a uniform density profile. There is also the possibility that every term remains invariant with respect to changes in $\lambda(\mathbf{r})$. In particular, if the fugacity of the system does not depend on λ, then the pressure, and not the density, is constant throughout the simulation box [13]. This assumption implies that

$$\begin{aligned}\text{Tr}\left\{f[\lambda]\frac{\delta H_{[\lambda]}}{\delta\lambda(\mathbf{r})}\right\} &= \text{Tr}\left\{f[\lambda]\textstyle\sum_{\alpha=1}^{N}(V_\alpha + V'^{\text{ext}}(\lambda(\mathbf{r})))\delta(\mathbf{R}_\alpha - \mathbf{r})\right\}\\ &= \left\langle\textstyle\sum_{\alpha=1}^{N} V_\alpha\delta(\mathbf{R}_\alpha - \mathbf{r})\right\rangle^{[\lambda]} + V'^{\text{ext}}(\lambda(\mathbf{r}))\rho^{[\lambda]}(\mathbf{r}) = 0\,,\end{aligned} \tag{I.21}$$

which allows us to evaluate V'^{ext} in terms of thermal averages over quantities that we know. From a practical viewpoint, we use the local equilibrium approximation discussed in Ref. [13]. When λ is sufficiently smooth, the thermal average is approximated to an average computed at a constant value of the switching field at a given position $\mathbf{r}$. That is, $\langle\cdots\rangle^{[\lambda]} \approx \langle\cdots\rangle^{\lambda}$ [13]. By integrating over space we obtain,

$$\langle V\rangle^{\lambda} + V'^{\text{ext}}(\lambda)N = 0\,, \tag{I.22}$$

with $V = \sum_\alpha V_\alpha$. Finally, we obtain [41]

$$V'^{\text{ext}}(\lambda)|_{\lambda=\lambda_\alpha} = -\langle V\rangle^{\lambda=\lambda_\alpha}/N \approx -\langle V\rangle_{\mathbf{R}_\alpha}\,. \tag{I.23}$$

By inserting this *drift* term in the force (I.18) we see that, on average, the term proportional to the gradient of λ vanishes. Therefore, we obtain a description of the system that on average satisfies Newton's third law and linear momentum conservation. This is an important issue because significant artifacts are introduced in the simulation if these conditions are not fulfilled [12]. Furthermore, by using this force, we are implicitly ensuring a constant pressure throughout the simulation box [41, 13].

We now return to adaptive resolution setups at constant density. To enforce this condition, we compute and apply, iteratively, an external force, dubbed *thermodynamic force* [17, 18], given by

$$\mathbf{F}_{n+1}^{\text{th}} = \mathbf{F}_n^{\text{th}} + \frac{c\nabla\rho(x)_n}{\rho_0}\,, \tag{I.24}$$

with c a parameter with units of energy. $\nabla\rho(x)_n$ is the gradient of the density profile computed at the n-th step of the iteration. The computation of this force converges when $\nabla\rho = 0$, that is, when the density becomes ρ_0 everywhere.
Hence, we have that the total external force acting on a molecule α, instantaneously present in the hybrid region, becomes

$$-\nabla_{\mathbf{R}_\alpha} V^{\text{ext}}(\lambda)|_{\lambda=\lambda_\alpha} = \langle V\rangle_{\mathbf{R}_\alpha} \nabla\lambda(\mathbf{R}_\alpha) + \mathbf{F}^{\text{th}}_\alpha \,. \tag{I.25}$$

The integral over space of this force will give us the external potential that, in virtue of Equation (I.17), equates the excess chemical potential, provided the density is constant and equal to ρ_0 throughout the simulation box. Due to the similitude between this approach and Kirkwood's thermodynamic integration (KTI) [23], in the following sections, we refer to this method to compute free energy differences as the spatially-resolved thermodynamic integration (`SPARTIAN`) method [22]. The comparison with KTI also implies that to compute μ^{exc} it is sufficient to enforce the target density at the AT and IG regions, which corresponds to the selection of a different thermodynamic path to compute the integral.
We conclude this section highlighting that the generalisation of this procedure to the study of mixtures is straightforward [40, 39]. In the following section, we will provide a few examples where excess chemical potentials of complex liquids and liquid mixtures are computed and compared with results available in the literature.

I.3 Simulation Results

In this section, we present the calculation of the excess chemical potential for SPC/E [5, 13, 62] water, Lennard-Jones mixtures and aqueous urea solutions [44, 36, 8]. The aim is to investigate convergence and performance of the calculation using various simulation setups.

I.3.1 Preliminaries

We use the LAMMPS [56] implementation of `SPARTIAN` described in Ref. [22, 18]. Namely, we discretise the hybrid region in bins of size $\delta\lambda$ and δr. To compute the forces, we use the iterative procedure described in Ref. [18]. We sample each term on the r.h.s. of Equation (I.25) separately, using independent time-intervals. At the end of every interval, the code generates two files containing $\langle V(\lambda)\rangle$ and $\mathbf{F}^{\text{th}}(r)$. As anticipated in the previous section, the analysis of a liquid mixture only requires to compute the corresponding forces acting on every individual component. Thus, in the case of a spherical atomistic region, with radius r_{AT} and thickness of the hybrid region d_{HY}, the excess chemical potential for the i-component is calculated as

$$\mu_i^{\text{exc}} = \int_0^1 d\lambda\, \langle V_i(\lambda)\rangle + \int_{r_{\text{AT}}}^{r_{\text{AT}}+d_{\text{HY}}} dr\, \mathbf{F}_i^{\text{th}}(r)\cdot\hat{n}\,, \tag{I.26}$$

with $\hat{n}$ a unit vector pointing away from the atomistic region. The switching function is defined as

$$\lambda(r) = \begin{cases} 1 & r \le r_{\text{AT}} \\ \cos^{14}\left(\frac{\pi(r-r_{\text{AT}})}{2d_{\text{HY}}}\right) & r_{\text{AT}} < r \le r_{\text{AT}} + d_{\text{HY}} \\ 0 & r > r_{\text{AT}} + d_{\text{HY}}\,, \end{cases} \tag{I.27}$$

with the exponent 14 selected for numerical convenience [22].
To compute the thermodynamic force (I.24), we use $c = 1.0$, 0.01 kcal/mol and $2\,\epsilon$ to modulate the contribution from $\nabla\rho$ in the case of SPC/E water, aqueous urea solutions and Lennard-Jones mixtures, respectively. To get a smooth force field, we convolute the density profile with Gaussian functions of width equal to three times the length of a hybrid region bin. During the simulation, it is possible that two molecules approach each other in the hybrid region while travelling towards the atomistic region. Once there, if they get too close to each other, the strong repulsion might create numerical instabilities. To avoid this problem, we introduce a capping radius for all the interactions. The corresponding values are 0.1 σ for Lennard-Jones mixtures and 0.5 Å for Van der Waals and electrostatic potentials in water and aqueous urea solutions [22].
We use the damped-shifted force potential model (DSF) [45, 47, 14, 15] to describe electrostatic interactions. To compare with the results of free energy differences obtained by using Ewald summation, we need to subtract a constant term [22]. In the case of rigid molecules using the SHAKE algorithm [40], this term corresponds to electrostatic intramolecular interactions between i, j atomic pairs, given by $q_i q_j / r_{ij}$ with q_i the charge of the i-th atom.

I.3.2 Size of the hybrid region

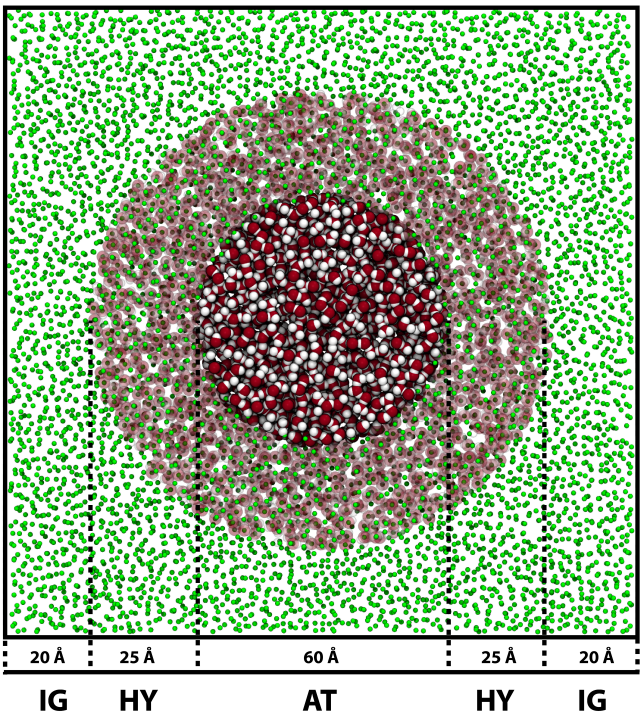

Figure I.2: Simulation snapshot showing the `SPARTIAN` configuration for a pure water calculation. A spherical atomistic (AT) region or radius 30 Å is embedded into a simulation box of linear size of 150 Å. The hybrid (HY) region is a spherical shell of maximum thickness 25 Å, and the ideal gas (IG) occupies the remaining free space.

We start with the calculation of $\mu^{\text{exc}}_{\text{WATER}}$ for pure water. We present the simulation setup used in Figure I.2. We consider a system of 111375 SPC/E [5, 13, 62] water molecules in a cubic simulation box, with linear size equal to 150.0 Å. A spherical atomistic (AT) region with radius 30 Å is embedded into a hybrid (HY) region of thickness varying between 10, 15, 20 and 25 Å. The starting point of the calculation corresponds to a fully atomistic simulation equilibrated for 25 ns at $T = 298.0$ K and $P = 1.0$ bar using the Nosé-Hoover thermostat and barostat with damping constant equal to 100 fs for both cases. After this step, the systems were equilibrated by 25 ns in the NVT ensemble using Langevin thermostat with damping parameter equal to 100 fs and a integration time step of 1.0 fs.

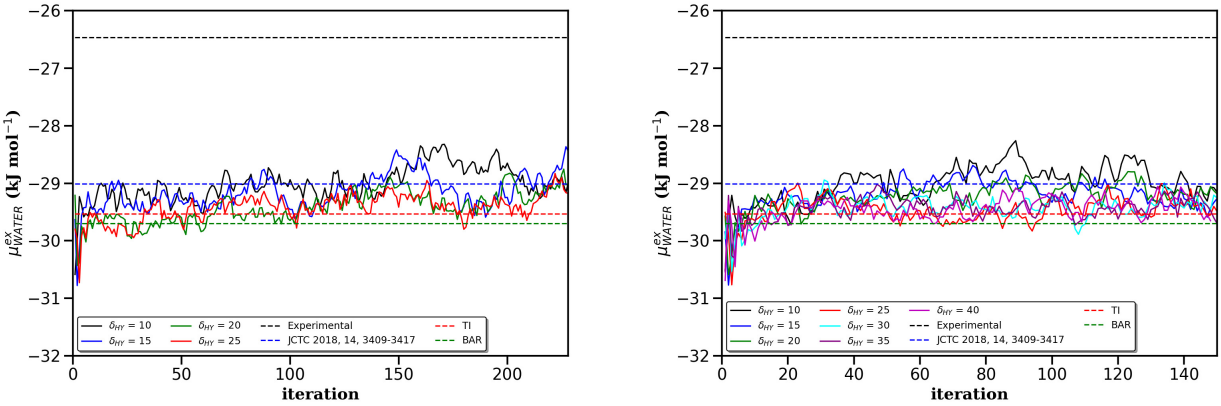

Figure I.3: (left) Excess chemical potential for water obtained for different hybrid region thicknesses. The linear size of the simulation box is 150 Å. (right) Excess chemical potential for water obtained for different hybrid region thicknesses. The linear size of the simulation box is 200 Å. In addition to reporting our previous result [22], we compare with the experimental value [3], and with computational results obtained by using thermodynamic integration (TI) [27] and the Bennett acceptance-ratio methods (BAR) [41].

By starting from the final fully-atomistic configuration, we start the `SPARTIAN` runs. We simultaneously average both contributions to the force (I.18) for 20 ps, then at the end of this period, we calculate and apply the forces. This cycle corresponds to a single iteration. After every iteration we use Equation (I.26) to obtain the value of $\mu^{\text{exc}}_{\text{WATER}}$. Results of this procedure, for different sizes of the HY region, are presented in Figure I.3(top). For reference, we also include the experimental value [3], as well as computational results obtained by using thermodynamic integration [27] and the Bennett acceptance-ratio methods [41].
It is apparent from the Figure that the final result is somewhat insensitive to the size of the HY region. With small values of δ_{Hy}, namely 10 and 15 Å, there are larger fluctuations in the result related to poor statistics. Once the sample size increases, $\delta_{\text{Hy}} = 20$ and 25 Å, the result convincingly converges after 100 iterations (2 ns) to the value obtained by using the state-of-the-art computational methods considered here.
This picture remains essentially consistent upon increasing the size of the system. Figure I.3(bottom) shows similar results for a simulation box of linear size equal to 200 Å. In this case, additionally to the values considered before, we can also investigate the convergence of the result for values

of $\delta_{\mathrm{HY}} = 30$, 35 and 40 Å. Similarly to the previous case, the result smoothly converges to the expected $\mu^{\mathrm{exc}}_{\mathrm{WATER}}$ value for the SPC/E model.
The results of this sub-section, averaged over the last twenty iterations, are summarised in Table I.1.

To conclude this section, we comment on how to choose the size of the HY region. To avoid

Table I.1: Excess chemical potential for water

δ_{HY} (Å)	μ_{ext} (kJ mol^{-1})	
	L=150 Å	L=200 Å
10.0	-28.79 ± 0.18	-28.93 ± 0.13
15.0	-28.91 ± 0.24	-29.31 ± 0.14
20.0	-29.08 ± 0.17	-29.17 ± 0.11
25.0	-29.27 ± 0.21	-29.41 ± 0.16
30.0		-29.36 ± 0.14
35.0		-29.40 ± 0.16
40.0		-29.38 ± 0.17

*standard deviation measured over the last twenty iterations.

artefacts arising from finite-size effects, the volume of the AT region should contain the volume defined by the correlation length of the system. Our results indicate that this size constraint also applies to the HY region. Hence, it is advisable to choose the linear size of the HY region larger than the correlation length of the system.

I.3.3 HPC performance

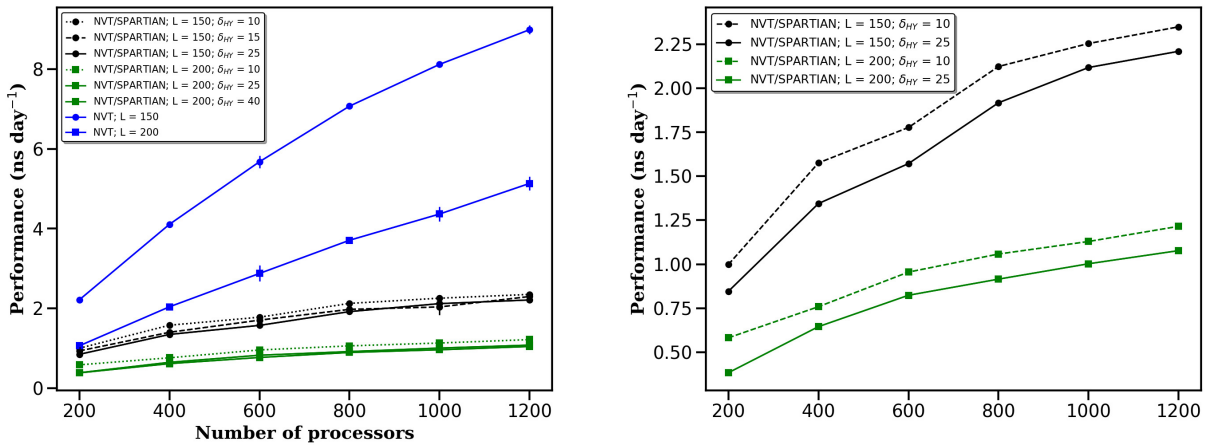

Figure I.4: (left) Performance of the `SPARTIAN` and NVT as a function of the number of processors. (right) Performance of the `SPARTIAN` and NVT as a function of the number of processors for hybrid region thickness equal to 10.0 Åand 25.0 Å

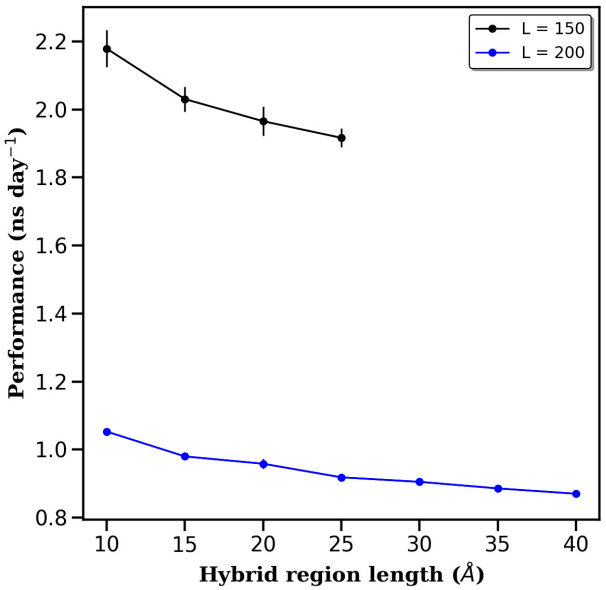

Figure I.5: Performance of the SPARTIAN calculation as a function of the hybrid region thickness. Calculations were performed by using 800 processors.

By using the water system considered in the previous sub-section, we perform a strong scaling analysis. Namely, we select a different number of processors and evaluate the performance of the calculation, measured in units of ns per day. Results are presented in Figure I.4(top) where it is apparent that fully-atomistic simulations, for both sizes included in this study (blue symbols), exhibit a better scaling than their SPARTIAN counterparts (black and green symbols). This behaviour is due to lack of optimisation in domain decomposition and load balancing in the SPARTIAN implementation. Processors assigned to the IG region run more idle than the ones assigned to the AT and HY regions thus hindering parallel processing and communication.
Upon increasing the size of the system, in both fully-atomistic and SPARTIAN calculations there is a substantial and proportionally similar decrease in computing efficiency. Furthermore, in the case of SPARTIAN calculations, as presented in Figure I.4(bottom), there is a slight decrease in efficiency upon increasing the size of the hybrid region while keeping the size of the atomistic region constant. Indeed, Figure I.5 illustrates the performance of the SPARTIAN calculation as a function of the hybrid region thickness, for systems with fixed linear sizes (150 and 200 Å), atomistic region size, and number of processors (800). Both tendencies confirm that the rather modest SPARTIAN performance is mainly due to the parallelisation and communication issues mentioned above. The default domain decomposition in LAMMPS consists of dividing the simulation box into three-dimensional equal-size subvolumes, where every subdomain gets assigned to one processor. In the SPARTIAN case, we combine different pair styles that result in a mismatch in computational load depending on whether particles are in the atomistic, hybrid or ideal gas regions. Hence, overall poor performance results from an increase in computational cost due to a uniform assignment of particles per processor.

To overcome this limitation, we use the shift scheme as implemented in LAMMPS [1]. Here, we assign a weight factor that tends to balance the computational load of each processor. A weight factor (> 1) associated with a given particle determines how much more computational resources are devoted to this particle than the other particles present in the simulation. To illustrate this point, we consider SPC/E water in a simulation box of size 150 Å, running on (i) 200, (ii) 800 and (iii) 1200 processors. By using a weight factor of ten between the particles in the AT and HY regions and the particles in the IG region, we obtain an increase in performance of (i) 60%, (ii) 47% and (iii) 24%, respectively.
These results are encouraging, however, a more important advancement, from both fundamental and technical perspectives, is related to the implementation of particle-insertion methods that allows us to significantly reduce the size of the IG region while effectively including an infinite size reservoir [22].

I.3.4 Binary Mixtures: Extremely Diluted Conditions

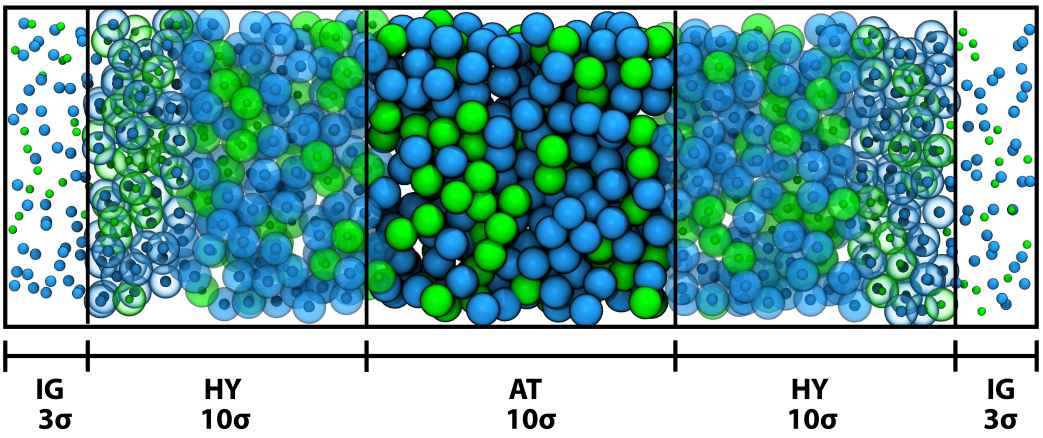

Figure I.6: Lateral view of the tetragonal simulation setup used to investigate diluted Lennard-Jones mixtures. The long axis of the box, where the adaptive resolution takes place, is aligned along the x-direction. The y-direction, with normal vector entering the page, is used to linearly increase the size of the system, thus increasing the number of particles (statistics) in the HY region.

The accuracy of the calculation of the excess chemical potential in the `SPARTIAN` method improves upon increasing the number of molecules present in the HY region, as it is apparent from Equation (I.18). Thus, we expect that the `SPARTIAN` calculation requires a major increase in HY region size to converge μ^{exc} for low-density liquids and highly-diluted mixtures. To investigate these conditions, we simulate A-B Lennard-Jones mixtures with $x_{\text{A}} = 0.1$ the mole fraction of A-molecules. In this case, we use a tetragonal simulation box, $L_x \times L_y \times L_z$ with its major axis, L_x, aligned along the x-direction (Figure I.6). The switching field $\lambda(x)$ is applied only along L_x and, in all cases considered, we keep constant the length of the AT, HY and IG regions. We also keep constant L_z. To increase the size of the system, and therefore the statistics for the `SPARTIAN` calculation, we

[1] https://lammps.sandia.gov/doc/balance.html

start from $L_y = L_z$ and keep increasing L_y. To sum up, we have a simulation box with $L_x = 36\ \sigma$, $L_z = 5\ \sigma$ and variable L_y.
We simulate systems with 3.6, 10.8, 36, 54, 100, 250, 300, 500, 1000 and 1500 ($\times\ 10^3$) particles. The results are expressed in Lennard-Jones reduced units. The force field parameters were chosen as $\sigma_{AA} = \sigma_{AB} = \sigma_{BB} = 1.0\ \sigma$, and $\epsilon_{AA} = \epsilon_{BB} = \epsilon_{AB} = 1.0\ \epsilon$, and the interactions were truncated to a cutoff radius equal to 2.5 σ. The systems were equilibrated in the NPT ensemble with $T^* = 2.0\ k_B\epsilon^{-1}$ and $P^* = 5.0\ \sigma^3\epsilon^{-1}$, by $1.0 \times 10^5\ \tau$, using Nosé-Hoover thermostat and barostat with damping coefficient equal to 10 τ and 100 τ, respectively and time step equal to 0.005 τ. Then, the systems were simulated in the NVT ensamble with $T^* = 2.0\ k_B\epsilon^{-1}$, by $1.0 \times 10^5\ \tau$, using Langevin thermostat with damping parameter equal to 100 τ and time step equal to 0.001 τ. In all cases, the `SPARTIAN` calculation were run with the atomistic region width and the hybrid region thickness equal to 10 σ and 10 σ, respectively, as shown in Figure I.6. The drift and thermodynamic forces were calculated every 10 τ.

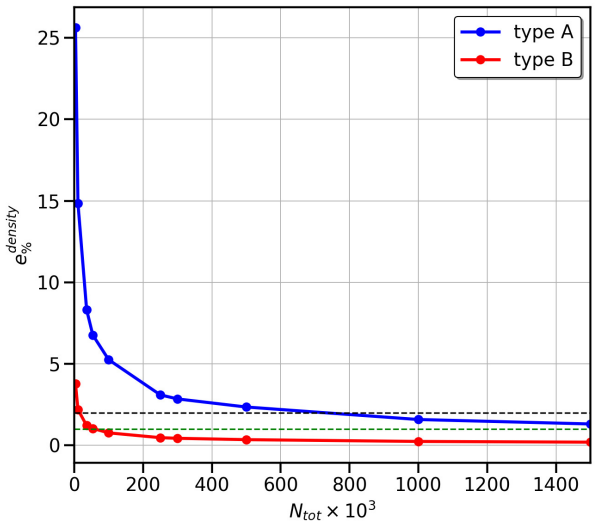

Figure I.7: Error (deviation from target value) in the density as a function of total number of particles. $N_A = 0.1\ N_{tot}$. Black and green dashed lines correspond to 2 % and 1 % limits, respectively.

One diagnostic tool to evaluate the quality of the `SPARTIAN` calculation is the density profile. Figure I.7 shows the error, written as a relative difference from the target density, as a function of the number of particles in the LJ mixture. As expected, due the high concentration of B-particles (0.9 N_{tot}), the deviation from the target value for this species is less than 2 %. This is the case for all the systems considered, except for the smallest system size (3600 particles). For the systems with larger number of particles, the error quickly decreases below 1 %. However, for A-particles, the deviation from the target density only reaches the mark of 2 % for systems composed of more than one million particles. It does not reach the 1 % limit even for the system with 1.5 million particles.
Figure I.8 shows the influence of the density in the calculation of the excess chemical potential using

SPARTIAN. Figure I.8(top) displays the excess chemical potential of B-particles, $\mu_{\mathrm{B}}^{\mathrm{exc}}$, as a function of the total number of particles in the system. It is apparent that $\mu_{\mathrm{B}}^{\mathrm{exc}}$ moderately fluctuates in the systems with small number of particles and that this fluctuation significantly decreases for larger systems. However, due to the low concentration of A-particles, and high deviation in their density profile with respect to the target density, the excess chemical potential, $\mu_{\mathrm{A}}^{\mathrm{exc}}$, fluctuates more distinctly for systems with total number of particles < 250000, as shown in Figure I.8(bottom). These results are summarised in Table I.2 where excess chemical potential values, averaged over the last twenty SPARTIAN iterations, are presented as a function of the system size.

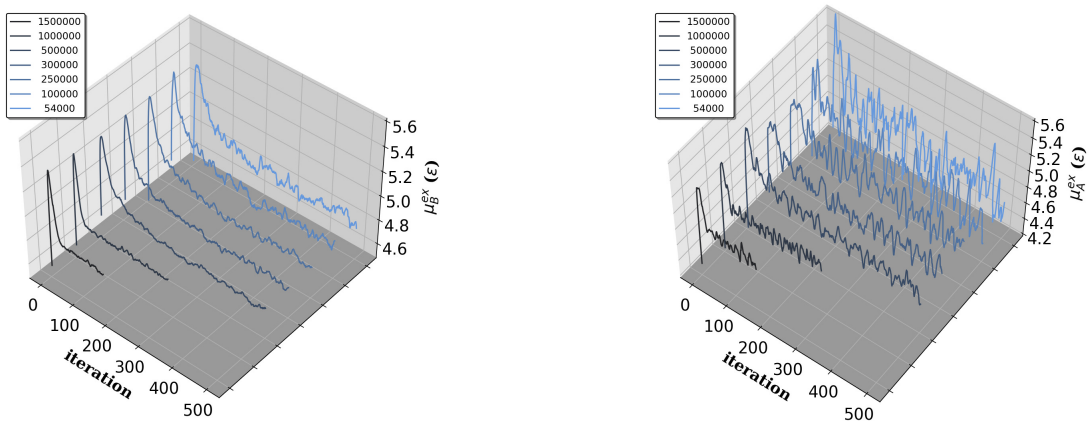

Figure I.8: (left) Excess chemical potential convergence for B-particles. $N_{\mathrm{B}} = 0.9\ N_{\mathrm{tot}}$. (right) Excess chemical potential convergence for A-particles. $N_{\mathrm{A}} = 0.1\ N_{\mathrm{tot}}$

Table I.2: Excess chemical potential (and standard deviation) for the LJ mixture as a function of system size. The data was obtained using the last twenty SPARTIAN iterations.

N_{tot}	$\mu_{exc}(\epsilon)$	
	A	B
3600	5.186 ± 0.671	5.100 ± 0.092
54000	4.567 ± 0.148	4.757 ± 0.021
100000	4.661 ± 0.147	4.746 ± 0.024
250000	4.718 ± 0.034	4.731 ± 0.010
300000	4.718 ± 0.098	4.715 ± 0.015
500000	4.653 ± 0.075	4.727 ± 0.008
1000000	4.782 ± 0.033	4.769 ± 0.004
1500000	4.777 ± 0.045	4.780 ± 0.005

Table I.3 and Figure I.9 show the relative error $\sigma^{\%}(\mu_i^{\mathrm{exc}})$ in the excess chemical potential for particles of type A and B, calculated over the last twenty SPARTIAN iterations. The value of the relative error for B-particles, $\sigma^{\%}(\mu_{\mathrm{B}}^{\mathrm{exc}})$, is less than 1.0 % for all systems considered (except for the smallest number of particles - 3600). However, for A-particles, the value of $\sigma^{\%}(\mu_{\mathrm{A}}^{\mathrm{exc}})$ only falls

above 1.0 % for systems with size equal or larger than one million particles, reaching the limit of 2.0 % for systems with $N_{\text{tot}} > 300{,}000$ particles.

Table I.3: Relative error in the excess chemical potential for LJ mixture

N_{tot}	$\sigma^{\%}(\mu_i^{\text{exc}})$	
	A	B
3600	12.9476	1.8074
54000	3.2379	0.4346
100000	3.1607	0.5141
250000	0.7209	0.2039
300000	2.0773	0.3241
500000	1.6033	0.1610
1000000	0.6850	0.0790
1500000	0.9340	0.1020

* Relative standard deviation for the last twenty iterations.

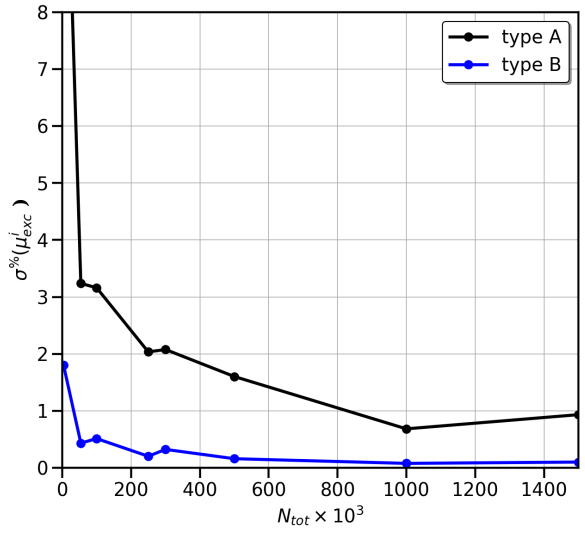

Figure I.9: Relative error in the excess chemical potential, calculated over the last twenty iterations, as a function of total number of particles. $N_{\text{A}} = 0.1\ N_{\text{tot}}$ and $N_{\text{B}} = 0.9\ N_{\text{tot}}$

I.3.5 Binary Mixtures: SPARTIAN and Kirkwood-Buff Analysis

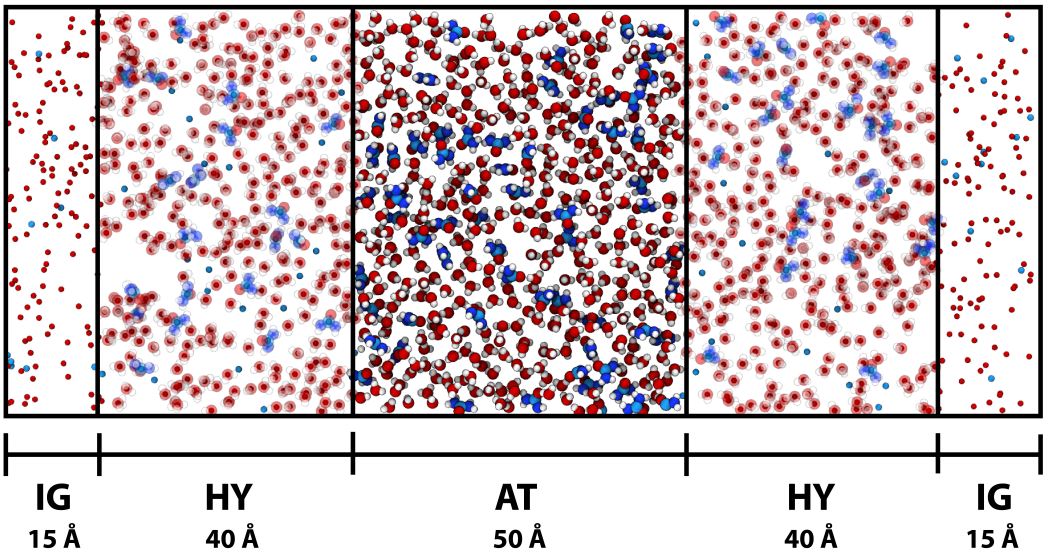

Figure I.10: Lateral view of a slice of the tetragonal simulation setup used to investigate aqueous urea solutions. The long axis of the box, where we apply the adaptive resolution, is aligned along the x-direction. The linear size in the y-direction is 62.0 Å (the slice's thickness is 5.0 Å)

The final example corresponds to the calculation of the excess chemical potential for a more realistic binary mixture. In this case, we consider aqueous urea solutions [44, 36, 8]. We expect that, when increasing the system's complexity, SPARTIAN becomes a robust alternative to existing computational methods.

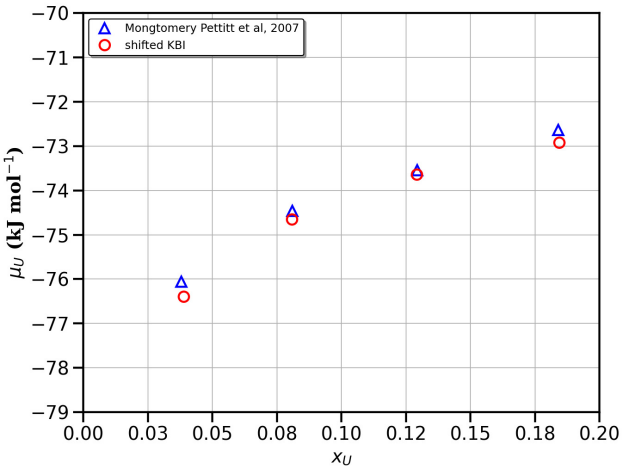

Figure I.11: Chemical potential of urea in aqueous solution as a function of urea mole fraction.

An initial configuration having 14500 SPC/E [5, 13, 62] water molecules and 2147 urea molecules [44]

representing an aqueous urea solution at 0.1292 mole fraction (6M) is generated and optimised by adjusting the y-direction of the simulation cell. The size of the resulting simulation box is $160\times62\times60$ Å^3. To equilibrate the system, we performed fully atomistic simulations in the NPT ensemble for 20 ns at 298.0 K at 1.0 bar pressure using Nosé-Hoover thermostat and barostat with damping coefficients 100 fs and 1000 fs, respectively. This step is followed by 10 ns simulation in the NVT ensemble using Langevin thermostat with a damping coefficient of 100 fs. The `SPARTIAN` setup is presented in Figure I.10. Every iteration takes 20 ps to be completed.
We also performed fully-atomistic simulations at different urea mole fractions: 0.0389 (2M), 0.0809 (4M), 0.1292 (6M) and 0.1844 (8M) and carried out a Kirkwood-Buff (KB) analysis [6, 19] to compute the shifted urea chemical potential as a function of solute concentration. We use the `SPARTIAN` calculation at 6M as a reference value for the KB analysis. Results are presented in Figure I.11 where a comparison with absolute values reported in the literature [24] is also included. Our results indicate that the combination of the KB-based method to obtain the trend with concentration, together with a more refined method like `SPARTIAN` to compute the reference value of the chemical potential, constitutes a useful tool for the calculation of free energy differences in complex molecular fluids and mixtures.

I.4 Concluding remarks

In the H-AdResS method [41, 40], the simulation of coexisting atomistic and ideal gas representations of a physical system is made possible by introducing a switching field that modulates the intermolecular potential. By including in the Hamiltonian an external potential, functional of the switching field, we build a grand potential for an inhomogeneous system. With this grand potential at hand, we find a parallel with DFT. Thus, we show that the external potential that balances the density across the simulation box is the excess chemical potential of the atomistic system [22]. Due to DFT's one-to-one correspondence between the density and the external potential, the use of one-body forces to impose a constant density profile, whose integral automatically gives the system's excess chemical potential, is compatible with a Hamiltonian description of the system. The resemblance with thermodynamic integration is apparent. Therefore, we dubbed the method `SPARTIAN` (spatially-resolved thermodynamic integration) [22].
We then use the `SPARTIAN` method to compute μ^{exc} for water, aqueous urea solutions and Lennard-Jones mixtures. Our results well compare with values reported in the literature, when available. We verify that upon increasing the size of the hybrid region, which improves the statistics while keeping constant the AT region's size, μ^{exc} consistently converges with a modest decrease in computing efficiency. Our results also show an excess computational overhead when compared to fully atomistic simulations. This overhead is mainly due to domain decomposition and communication issues that result from an inefficient assignment of processors in the IG region. A preliminary experiment shows that `SPARTIAN` efficienty improves with a moderate fine-tuning effort aiming at optimising domain decomposition features. Nevertheless, the computation of chemical potentials is efficient and accurate. For water and aqueous urea solutions, we get a well-converged result after a few `SPARTIAN` iterations that correspond to simulate the system for just a couple of ns. Furthermore, size constraints are not an issue in `SPARTIAN` thanks to the particle-insertion method that we have recently developed, implemented and validated [22]. We also consider highly-diluted conditions in which we expect the `SPARTIAN` method to underperform compared to state-of-the-art

computational methods. Still, our results display a convincing convergence of the excess chemical potential for all species involved in the system under such conditions. However, this requires a substantial increase in the system size. Hence, depending on the system's complexity, the choice of a computational method becomes a trade-off between efficiency, user-friendliness and accuracy. In particular, standard thermodynamic integration or particle-insertion-based methods applied to molecular mixtures require a careful design of an alchemical path. In the `SPARTIAN` case, once we obtain the initial configuration (following standard molecular dynamics) and decide the parameters, the simulation automatically finds μ^{exc}. Indeed, depending on the problem at hand, a combination of methods might result advantageous. In case of investigating a solution's phase diagram, where it is necessary to compute μ^{exc} as a function of the solute's concentration, we propose to combine a Kirkwood-Buff theory approach with only one `SPARTIAN` calculation at a given (preferably high) concentration.
Finally, the thermodynamically consistent coupling of atomistic and ideal representations opens various interesting research avenues [38]. A straightforward application concerns the calculation of free energy differences [20]. Moreover, non-equilibrium conditions [42] can be readily imposed on the ideal gas reservoir and investigated on the atomistic region [22]. Finally, this and similar ideas originated from the adaptive resolution method could also be incorporated into QM/MM approaches [5, 10].

Acknowledgement

The authors thank Nancy C. Forero-Martinez for her critical reading of the manuscript. The authors gratefully acknowledge funding from SFB-TRR146 of the German Research Foundation (DFG). R.P. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant 758588). Simulations have been performed on the THINC cluster at the Max Planck Institute for Polymer Research and on the COBRA cluster of the Max Planck Computing and Data Facility.

Bibliography

[1] Animesh Agarwal, Jinglong Zhu, Carsten Hartmann, Han Wang, and Luigi Delle Site. Molecular dynamics in a grand ensemble: Bergmann–lebowitz model and adaptive resolution simulation. *New J. Phys.*, 17(8):083042, aug 2015.

[2] Nathan Argaman and Guy Makov. Density functional theory: An introduction. *American Journal of Physics*, 68(1):69–79, 2000.

[3] A. Ben-Naim and Y. Marcus. Solvation thermodynamics of nonionic solutes. *The Journal of Chemical Physics*, 81(4):2016–2027, 1984.

[4] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[5] Jelle M. Boereboom, Raffaello Potestio, Davide Donadio, and Rosa E. Bulo. Toward hamiltonian adaptive qm/mm: Accurate solvent structures using many-body potentials. *Journal of Chemical Theory and Computation*, 12(8):3441–3448, 2016. PMID: 27332140.

[6] R. Cortes-Huerto, K. Kremer, and R. Potestio. Communication: Kirkwood-buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *The Journal of Chemical Physics*, 145(14):141103, 2016.

[7] Liem X. Dang and B. Montgomery. Pettitt. Simple intramolecular model potentials for water. *J. Phys. Chem.*, 91(12):3349–3354, 1987.

[8] Tiago E. de Oliveira, Paulo A. Netz, Kurt Kremer, Christoph Junghans, and Debashish Mukherji. C-ibi: Targeting cumulative coordination within an iterative protocol to derive coarse-grained models of (multi-component) complex fluids. *J. Chem. Phys.*, 144(17):174106, 2016.

[9] Rafael Delgado-Buscalioni, Kurt Kremer, and Matej Praprotnik. Concurrent triple-scale simulation of molecular liquids. *J. Chem. Phys.*, 128(11):114110, 2008.

[10] Luigi Delle Site. Simulation of many-electron systems that exchange matter with the environment. *Advanced Theory and Simulations*, 1(10):1800056, 2018.

[11] Luigi Delle Site, Christian Krekeler, John Whittaker, Animesh Agarwal, Rupert Klein, and Felix Höfling. Molecular dynamics of open systems: Construction of a mean-field particle reservoir. *Adv. Theor. Sim.*, 2(5):1900014, 2019.

[12] Burkhard Dünweg. Molecular dynamics algorithms and hydrodynamic screening. *J. Chem. Phys.*, 99(9):6977–6982, 1993.

[13] P. Español, R. Delgado-Buscalioni, R. Everaers, R. Potestio, D. Donadio, and K. Kremer. Statistical mechanics of hamiltonian adaptive resolution simulations. *The Journal of Chemical Physics*, 142(6):064115, 2015.

[14] G. Fanourgakis. An extension of wolf's method for the treatment of electrostatic interactions: Application to liquid water and aqueous solutions. *J. Phys. Chem. B*, 119(5):1974–1985, 2015.

[15] Christopher J. Fennell and J. Daniel Gezelter. Is the ewald summation still necessary? pairwise alternatives to the accepted standard for long-range electrostatics. *J. Chem. Phys.*, 124(23), 2006.

[16] S. Fritsch, C. Junghans, and K. Kremer. Structure formation of toluene around c60: Implementation of the adaptive resolution scheme (adress) into gromacs. *J. Chem. Theory Comput.*, 8(2):398–403, FEB 2012.

[17] S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, and K. Kremer. Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir. *Phys. Rev. Lett.*, 108:170602, Apr 2012.

[18] M. Heidari, R. Cortes-Huerto, D. Donadio, and R. Potestio. Accurate and general treatment of electrostatic interaction in hamiltonian adaptive resolution simulations. *The European Physical Journal Special Topics*, 225(8):1505–1526, 2016.

[19] M. Heidari, K. Kremer, R. Potestio, and R. Cortes-Huerto. Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations. *Molecular Physics*, 116(21-22):3301–3310, 2018.

[20] Maziar Heidari, Robinson Cortes-Huerto, Raffaello Potestio, and Kurt Kremer. Steering a solute between coexisting solvation states: Revisiting nonequilibrium work relations and the calculation of free energy differences. *The Journal of Chemical Physics*, 151(14):144105, 2019.

[21] Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Spatially resolved thermodynamic integration: An efficient method to compute chemical potentials of dense fluids. *Journal of Chemical Theory and Computation*, 14(7):3409–3417, 2018. PMID: 29874069.

[22] Maziar Heidari, Kurt Kremer, Ramin Golestanian, Raffaello Potestio, and Robinson Cortes-Huerto. Open-boundary hamiltonian adaptive resolution. from grand canonical to non-equilibrium molecular dynamics simulations. *The Journal of Chemical Physics*, 152(19):194104, 2020.

[23] John G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3(5):300–313, 1935.

[24] H. Kokubo, J. Rösgen, D. Wayne Bolen, and B. Montgomery Pettitt. Molecular basis of the apparent near ideality of urea solutions. *Biophys. J.*, 93:3392, 2007.

[25] Kreis, K., Fogarty, A.C., Kremer, K., and Potestio, R. Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations. *Eur. Phys. J. Special Topics*, 224(12):2289–2304, 2015.

[26] N. David Mermin. Thermal properties of the inhomogeneous electron gas. *Phys. Rev.*, 137:A1441–A1443, Mar 1965.

[27] Zoltan Mester and Athanassios Z Panagiotopoulos. Mean ionic activity coefficients in aqueous nacl solutions from molecular dynamics simulations. *J. Chem. Phys.*, 142(4):044507, 2015.

[28] Debashish Mukherji and Kurt Kremer. Coil–Globule–Coil Transition of PNIPAm in Aqueous Methanol: Coupling All-Atom Simulations to Semi-Grand Canonical Coarse-Grained Reservoir. *Macromolecules*, 46(22):9158–9163, 2013.

[29] Debashish Mukherji, Nico F. A. van der Vegt, and Kurt Kremer. Preferential solvation of triglycine in aqueous urea: An open boundary simulation approach. *J. Chem. Theory Comput.*, 8(10):3536–3541, 2012.

[30] M. Oettel. Classical density functional theory : Exact density distribution of hard rods between walls in 1 d. 2018.

[31] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.*, 117(1):1 – 19, 1995.

[32] Raffaello Potestio, Pep Español, Rafael Delgado-Buscalioni, Ralf Everaers, Kurt Kremer, and Davide Donadio. Monte carlo adaptive resolution simulation of multicomponent molecular liquids. *Phys. Rev. Lett.*, 111:060601, Aug 2013.

[33] Raffaello Potestio, Sebastian Fritsch, Pep Español, Rafael Delgado-Buscalioni, Kurt Kremer, Ralf Everaers, and Davide Donadio. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.*, 110:108301, Mar 2013.

[34] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.*, 123(22):224106–14, 2005.

[35] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids. *Phys. Rev. E*, 73:066701, 2006.

[36] M. Praprotnik, L. Delle Site, and K. Kremer. A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation. *J. Chem. Phys.*, 126:134902, 2007.

[37] M. Praprotnik, L. Delle Site, and K. Kremer. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Ann. Rev. Phys. Chem.*, 59(1):545–571, 2008.

[38] S. Qi, H. Behringer, T. Raasch, and F. Schmid. A hybrid particle-continuum resolution method and its application to a homopolymer solution. *The European Physical Journal Special Topics*, 225(8):1527–1549, 2016.

[39] R. Roth. Introduction to density functional theory of classical systems: Theory and applications. 2006.

[40] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327 – 341, 1977.

[41] J. Sauter and A. Grafmueller. Predicting the chemical potential and osmotic pressure of polysaccharide solutions by molecular simulations. *J. Chem. Theory Comput.*, 12(9):4375–4384, 2016.

[42] S. Stalter, L. Yelash, N. Emamy, A. Statt, M. Hanke, M. Lukáĉová-Medvid'ová, and P. Virnau. Molecular dynamics simulations in hybrid particle-continuum schemes: Pitfalls and caveats. *Comput. Phys. Comm.*, 224:198 – 208, 2018.

[43] Han Wang, Carsten Hartmann, Christof Schütte, and Luigi Delle Site. Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique. *Phys. Rev. X*, 3:011018, Mar 2013.

[44] Samantha Weerasinghe and Paul E. Smith. A kirkwood-buff derived force field for mixtures of urea and water. *J. Phys. Chem. B*, 107(16):3891–3898, 2003.

[45] D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht. Exact method for the simulation of coulombic systems by spherically truncated, pairwise 1/r summation. *J. Chem. Phys.*, 110(17):8254–8282, 1999.

[46] Yujie Wu, Harald L. Tepper, and Gregory A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.*, 124(2):024503, 2006.

[47] D. Zahn, B. Schilling, and S. M. Kast. Enhancement of the wolf damped coulomb potential: Static, dynamic, and dielectric properties of liquid water from molecular simulation. *J. Phys. Chem. B*, 106(41):10725–10732, 2002.

II. Finite-size scaling and thermodynamics of model supercooled liquids: long-range concentration fluctuations and the role of attractive interactions

Outline

Bibliographic Information

Atreyee Banerjee, Mauricio Sevilla, Joseph F. Rudzinski, and Robinson Cortes-Huerto. Finite-size scaling and thermodynamics of model supercooled liquids: long-range concentration fluctuations and the role of attractive interactions. *Soft Matter*, 18:2373–2382, 2022.

Author contribution

Atreyee Banerjee and Mauricio Sevilla have contributed equally to this work.

§ Atreyee Banerjee: Conceptualization, methodology, formal analysis, writing – original draft.

§ Mauricio Sevilla: Methodology, software, formal analysis, visualisation.

§ Joseph F. Rudzinski: investigation, writing – review & editing.

§ Robinson Cortes-Huerto: conceptualisation, methodology, writing – original draft, review & editing.

Abstract

We compute partial structure factors, Kirkwood-Buff integrals (KBIs) and chemical potentials of model supercooled liquids with and without attractive interactions. We aim at investigating whether relatively small differences in the tail of the radial distribution functions result in contrasting thermodynamic properties. Our results suggest that the attractive potential favours the nucleation of long-range structures. Indeed, upon decreasing temperature, Bathia-Thornton structure factors display anomalous behaviour in the $k \to 0$ limit. KBIs extrapolated to the thermodynamic limit confirm this picture, and excess coordination numbers identify the anomaly with long-range concentration fluctuations. By contrast, the purely repulsive system remains perfectly miscible for the same temperature interval and only reveals qualitatively similar concentration fluctuations in the crystalline state. Furthermore, differences in both isothermal compressibilities and chemical potentials show that thermodynamics is not entirely governed by the short-range repulsive part of the interaction potential, emphasising the nonperturbative role of attractive interactions. Finally, at higher density, where both systems display nearly identical dynamical properties and repulsive interactions become dominant, the anomaly disappears, and both systems also exhibit similar thermodynamic properties.

II.1 Introduction

The supercooled state challenges our understanding of the theory of liquids. In particular, the connection between dynamics, which varies considerably upon supercooling, and structure, which appears to remain essentially unchanged, is the subject of intense research. [10, 59, 11, 34, 29, 35, 6] Model systems with reduced complexity, still retaining essential physical features, provide a direct route to investigate this problem. For example, Kob–Andersen mixtures [31] with purely repulsive Weeks-Chandler-Andersen interactions (KAWCA) [64] exhibit substantially different dynamics compared to their Lennard-Jones counterpart (KALJ). [31] By contrast, their structure, investigated from the point of view of radial distribution functions, is somewhat similar. [6, 7]
The connection between pair correlations and dynamical properties has been extensively investigated. [21, 7] On the one hand, a variety of studies conclude that two-body contributions are not enough to account for the difference in dynamics between the KAWCA and KALJ systems. Perhaps the most well-known example is mode-coupling theory, based on pair correlation functions, which underestimates these dynamical differences. [7] Additionally, deviations in many-body structural descriptors such as triplet [14] and point-to-set correlations, [25] as well as bond-order distributions [60] and the packing capabilities of local particle arrangements, [59] have been observed between the KALJ and KAWCA systems. These results indicate that higher-order features may be necessary to resolve the difference in their dynamical properties. [54]
On the other hand, several studies indicate that two-body structure is enough to describe particular aspects of the dynamics of model supercooled liquids. For example, features based on the pair structure have been used to predict diffusion constants from short-time trajectories of the KALJ model. [16, 12, 50] Concerning the comparison between models, Bhattacharyya and coworkers [7, 2] directly explored structure-dynamics relationships in KALJ and KAWCA systems. In particular, they used the Adam–Gibbs relation, [1] to connect relaxation time to the configurational entropy. Their results demonstrated that the two-body contribution to the entropy plays a significant role in distinguishing the dynamics of the two systems.

To further contribute to the discussion, recent research efforts have focused on the detailed characterization of the liquid's two-body structure. In particular, *softness parameters*, defined via weighted integrals of pair-correlation functions [15, 34] or multi-dimensional integrals of partial structure factors, [36] respond to minor structural changes and can accurately describe dynamical differences. However, either non-trivial reweighting procedures or combinations of local and nonlocal terms prevents an unambiguous identification of the dominant, short- versus long-range, contributions to the resulting structure-dynamics relationship.
The potentially dominant role of short-range pair correlations brings with it yet another dilemma. According to perturbation theory, short-range repulsive interactions mostly dominate the liquid's structure. [64] By contrast, based on Kirkwood-Buff theory, [28] long-range fluctuations in the tail of the pair correlation function have a significant effect on the system's solvation thermodynamics. [51, 47, 56, 32, 13] The studies mentioned above investigating KALJ and KAWCA dynamics have mainly focused on short-range contributions. Nevertheless, evidence for the nucleation of long-range structures in glassy systems at low temperatures [13, 39, 51, 65] highlights the necessity to carefully address this point. Finite-size effects present in computer simulations dramatically affect the tail of the pair correlation function and the $k \to 0$ limit of the structure factor, i.e., the long-range structure properties, which in turn sensitively impact thermodynamic quantities. Consequently, a careful evaluation of finite-size effects becomes critical for investigating these properties in the supercooled regime.
In this paper, we investigate various thermodynamic properties of KALJ and KAWCA a − b mixtures in the supercooled liquid state. We calculate structure factors of density, $S_{\rho\rho}(k)$, and concentration, $S_{cc}(k)$, while highlighting the $k \to 0$ limit. The KALJ liquid exhibits anomalous behaviour reflected in a major increase in concentration fluctuations. This anomaly closely resembles the nucleation of nanometric clusters reported by Fischer in low-temperature ortho-terphenyl, [13, 39] and it has been recently identified as a general feature present in polydisperse colloidal models. [29] By contrast, the purely repulsive KAWCA system remains perfectly miscible in the supercooled state. A finite-size Kirkwood–Buff analysis confirms this picture by enabling the precise identification of the $k \to 0$ limit. Furthermore, we show that the isothermal compressibility and chemical potential of the two models exhibit similar trends with temperature, apart from constant shifts. These differences highlight the nonperturbative role of attractive interactions in the system. To sum up, we demonstrate that seemingly small differences in the tail of the radial distribution function result in significantly different structural and thermodynamic properties for supercooled systems with and without attractive interactions.
The paper is organised as follows: we provide the computational details in Sec. II.2, present the results in Sec. II.3 and conclude in Sec. II.4.

II.2 Computational Details

We have simulated the Kob–Andersen model, which is a binary mixture (80:20) of Lennard-Jones (KALJ) particles. [31] The inter-atomic pair potential between species α and β, $U_{\alpha\beta}(r)$, with $\alpha, \beta = \mathrm{a}, \mathrm{b}$ is described by a shifted and truncated Lennard–Jones potential[61], as given by:

$$U_{\alpha\beta}(r) = \begin{cases} U_{\alpha\beta}^{(LJ)}(r; \sigma_{\alpha\beta}, \epsilon_{\alpha\beta}) - U_{\alpha\beta}^{(LJ)}(r_{\alpha\beta}^{(c)}; \sigma_{\alpha\beta}, \epsilon_{\alpha\beta}), & r \leq r_{\alpha\beta}^{(c)} \\ 0, & r > r_{\alpha\beta}^{(c)} \end{cases} \tag{II.1}$$

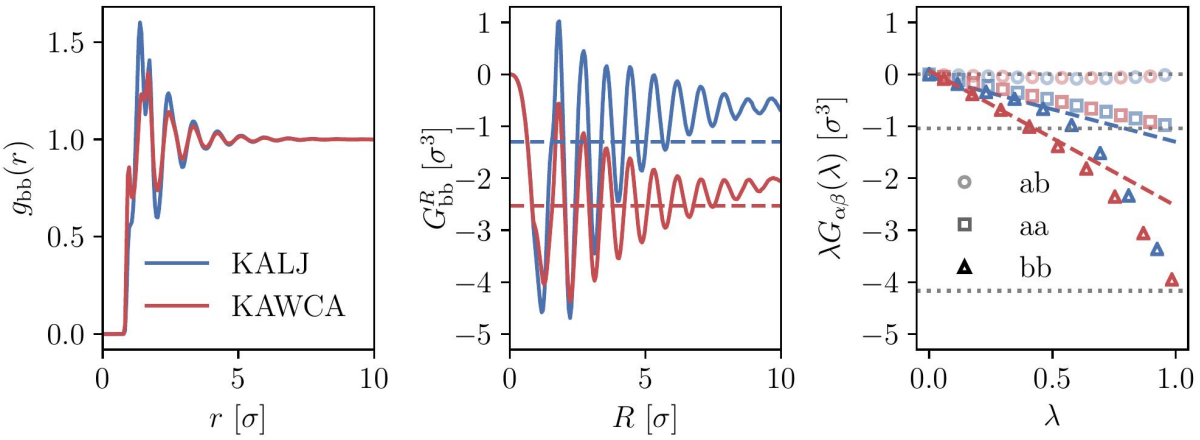

Figure II.1: Differences between KALJ and KAWCA systems in terms of the $g_{bb}(r)$ component and the KBIs at $T = 0.45\epsilon/k_{\rm B}$. (a) Differences between the RDF for the low-concentration b-component of the mixture seem to be small and mostly coming from the local structure of the fluid. (b) $G^R_{\rm bb}$ as obtained from Eq. (II.3) shows a different short-range behaviour and, more importantly, the tails do not converge due to finite-size effects. (c) KBIs obtained using the method described in Ref. [13] (Eq. (II.4)). The KBIs in the thermodynamic limit $G^\infty_{\alpha\beta}$ are obtained from the slope of a linear fitting of the region $0 < \lambda < 0.3$. This straight line is indicated for the bb case. Horizontal, dark lines correspond to the asymptotic limit $-\delta_{\alpha\beta}/\rho_\alpha$ with $\delta_{\alpha\beta}$ the Kronecker delta and ρ_α the number density of the α-species. The $G^\infty_{\alpha\beta}$ values obtained in this way are plotted in panel (b) as horizontal lines.

where $U^{(LJ)}_{\alpha\beta}(r;\sigma_{\alpha\beta},\epsilon_{\alpha\beta}) = 4\epsilon_{\alpha\beta}[(\sigma_{\alpha\beta}/r)^{12} - (\sigma_{\alpha\beta}/r)^6]$ and $r^{(c)}_{\alpha\beta}$ is equal to $2.5\sigma_{\alpha\beta}$ for LJ system and $r^{(c)}_{\alpha\beta}$ is equal to the position of the minimum of $U^{(LJ)}_{\alpha\beta}$ for the WCA systems (KAWCA). [64] We have added a linear correction so that both the potential and the force go to zero continuously at the cutoff distance. [61] We have used LJ natural units, such that length, temperature and time are measured in $\sigma_{\rm aa}$, $k_{\rm B}T/\epsilon_{\rm aa}$ and $\tau = \sqrt{(m_{\rm a}\sigma^2_{\rm aa}/\epsilon_{\rm aa})}$, respectively. For all the simulations, we have used the following interaction parameters $\sigma_{\rm aa} = 1.0\ \sigma$, $\sigma_{\rm ab} = 0.8\sigma$, $\sigma_{\rm bb}$ =0.88σ, $\epsilon_{\rm aa} = 1.0\ \epsilon$, $\epsilon_{\rm ab} = 1.5\epsilon$, $\epsilon_{\rm bb} = 0.5\epsilon$, $m_{\rm a} = m_{\rm b} = 1.0m$.

We have performed two different sets of simulations: the first for the calculation of structural properties, and the second for the calculation of chemical potentials, which employed a different box geometry and number of particles. All simulations have been carried out using the LAMMPS molecular dynamics software [41]. We have performed the first set of simulations in a cubic box with periodic boundary conditions in the canonical ensemble (NVT), using the Nosé-Hoover thermostat [19] with an integration timestep of 0.005τ and a time constant of 100 timesteps. The system is composed of $N = 23328$ particles, with $N_{\rm a} = 18664$ particles of type a. We have simulated this system at two different densities, $\rho = 1.2/\sigma^3$ and $1.6/\sigma^3$ for different temperatures, as specified in the main text. Starting from the high temperature case, the final configuration of the simulation has been used as an initial configuration for the simulation one (temperature) step below. The same procedure has been followed for the KALJ and KAWCA systems. For all state points, three to five independent simulations with run lengths $> 100\tau_\alpha$ (τ_α is the α-relaxation time estimated from Ref. [2]) have been performed.

To calculate the excess chemical potential, we have used the LAMMPS[41] implementation

of SPARTIAN already described in Ref. [22]. The SPARTIAN method, a variant of the adaptive resolution method [44, 45, 46, 47, 41, 40], simulates the coexistence of an atomistic system to its ideal gas representation at a constant density and temperature. We have computed the excess chemical potential of the system as the external potential required to balance the density across the simulation box. To guarantee enough statistics, we have used a slab geometry (An anisotropic box with $L_x = 36\sigma$, $L_y = 578\sigma$ and $L_z = 10\sigma$), also with periodic boundary conditions and at density $\rho = 1.2/\sigma^3$, resulting in a system with $N = 250000$ and $N_a = 200000$. The same protocol as described above has been used to quench the system before performing the SPARTIAN calculation. For the SPARTIAN method calculation, we have considered an slab geometry with atomistic region of length of 10σ and hybrid regions of linear size 10σ aligned along the x direction.
After equilibration, we have performed the SPARTIAN calculations in the canonical ensemble (NVT), using a Langevin thermostat with $dt = 0.001\tau$ and damping parameter of 10τ. In order to get the correct density profiles and therefore, chemical potential, we have simulated for 3×10^6 simulation steps.

II.3 Results and discussions

II.3.1 Kirkwood Buff analysis

We consider temperatures in the range $0.45\epsilon/k_B \leq T \leq 6\epsilon/k_B$ for KALJ system and $0.3\epsilon/k_B \leq T \leq 6\epsilon/k_B$ for KAWCA system (See Section II.2). Visual inspection of the radial distribution functions (RDFs) for both systems reveals that they are almost indistinguishable (Figure S1), and only the RDF $g_{bb}(r)$ for the minor component shows relatively small differences, visible at $r < 3\sigma$ (Figure II.1(a)). [40, 2] However, this direct comparison is misleading: a few thermodynamic quantities are quite sensitive to small fluctuations in the tail of the RDFs.
One such quantities are the Kirkwood–Buff integrals (KBIs), [28] which relate the microscopic structure of a liquid mixture to its solvation thermodynamics. For a multi-component system of species α and β, in equilibrium at temperature T, the KBIs in the thermodynamic limit (TL) take the form

$$G^{\infty}_{\alpha\beta} = 4\pi \int_0^{\infty} dr\, r^2 (g_{\alpha\beta}(r) - 1)\,, \tag{II.2}$$

where $g_{\alpha\beta}$ is the radial distribution for an infinite, open system. Here, it is obvious from Eq. II.2 that small deviations for large r result in important contributions to $G_{\alpha\beta}$. In computer simulations, usually far from the thermodynamic limit, Equation (II.2) is often approximated as

$$G^{R}_{\alpha\beta} = 4\pi \int_0^{R} dr\, r^2 (g^{c}_{\alpha\beta}(r) - 1)\,, \tag{II.3}$$

where $g^c_{\alpha\beta}(r)$ is the RDF of the closed, finite, system and R is a truncation radius. It is essential to choose R larger than the correlation length of the system. Nevertheless, this expression seldom converges due to different finite-size effects. Here, it is already clear that G^R_{bb} for the KALJ and KAWCA systems displays different behaviour (See Figure II.1(b)).
By explicitly including finite-size effects due to the thermodynamic ensemble and the finite inte-

gration domains, we compute the KBIs as [13]

$$\lambda G_{\alpha\beta}(\lambda) = \lambda G^{\infty}_{\alpha\beta}\left[1-\lambda^3\right] - \lambda^4 \frac{\delta_{\alpha\beta}}{\rho_\alpha} + \frac{c_{\alpha\beta}}{V_0^{\frac{1}{3}}}\,, \tag{II.4}$$

with $\lambda \equiv (V/V_0)^{\frac{1}{3}}$, $G^{\infty}_{\alpha\beta}$ being the value of the KBIs in the thermodynamic limit, $c_{\alpha\beta}$ a constant with units of length, $\delta_{\alpha\beta}$ the Kronecker delta and ρ_α the number density of the α-species. We can compute $G_{\alpha\beta}(\lambda)$, the KBIs for a subdomain of volume V inside a simulation box of volume V_0, in terms of fluctuations of the number of particles [51, 47, 56, 32, 13, 23, 24]

$$G_{\alpha\beta}(\lambda) = V\left(\frac{\langle N_\alpha N_\beta\rangle' - \langle N_\alpha\rangle'\langle N_\beta\rangle'}{\langle N_\alpha\rangle'\langle N_\beta\rangle'} - \frac{\delta_{\alpha\beta}}{\langle N_\alpha\rangle'}\right), \tag{II.5}$$

where $G_{\alpha\beta}(\lambda) \equiv G_{\alpha\beta}(V;V_0)$ and the average number of α-particles, $\langle N_\alpha\rangle' \equiv \langle N_\alpha\rangle_{V,V_0}$, depends on both the subdomain and simulation box volumes. Figure II.1(c) shows the results obtained from Eqs (II.4) and (II.5) for the KALJ and KAWCA systems at $T = 0.45\epsilon/k_{\mathrm{B}}$. These curves are rather similar in both cases, with a major difference appearing for the G^{∞}_{bb} component, which can be obtained as the slope of a linear fit of $G_{bb}(\lambda)$ within the region $\lambda < 0.3$. The resulting values of G^{∞}_{bb} are plotted as dashed lines in Figure II.1(b) to indicate the value at which the KBIs should converge.

II.3.2 Density and concentration structure factors

As anticipated, fluctuations in the tail of the radial distribution function affect the long-range structure of the fluid. Hence, to investigate these effects, we compute partial structure factors

$$S_{\alpha\beta}(k) = x_\alpha\delta_{\alpha\beta} + 4\pi x_\alpha x_\beta \rho \int_0^\infty dr\, r^2 \frac{\sin kr}{kr}(g_{\alpha\beta}(r)-1)\,, \tag{II.6}$$

where k is the norm of a reciprocal-lattice vector, $\rho = \rho_{\mathrm{a}} + \rho_{\mathrm{b}}$ is the total number density and $x_\alpha = N_\alpha/N$ is the mole fraction of the α-species. To avoid numerical instabilities at the low k limit [42], we compute the structure factor directly from the simulated trajectory [7] using the following expression as,

$$S_{\alpha\beta}(\mathbf{k}) = \frac{1}{N}\left\langle \sum_{i\in\alpha}^{N_\alpha}\sum_{j\in\beta}^{N_\beta} \exp(-i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j))\right\rangle, \tag{II.7}$$

where α and β denote the species, and the indexes i and j run over particles belonging to α and β, respectively. The average runs over the values of $\mathbf{k}$ such that $|\mathbf{k}| = k$ and over the ensemble. We follow the standard procedure in computer simulations to compute the structure factor. Namely, for a cubic simulation box of linear size L with periodic boundary conditions, we discretise the reciprocal space by considering wavevectors $\mathbf{k} = 2\pi(n_x, n_y, n_z)/L$ with n_x, n_y and n_z integer values. We use a maximum number $n_x^{\max} = n_y^{\max} = n_z^{\max} = 76$.
Partial structure factors are difficult to interpret for liquid mixtures. Hence, we focus on density, $S_{\rho\rho}(k)$, and concentration, $S_{cc}(k)$, structure factors [9] which carry a direct physical meaning. [33]

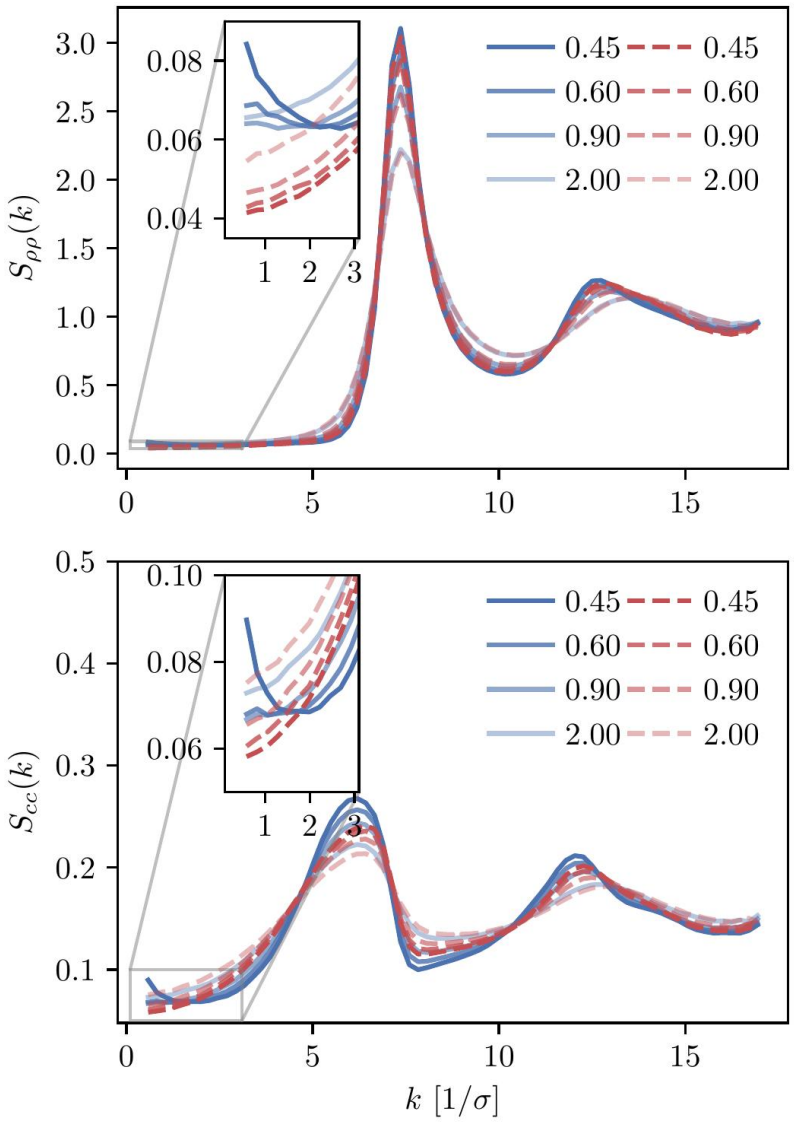

Figure II.2: Density, $S_{\rho\rho}$ (top), and concentration, S_{cc} (bottom), structure factors for both KALJ (blue) and KAWCA (red) systems for the temperatures considered here.

$S_{\rho\rho}(k)$ and $S_{cc}(k)$ describe the correlation of density and concentration fluctuations in the liquid mixture. They are defined as

$$\begin{aligned} S_{\rho\rho}(k) &= S_{\mathrm{aa}}(k) + S_{\mathrm{bb}}(k) + 2S_{\mathrm{ab}}(k)\,, \\ S_{cc}(k) &= x_{\mathrm{b}}^2 S_{\mathrm{aa}}(k) + x_{\mathrm{a}}^2 S_{\mathrm{bb}}(k) - 2x_{\mathrm{a}}x_{\mathrm{b}}S_{\mathrm{ab}}(k)\,. \end{aligned} \tag{II.8}$$

For large k-values, the behaviour of $S_{\rho\rho}$ and S_{cc} is rather similar for both systems (See Figure II.2). This includes a first peak at $k_0 \approx 7.13/\sigma$, followed by a second peak at approximately $1.7k_0$ that develops at low temperatures. This second peak is associated with the nucleation of structural motifs that precede the complete crystallisation of the system. As it has been reported for various metallic glasses, the splitting of this second peak [30] results from the optimal facet-sharing configurations of such structural (icosahedral and tetrahedral) motifs that grow upon

decreasing temperature. [62, 18] In our particular case, we do not observe this feature down to $T = 0.45\epsilon/k_B$, thus confirming that both systems remain liquid-like. Concerning the difference between the KALJ and the KAWCA systems, the first and second peaks in the S_{cc} show slightly more structure for the KALJ system at $T = 0.45\epsilon/k_B$, as expected from the RDF (See Figure II.1(a)).

Perhaps more interesting, it is apparent from the inset in Figure II.2 that the KALJ and KAWCA systems show substantially different behaviour in the region of small k (large r). On the one hand, the KAWCA liquid behaves like a normal liquid with monotonically decreasing density fluctuations upon decreasing temperature. On the other hand, the KALJ system exhibits anomalous behaviour, similar to SAXS curves obtained for ortho-terphenyl [39] and supercooled water, [27] with clear density fluctuations starting around $k \sim 2/\sigma$ ($r \sim 3\sigma$) appearing at temperatures lower than the onset temperature of glassy dynamics $T = 1\epsilon/k_B$ (See Figure S2). [3] These results indicate that the two systems display stark structural differences in the supercooled regime, with clear long-range density domains ($r > 3\sigma$) induced by the presence of attractive interactions in the KALJ mixture.

The extrapolation to the $k \to 0$ limit by using Eq. (II.6) or (II.7) is not trivial because finite-size effects in the simulation affect the precision in computing structure factors as we approach the linear size of the simulation box. In the next subsection, we use the relation between the structure factor in the limit $k \to 0$ and the KBIs to investigate this limiting case in more detail.

II.3.3 KBIs and the $k \to 0$ limit

Similar to the single component case, the extrapolation to the $k \to 0$ limit provides useful physical information. [30] Here, we use the relation between the limit $k \to 0$ in the structure factor, evaluated in Eq. (II.6), and the KBIs in the thermodynamic limit, Eq. (II.2). We obtain

$$\lim_{k\to 0} S_{\alpha\beta}(k) = x_\alpha \delta_{\alpha\beta} + \rho_\alpha x_\beta G^{\infty}_{\alpha\beta}\,, \tag{II.9}$$

thus

$$\begin{aligned}\lim_{k\to 0} S_{\rho\rho}(k) &= \rho_{\mathrm{a}} x_{\mathrm{a}} G^{\infty}_{\mathrm{aa}} + \rho_{\mathrm{b}} x_{\mathrm{b}} G^{\infty}_{\mathrm{bb}} + 2\rho_{\mathrm{a}} x_{\mathrm{b}} G^{\infty}_{\mathrm{ab}} + 1 \\ &= \rho k_{\mathrm{B}} T \kappa_T + \delta^2 \lim_{k\to 0} S_{cc}(k)\,.\end{aligned} \tag{II.10}$$

The last relation in Eq.(II.10) gives two contributions that allows us to connect long-range density fluctuations to both the isothermal compressibility κ_T of the system and to concentration fluctuations modulated by the difference in partial molar volumes $v_{\mathrm{a}} - v_{\mathrm{b}}$, with $\delta = \rho(v_{\mathrm{a}} - v_{\mathrm{b}})$. [30] The isothermal compressibility and the partial molar volumes can also be written in terms of the KBIs, namely:

$$\kappa_T = \frac{1 + \rho_{\mathrm{a}} G^{\infty}_{\mathrm{aa}} + \rho_b G^{\infty}_{\mathrm{bb}} + \rho_{\mathrm{a}}\rho_b (G^{\infty}_{\mathrm{aa}} G^{\infty}_{\mathrm{bb}} - G^{\infty\,2}_{\mathrm{ab}})}{k_B T \eta}\,, \tag{II.11}$$

and

$$\begin{aligned} v_{\mathrm{a}} &= \frac{1 + \rho_{\mathrm{b}}(G^{\infty}_{\mathrm{bb}} - G^{\infty}_{\mathrm{ab}})}{\eta}\,, \\ v_{\mathrm{b}} &= \frac{1 + \rho_{\mathrm{a}}(G^{\infty}_{\mathrm{aa}} - G^{\infty}_{\mathrm{ab}})}{\eta}\,, \end{aligned} \tag{II.12}$$

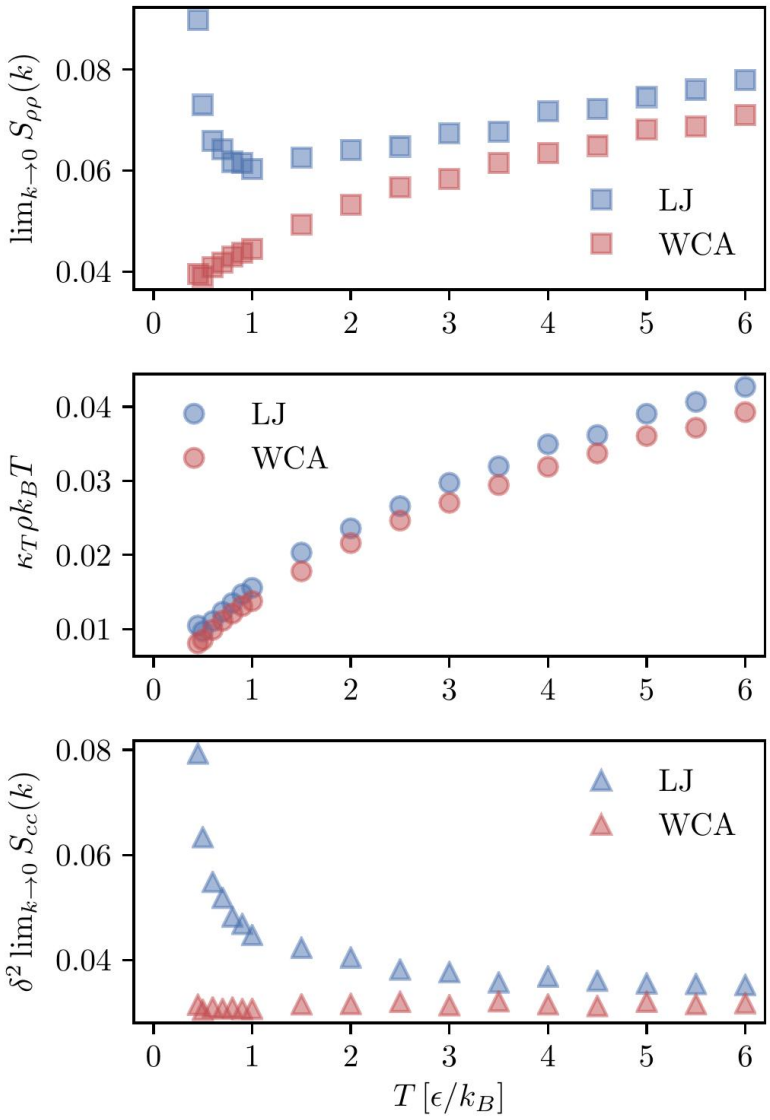

Figure II.3: Density-density correlation function $S_{\rho\rho}(k)$. (Upper panel) $\lim_{k\to 0} S_{\rho\rho}(k)$ obtained from the KBIs (Eq. (II.10)). At high temperature, both systems present a similar monotonically decreasing behaviour upon decreasing temperature. At the onset temperature of glassy dynamics ($T = 1\epsilon/k_{\mathrm{B}}$), [3] the data for the KALJ system shows an inflexion point which signals density-density correlations visible for distances longer than $r = 2.5\sigma$. Individual components of $\lim_{k\to 0} S_{\rho\rho}(k)$: (Middle panel) $\kappa_T \rho K_{\mathrm{B}} T$ and (Lower Panel) $\delta^2 \lim_{k\to 0} S_{cc}(k)$ with $\delta = \rho(v_{\mathrm{a}} - v_{\mathrm{b}})$ the product of the total density with the difference in partial molar volumes. It is apparent that the contrast in $S_{\rho\rho}$ originates from major concentration fluctuations present in the KALJ system, as indicated by $S_{cc}(k)$.

where $\eta = \rho_{\mathrm{a}} + \rho_{\mathrm{b}} + \rho_{\mathrm{a}}\rho_b(G^{\infty}_{\mathrm{aa}} + G^{\infty}_{\mathrm{bb}} - 2G^{\infty}_{\mathrm{ab}})$.

We use the definition in Eqs (II.7) and (II.8) to compute $S_{\rho\rho}(k)$, and compare with the $\lim_{k\to 0} S_{\rho\rho}(k)$ obtained from the KBIs (Eq. (II.10)). The results, presented in Figure II.3 (Top panel), confirm the information given by the partial structure factors (See Figure S2 for a com-

parison between the values obtained from the structure factor and the KBIs). Namely, in contrast to the KAWCA system, the KALJ system exhibits increasingly large density fluctuations upon decreasing temperature. To investigate the origin of the anomaly, we investigate the contributions to $S_{\rho\rho}$ separately as given by the r.h.s. of Eq. (II.10). The middle and lower panels of Figure II.3 splits $S_{\rho\rho}$ into isothermal compressibility and concentration fluctuation terms, respectively. There, it is apparent that the anomalous behaviour exhibited by the KALJ system at low k values is due to the formation of long-range concentration domains (red and blue triangles). By contrast, the isothermal compressibility contribution remains nearly the same for both systems (red and blue circles).

The anomaly observed in the limit $k \to 0$ in Figure II.2 has also been reported in Ref. [26]. The authors suggest a plausible explanation involving the gas-liquid phase separation of the KALJ system. However, we note that the gas-liquid coexistence region for the KALJ system is still far from the point $\rho = 1.2/\sigma^3$, $T = 0.45\epsilon/k_{\mathrm{B}}$ [52, 58] (See also the discussion in Sec. IV in Ref. [26]). Moreover, we observe the non-monotonic behaviour starting just below the onset temperature of glassy dynamics $T = 1\epsilon/k_{\mathrm{B}}$, which is even farther away from the coexistence region. Moreover, the virial part of the pressure remains positive even at the lowest temperature considered for the KALJ model (See Fig. 1 in Ref. [8] and Figure S3). For the KAWCA model, as expected, the system's pressure is systematically higher than the KALJ pressure due to the absence of attractive interactions. In general, the positive pressure of our simulated state points already suggests that the anomalous behaviour is not due to gas-liquid coexistence.

Plots of the excess coordination number ($N_{\alpha\beta} = \rho_\beta G^{\infty}_{\alpha\beta}$), which gives the change in the number of

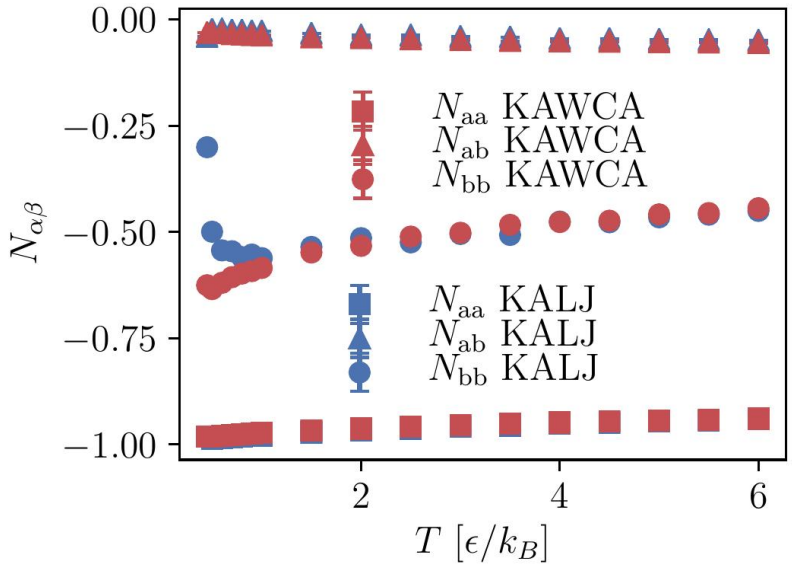

Figure II.4: Excess coordination number ($N_{\alpha\beta} = \rho_\beta G^{\infty}_{\alpha\beta}$) as a function of temperature for both KALJ and KAWCA systems. N_{ab} close to zero corresponds to a preferential a − b effective interaction. Below the onset temperature of glassy dynamics upon cooling, N_{bb} gets close to zero for the KALJ system, indicating a growing preferential b − b effective interaction, ultimately leading to phase segregation.

α particles when one β particle is removed from the system, as a function of temperature provide a clear insight (See Figure II.4). As expected from the model, the effective interaction between a and

b particles is favoured in both systems at all temperatures: excess coordination numbers are close to zero. Below the onset temperature of glassy dynamics, the excess coordination number shows a collective tendency for the KALJ mixtures to increase b − b effective interactions upon cooling. This tendency is not apparent for the a particles because they are the majority component in the mixture. More importantly, this demixing propensity is not observed in the KAWCA case. We underline here that these concentration domains for the KALJ system resemble the behaviour discovered by Fischer [13] for supercooled ortho-terphenyl. Namely, anomalies in the structure factor at low k-values, which are not commensurate with the isothermal compressibility, are connected to the nucleation of nanometric structures. [57] Furthermore, our results agree with recent theoretical efforts demonstrating that the low k portion of the structure factor for polydisperse colloidal systems can be separated into a compressibility contribution and a term related to composition fluctuations. [29]

We now focus on the isothermal compressibility (Eq. II.11). In Figure II.5, we present a log-log

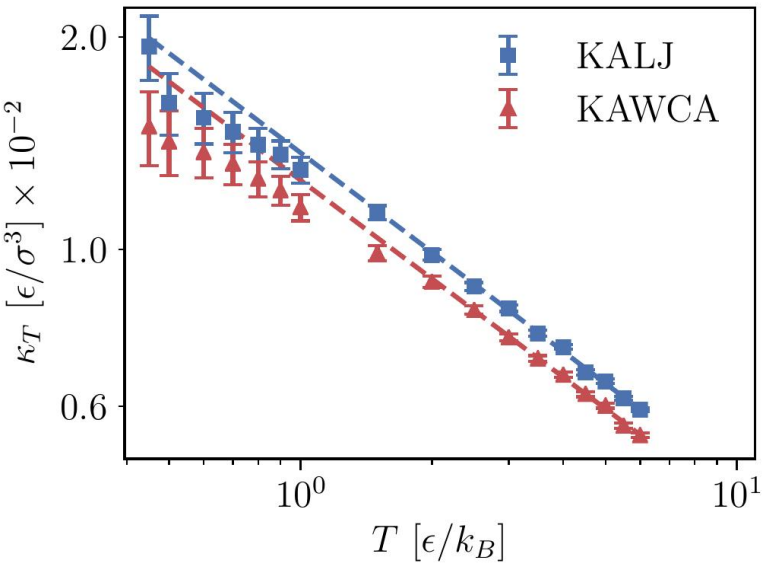

Figure II.5: Bulk isothermal compressibility κ_T, calculated from Eq. (II.11), as a function of temperature for KALJ and KAWCA systems (Log-log representation). We observe that a power law relationship holds as $\kappa_T = \kappa_T^0 T^{-\gamma}$ with $\gamma = 0.46 \pm 0.01$ and 0.45 ± 0.01 for KALJ and KAWCA, respectively. The dashed lines are the corresponding power law fitting.

plot of κ_T vs T, where it is apparent that the KALJ system is systematically more compressible than the KAWCA system at all temperatures considered here. Hence, it is again clear that small differences in the tail of the RDFs result in sizeable differences in their thermodynamic properties. Furthermore, a power-law behaviour $\kappa_T = \kappa_T^0 T^{-\gamma}$ is apparent with $\gamma = 0.46 \pm 0.01$ for the KALJ system and $\gamma = 0.45 \pm 0.01$ for the KAWCA system. Below the onset temperature of glassy dynamics, both systems deviate from this power law and become comparatively less compressible in the deeply supercooled region. One would expect that, for a system undergoing a gas-liquid separation, compressibility substantially increases upon approaching the gas-liquid region. By contrast, the behaviour observed in Figure II.5 suggests that this is not the case. Finally, the existence of this power law, including the low-temperature deviations [27], is somewhat similar ($\gamma = 0.40 \pm 0.01$) [56] to the one observed experimentally in supercooled water.

II.3.4 Chemical potential

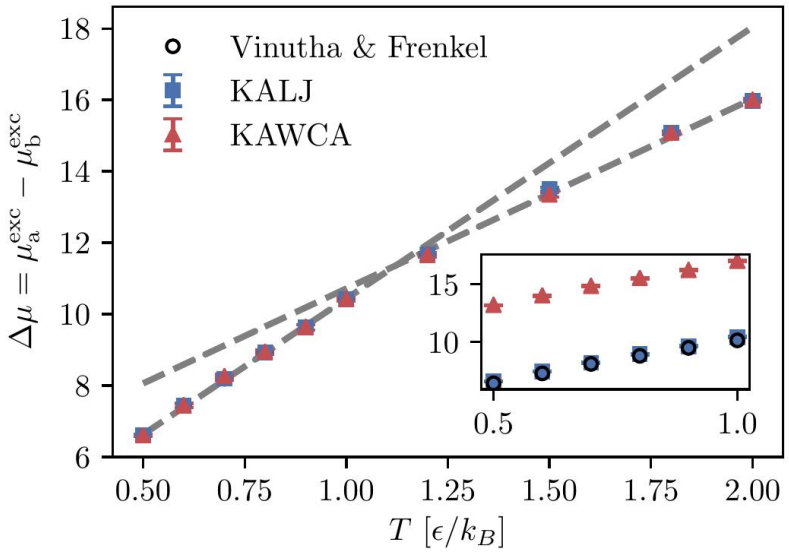

Figure II.6: Difference of excess chemical potentials between species a and b for both, KALJ and KAWCA systems. The KAWCA system results were shifted by a constant in order to mach the lowest temperature $T = 0.5\epsilon/k_B$, indicating that the potential energy can be approximated to $U_{LJ} \approx U_{WCA} + U_{Attractive}$. A change in the behaviour with T, indicated by the dashed-grey lines, is apparent at the onset temperature of glassy dynamics. The inner plot shows the difference of chemical potential for KALJ and KAWCA (without shifting). Results for the KALJ system in the temperature range $0.5 - 1.0\epsilon/k_B$ well compare with results available in the literature. [63]

Finally, we compute the excess chemical potential for both systems (Figure II.6) using the `SPARTIAN` method. [22] Recent calculations of the chemical potential for the KALJ system in the range of temperature $0.5\epsilon/k_B < T < 1.0\epsilon/k_B$ are in excellent agreement with our results. [63] At the onset temperature of glassy dynamics, there is a transition between two regimes, reflecting the tendency for the system to minimise its free energy. The fact that the curves for the KALJ and the KAWCA systems are identical up to a constant factor is a consequence of writing the LJ potential energy as $U_{LJ} \approx U_{WCA} + U_{Attractive}$. This expression lies at the foundation of perturbation theory that assumes that $U_{Attractive}$ is very small compared to U_{WCA}. However, the sizeable difference in chemical potential ($\approx 5\epsilon$) indicates that this approximation does not hold in this case.
Similarly to other thermodynamic properties like excess [2] and configurational entropy [7], isothermal compressibility and chemical potential results confirm that perturbation theory is not valid for the KALJ and KAWCA systems at $\rho = 1.2/\sigma^3$ since attractive interactions induce non-perturbative structural effects. In the following section, we investigate these systems at higher density, namely $\rho = 1.6/\sigma^3$, where we expect that repulsive interactions play an increasingly dominant role. [6, 7]

II.3.5 KALJ and KAWCA mixtures at $\rho = 1.6/\sigma^3$

We perform a similar thermodynamic analysis for KALJ and KAWCA systems at $\rho = 1.6/\sigma^3$. Our results show that density and concentration structure factors (Figure II.7) are nearly identical for

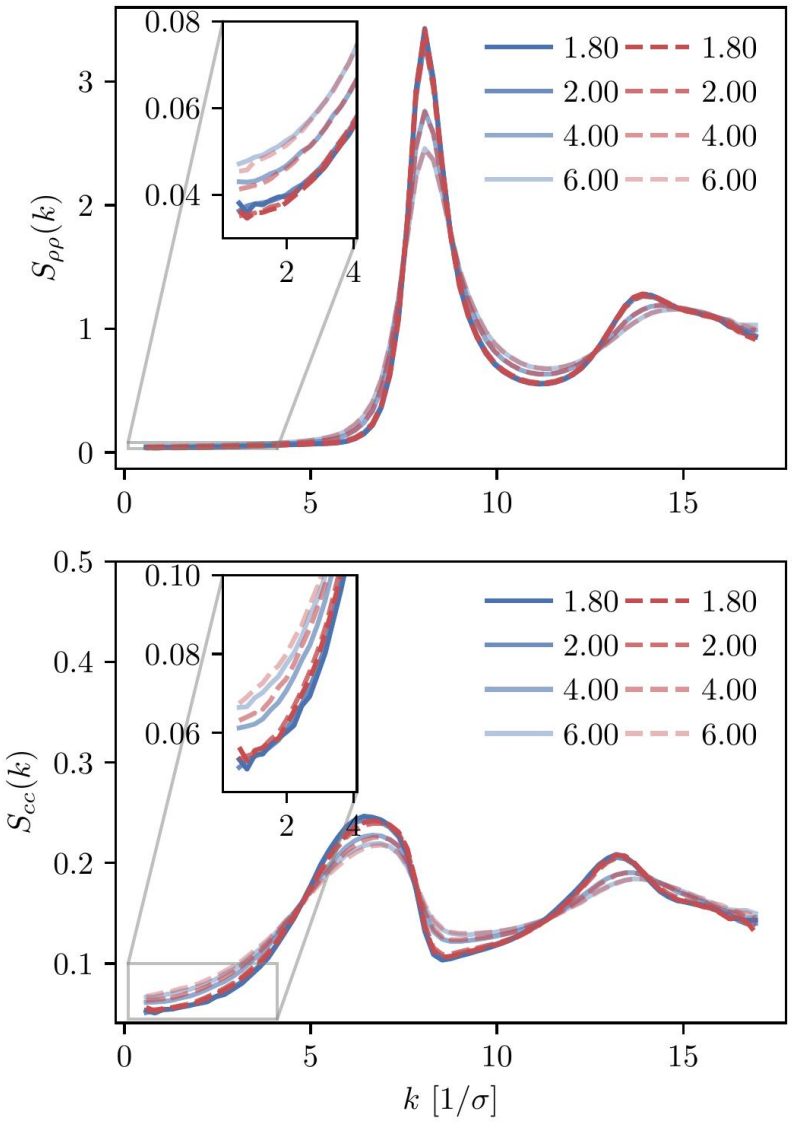

Figure II.7: Density and concentration structure factors, $S_{\rho\rho}$ and S_{cc}, for the KALJ and KAWCA systems at a higher density ($\rho = 1.6$), in the range of temperature considered here.

both systems in the range of temperature considered. Structure factors in the limit $k \to 0$, in particular, show no evidence for the nucleation of long-range structures. Dynamical properties for these mixtures available in the literature [8, 2, 5] reveal that both systems exhibit similar structural and dynamical properties at this density. Therefore, we conclude that long-range concentration fluctuations might be closely connected to the significant mismatch between dynamical properties of the two systems at $\rho = 1.2/\sigma^3$.

Concerning the isothermal compressibility (Figure II.8), the two systems are essentially indistinguishable in the whole temperature range. As a reference, the onset temperature of glassy dynamics for KALJ and KAWCA systems at this density is close to $2.80\epsilon/k_B$. [3] The two systems hence behave similarly well below the onset temperature of glassy dynamics. This result highlights the dominant role played by attractive interactions in determining thermodynamic properties of

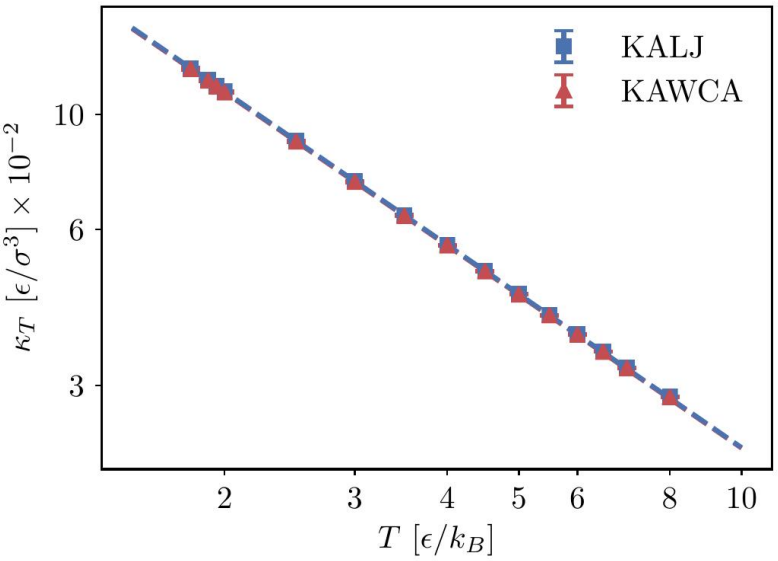

Figure II.8: Bulk isothermal compressibility κ_T for the KALJ and KAWCA systems at a higher density ($\rho = 1.6/\sigma^3$)

high-density liquids.

II.3.6 Crystallisation of the KAWCA system

In the last section, we investigate the crystallisation of the KAWCA system at $\rho = 1.2/\sigma^3$. We further decrease the temperature down to $T = 0.35\epsilon/k_\mathrm{B}$. Figure II.9 shows snapshots of the system at $T = 0.45\epsilon/k_\mathrm{B}$ (Top) and $T = 0.35\epsilon/k_\mathrm{B}$ (Bottom). It is apparent from the figure that the system at $T = 0.45\epsilon/k_\mathrm{B}$ appears like a miscible liquid. Conversely, the system at $T = 0.35\epsilon/k_\mathrm{B}$ shows crystalline domains with a clear tendency for phase-segregation. Density and concentration structure factors (Figure II.10) enable us to validate this crystallisation-demixing scenario. In particular, we observe the splitting of the second peak of $S_{\rho\rho}(k)$ at $1.7k_0$ with $k_0 \approx 7.13/\sigma$ that indicates the presence of facet-sharing domains between neighbouring crystalline regions. Perhaps more interesting, the structure factors at $k < 2/\sigma$ show a marked formation of long-range domains qualitatively similar to the ones present for the KALJ case below onset temperature.

These results indicate that the crystallisation of the KAWCA system is a process driven by phase segregation. We note that similar behaviour to that observed for the low-temperature density and concentration structure factors for KAWCA has also been observed in polydisperse glass-forming systems. [38] Furthermore, recent GPU simulations report the crystallisation of the KALJ system [26] due to composition fluctuations similar to the ones investigated in this work. Indeed, our results also support previous claims pointing out demixing as a precursor for crystalisation in the modified KA model [37] and in liquid metals. [17]

We conclude here that the presence of long-range concentration fluctuations is a qualitatively common feature for both KALJ and KAWCA systems. Perhaps more important, it is not directly related to gas-liquid coexistence present in the KALJ system. Indeed, in the crystalline state ($T = 0.35\epsilon/k_\mathrm{B}$), the KAWCA system exhibits a significant growth of concentration fluctuations, apparent in the lower panel in Figure II.10. The attractive interactions present in the KALJ sys-

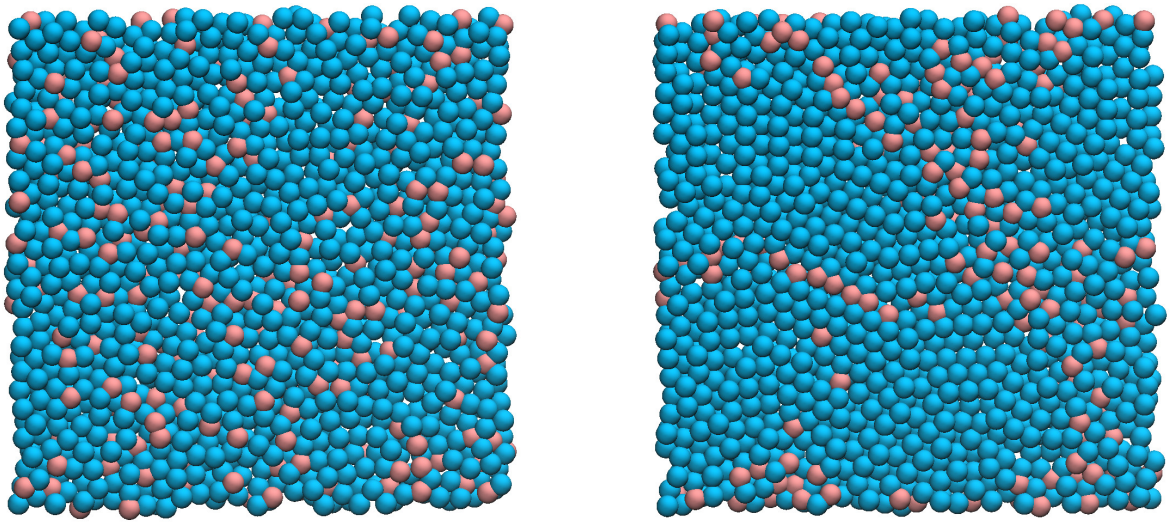

Figure II.9: Snapshot of the KAWCA system at $T = 0.45\epsilon/k_{\mathrm{B}}$ (Top) and $T = 0.35\epsilon/k_{\mathrm{B}}$ (Bottom).

tem favour the nucleation of concentration fluctuations starting at relatively high temperatures ($T = 1.00\epsilon/k_{\mathrm{B}}$). Although the KAWCA system crystallises in our simulation timescale as its dynamics are considerably faster than for the KALJ system, our results suggest that both systems crystallise upon demixing by following a similar pathway.

II.4 Conclusions

We compute various thermodynamic properties of model supercooled liquids, with (KALJ) and without (KAWCA) attractive interactions at density $\rho = 1.2/\sigma^3$. We aim at studying whether fluctuations in the tail of the two-body correlation function induce significant thermodynamic differences between the two systems. Density and concentration structure factors in the limit $k \to 0$ indicate that the KALJ system exhibits anomalous structural behaviour that we identify as the nucleation of long-range concentration domains. Conversely, the KAWCA system behaves like a normal liquid, with density and concentration structure factors decreasing monotonically. A finite-size Kirkwood-Buff analysis used to extrapolate to the $k \to 0$ limit confirms this picture. Differences in isothermal compressibilities and chemical potentials highlight the non-perturbative role of attractive interactions. Results of the crystallisation of the KAWCA system suggest that the anomaly, enhanced by the presence of attractive interactions, is a common feature of both models. All our results indicate that these long-range concentration fluctuations are not connected to gas-liquid coexistence, implying that demixing precedes crystallisation in both systems. Finally, upon increasing density ($\rho = 1.6/\sigma^3$), where KALJ and KAWCA systems show similar dynamical properties, the KALJ anomaly disappears, and both systems exhibit nearly identical thermodynamic properties. Hence, we speculate that there might be a connection between large-scale concentration fluctuations and the significant dynamical slow down of the KALJ system in the deeply supercooled regime.

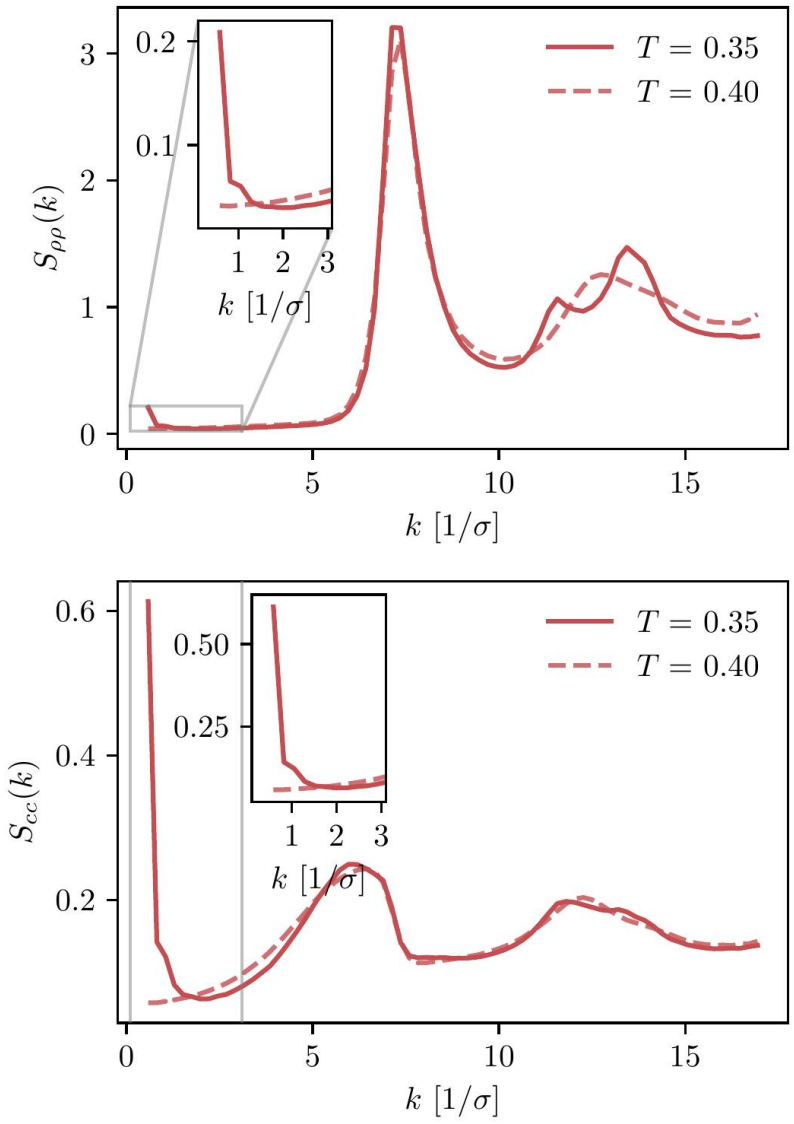

Figure II.10: Density and concentration structure factors, $S_{\rho\rho}$ and S_{cc}, for the KAWCA system in the temperature range $0.3\epsilon/k_{\mathrm{B}} < T < 0.45\epsilon/k_{\mathrm{B}}$.

Author Contributions

Atreyee Banerjee: Conceptualization, Methodology, Formal Analysis, Writing - Original Draft **Mauricio Sevilla**: Methodology, Software, Formal Analysis, Visualisation **Joseph F. Rudzinski**: Investigation, Writing - Review & Editing **Robinson Cortes-Huerto**: Conceptualisation, Methodology, Writing - Original Draft, Review & Editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank Kurt Kremer for his insightful discussions and his critical reading of the manuscript. They are also grateful to Pietro Ballone, Burkhard Dünweg, Smarajit Karmakar and Werner Steffen for their valuable feedback and suggestions. R.C.-H. thankfully acknowledge funding from SFB-TRR146 of the German Research Foundation (DFG).

Bibliography

[1] Gerold Adam and Julian H Gibbs. On the temperature dependence of cooperative relaxation properties in glass-forming liquids. *J. Chem. Phys.*, 43(1):139–146, 1965.

[2] Atreyee Banerjee, Manoj Kumar Nandi, Srikanth Sastry, and Sarika Maitra Bhattacharyya. Effect of total and pair configurational entropy in determining dynamics of supercooled liquids over a range of densities. *J. Chem. Phys.*, 145(3):034502, 2016.

[3] Atreyee Banerjee, Manoj Kumar Nandi, Srikanth Sastry, and Sarika Maitra Bhattacharyya. Determination of onset temperature from the entropy for fragile to strong liquids. *J. Chem. Phys.*, 147(2):024504, 2017.

[4] Atreyee Banerjee, Shiladitya Sengupta, Srikanth Sastry, and Sarika Maitra Bhattacharyya. Role of structure and entropy in determining differences in dynamics for glass formers with different interaction potentials. *Phys. Rev. Lett.*, 113(22):225701, 2014.

[5] Atreyee Banerjee and David J Wales. Fragility and correlated dynamics in supercooled liquids. *J. Chem. Phys.*, 153(12):124501, 2020.

[6] Ludovic Berthier and Gilles Tarjus. Nonperturbative effect of attractive forces in viscous liquids. *Phys. Rev. Lett.*, 103:170601, Oct 2009.

[7] Ludovic Berthier and Gilles Tarjus. Critical test of the mode-coupling theory of the glass transition. *Phys. Rev. E*, 82(3):031502, 2010.

[8] Ludovic Berthier and Gilles Tarjus. The role of attractive forces in viscous liquids. *J. Chem. Phys.*, 134(21):214503, 2011.

[9] A. B. Bhatia and D. E. Thornton. Structural aspects of the electrical resistivity of binary alloys. *Phys. Rev. B*, 2:3004–3012, Oct 1970.

[10] Emanuele Boattini, Susana Marín-Aguilar, Saheli Mitra, Giuseppe Foffi, Frank Smallenburg, and Laura Filion. Autonomously revealing hidden local structures in supercooled liquids. *Nat. Comm.*, 11(1):5479, 2020.

[11] Joyjit Chattoraj and Massimo Pica Ciamarra. Role of attractive forces in the relaxation dynamics of supercooled liquids. *Phys. Rev. Lett.*, 124:028001, Jan 2020.

[12] Massimo Pica Ciamarra, Raffaele Pastore, and Antonio Coniglio. Particle jumps in structural glasses. *Soft Matter*, 12(2):358–366, 2016.

[13] R. Cortes-Huerto, K. Kremer, and R. Potestio. Communication: Kirkwood-Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *J. Chem. Phys.*, 145(14):141103, OCT 14 2016.

[14] Daniele Coslovich. Static triplet correlations in glass-forming liquids: A molecular dynamics study. *J. Chem. Phys.*, 138(12):12A539, 2013.

[15] E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, and A. J. Liu. Identifying structural flow defects in disordered solids using machine-learning methods. *Phys. Rev. Lett.*, 114:108001, Mar 2015.

[16] Vanessa K. de Souza and David J. Wales. Energy landscapes for diffusion: Analysis of cage-breaking processes. *J. Chem. Phys*, 129(16):164507, 2008.

[17] Caroline Desgranges and Jerome Delhommelle. Unraveling the coupling between demixing and crystallization in mixtures. *Journal of the American Chemical Society*, 136(23):8145–8148, 2014.

[18] Caroline Desgranges and Jerome Delhommelle. Unusual crystallization behavior close to the glass transition. *Phys. Rev. Lett.*, 120:115701, Mar 2018.

[19] Denis J Evans and Brad Lee Holian. The nose–hoover thermostat. *J. Chem. Phys.*, 83(8):4069–4074, 1985.

[20] E.W. Fischer. Light scattering and dielectric studies on glass forming liquids. *Physica A: Statistical Mechanics and its Applications*, 201(1):183–206, 1993.

[21] W Götze and L Sjögren. The glass transition singularity. *Zeitschrift für Physik B Condensed Matter*, 65(4):415–427, 1987.

[22] Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Spatially resolved thermodynamic integration: An efficient method to compute chemical potentials of dense fluids. *J. Chem. Theo. Comp.*, 14(7):3409–3417, 2018.

[23] Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto. Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations. *Mol. Phys.*, 116(21-22):3301–3310, 2018.

[24] Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto. Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: Revisiting the spatial block analysis method. *Entropy*, 20(4):222, 2018.

[25] Glen M Hocky, Thomas E Markland, and David R Reichman. Growing point-to-set length scale correlates with growing relaxation times in model supercooled liquids. *Phys. Rev. Lett.*, 108(22):225506, 2012.

[26] Trond S. Ingebrigtsen, Jeppe C. Dyre, Thomas B. Schrøder, and C. Patrick Royall. Crystallization instability in glass-forming mixtures. *Phys. Rev. X*, 9:031016, Aug 2019.

[27] Kyung Hwan Kim, Alexander Späh, Harshad Pathak, Fivos Perakis, Daniel Mariedahl, Katrin Amann-Winkel, Jonas A Sellberg, Jae Hyuk Lee, Sangsoo Kim, Jaehyun Park, et al. Maxima in the thermodynamic response and correlation functions of deeply supercooled water. *Science*, 358(6370):1589–1593, 2017.

[28] John G Kirkwood and Frank P Buff. The statistical mechanical theory of solutions. i. *J. Chem. Phys.*, 19(6):774–777, 1951.

[29] L. Klochko, J. Baschnagel, J. P. Wittmer, O. Benzerara, C. Ruscher, and A. N. Semenov. Composition fluctuations in polydisperse liquids: Glasslike effects well above the glass transition. *Phys. Rev. E*, 102:042611, Oct 2020.

[30] W. Knoll and S. Steeb. *Z. Naturforsch.*, 33a:472–479, 1978.

[31] Walter Kob and Hans C Andersen. Testing mode-coupling theory for a supercooled binary lennard-jones mixture. ii. intermediate scattering function and dynamic susceptibility. *Phys. Rev. E*, 52(4):4134, 1995.

[32] Peter Krüger, Sondre K Schnell, Dick Bedeaux, Signe Kjelstrup, Thijs JH Vlugt, and Jean-Marc Simon. Kirkwood–buff integrals for finite volumes. *J. Phys. Chem. Lett.*, 4(2):235–238, 2013.

[33] P. Kumari, V. V. S. Pillai, D. Gobbo, P. Ballone, and A. Benedetto. The transition from salt-in-water to water-in-salt nanostructures in water solutions of organic ionic liquids relevant for biological applications. *Phys. Chem. Chem. Phys.*, 23:944–959, 2021.

[34] F. P. Landes, Giulio Biroli, Olivier Dauchot, Andrea J. Liu, and David R. Reichman. Attractive versus truncated repulsive supercooled liquids: The dynamics is encoded in the pair correlation function. *Phys. Rev. E*, 101:010602, Jan 2020.

[35] Mathieu Leocmach and Hajime Tanaka. Roles of icosahedral and crystal-like order in the hard spheres glass transition. *Nat. Comm.*, 3(1):974, 2012.

[36] Manoj Kumar Nandi and Sarika Maitra Bhattacharyya. Microscopic theory of softness in supercooled liquids. *Phys. Rev. Lett.*, 126(20):208001, 2021.

[37] Ujjwal Kumar Nandi, Atreyee Banerjee, Suman Chakrabarty, and Sarika Maitra Bhattacharyya. Composition dependence of the glass forming ability in binary mixtures: The role of demixing entropy. *The Journal of chemical physics*, 145(3):034503, 2016.

[38] Andrea Ninarello, Ludovic Berthier, and Daniele Coslovich. Models and algorithms for the next generation of glass transition studies. *Phys. Rev. X*, 7:021039, Jun 2017.

[39] A. Patkowski, Th. Thurn-Albrecht, E. Banachowicz, W. Steffen, P. Bösecke, T. Narayanan, and E. W. Fischer. Long-range density fluctuations in orthoterphenyl as studied by means of ultrasmall-angle x-ray scattering. *Phys. Rev. E*, 61:6909–6913, Jun 2000.

[40] Ulf R Pedersen, Thomas B Schrøder, and Jeppe C Dyre. Repulsive reference potential reproducing the dynamics of a liquid with attractions. *Phys. Rev. Lett.*, 105(15):157801, 2010.

[41] S. J. Plimpton. *J. Comput. Phys.*, **117**:1, 1995.

[42] Raffaello Potestio, Pep Español, Rafael Delgado-Buscalioni, Ralf Everaers, Kurt Kremer, and Davide Donadio. Monte carlo adaptive resolution simulation of multicomponent molecular liquids. *Phys. Rev. Lett.*, 111:060601, Aug 2013.

[43] Raffaello Potestio, Sebastian Fritsch, Pep Español, Rafael Delgado-Buscalioni, Kurt Kremer, Ralf Everaers, and Davide Donadio. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.*, 110:108301, Mar 2013.

[44] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.*, 123(22):224106–14, 2005.

[45] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids. *Phys. Rev. E*, 73:066701, 2006.

[46] M. Praprotnik, L. Delle Site, and K. Kremer. A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation. *J. Chem. Phys.*, 126:134902, 2007.

[47] M. Praprotnik, L. Delle Site, and K. Kremer. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Ann. Rev. Phys. Chem.*, 59(1):545–571, 2008.

[48] F L Román, J A White, and S Velasco. Fluctuations in an equilibrium hard-disk fluid: Explicit size effects. *J. Chem. Phys.*, 107:4635, 1997.

[49] M Rovere, D W Heermann, and K Binder. The gas-liquid transition of the two-dimensional lennard-jones fluid. *J. Phys.: Condens. Matter*, 2(33):7009, 1990.

[50] Joseph F Rudzinski, Marc Radu, and Tristan Bereau. Automated detection of many-particle solvation states for accurate characterizations of diffusion kinetics. *J. Chem. Phys.*, 150:024102, 2019.

[51] Philip S Salmon, Richard A Martin, Philip E Mason, and Gabriel J Cuello. Topological versus chemical ordering in network glasses at intermediate and extended length scales. *Nature*, 435(7038):75–78, 2005.

[52] Srikanth Sastry. Liquid limits: Glass transition and liquid-gas spinodal boundaries of metastable liquids. *Physical Review Letters*, 85(3):590, 2000.

[53] Sondre K Schnell, Thijs JH Vlugt, Jean-Marc Simon, Dick Bedeaux, and Signe Kjelstrup. Thermodynamics of a small system in a μt reservoir. *Chem. Phys. Lett.*, 504(4-6):199–201, 2011.

[54] Francesco Sciortino and Walter Kob. Debye-waller factor of liquid silica: Theory and simulation. *Physical review letters*, 86(4):648, 2001.

[55] Felix Sedlmeier, Dominik Horinek, and Roland R Netz. Spatial correlations of density and structural fluctuations in liquid water: A comparative simulation study. *J. Am. Chem. Soc,*, 133(5):1391–1398, 2011.

[56] Alexander Späh, Harshad Pathak, Kyung Hwan Kim, Fivos Perakis, Daniel Mariedahl, Katrin Amann-Winkel, Jonas A. Sellberg, Jae Hyuk Lee, Sangsoo Kim, Jaehyun Park, Ki Hyun Nam, Tetsuo Katayama, and Anders Nilsson. Apparent power-law behavior of water's isothermal

compressibility and correlation length upon supercooling. *Phys. Chem. Chem. Phys.*, 21:26–31, 2019.

[57] Jacob D. Stevenson and Peter G. Wolynes. The ultimate fate of supercooled liquids. *J. Phys. Chem. A*, 115(16):3713–3719, 2011.

[58] Vincent Testard, Ludovic Berthier, and Walter Kob. Influence of the glass transition on the liquid-gas spinodal decomposition. *Physical review letters*, 106(12):125702, 2011.

[59] Hua Tong and Hajime Tanaka. Role of attractive interactions in structure ordering and dynamics of glass-forming liquids. *Phys. Rev. Lett.*, 124:225501, Jun 2020.

[60] Søren Toxvaerd. Role of the attractive forces in a supercooled liquid. *Phys. Rev. E*, 103:022611, Feb 2021.

[61] Søren Toxvaerd and Jeppe C. Dyre. Communication: Shifted forces in molecular dynamics. *J. Chem. Phys.*, 134(8):081102, 2011.

[62] Benjamin W. van de Waal. On the origin of second-peak splitting in the static structure factor of metallic glasses. *Journal of Non-Crystalline Solids*, 189(1):118–128, 1995.

[63] HA Vinutha and Daan Frenkel. Computation of the chemical potential and solubility of amorphous solids. *J. Chem. Phys.*, 154(12):124502, 2021.

[64] John D Weeks, David Chandler, and Hans C Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.*, 54(12):5237–5247, 1971.

[65] Zhen Zhang and Walter Kob. Revealing the three-dimensional structure of liquids using four-point correlation functions. *Proc. Nat. Aca. Sci.*, 117(25):14032–14037, 2020.

III. Connecting density fluctuations and Kirkwood-Buff integrals for finite-size systems

Outline

Bibliographic Information

Mauricio Sevilla and Robinson Cortes-Huerto. Connecting density fluctuations and Kirkwood–Buff integrals for finite-size systems. *The Journal of Chemical Physics*, 156(4):044502, 01 2022.

Author contribution

§ Mauricio Sevilla: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Software (lead); Writing – original draft (equal); Writing – review & editing (equal).

§ Robinson Cortes-Huerto: Conceptualization (lead); Formal analysis (equal); Investigation (equal); Project administration (lead); Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal).

Abstract

Kirkwood-Buff integrals (KBI) connect the microscopic structure and thermodynamic properties of liquid solutions. KBI are defined in the grand canonical ensemble and evaluated assuming the thermodynamic limit (TL). In order to reconcile analytical and numerical approaches, finite-size KBI have been proposed in the literature, resulting in two strategies to obtain their TL values from computer simulations. (i) The spatial block-analysis method in which the simulation box is divided into subdomains of volume V to compute fluctuations of the number of particles. (ii) A direct integration method where a corrected radial distribution function and a kernel that accounts for the geometry of the integration subvolumes are combined to obtain KBI as a function of V. In this work, we propose a method that connects both strategies into a single framework. We start from the definition of finite-size KBI, including the integration subdomain and an asymptotic correction to the radial distribution function, and solve them in Fourier space where periodic boundary conditions are trivially introduced. The limit $q \to 0$, equivalent to the value of the KBI in the TL, is obtained via the spatial block-analysis method. When compared to the latter, our approach gives nearly identical results for all values of V. Moreover, all finite-size effect contributions (ensemble, finite-integration domains and periodic boundary conditions) are easily identifiable in the calculation. This feature allows us to analyse finite-size effects independently and extrapolate the results of a single simulation to different box sizes. To validate our approach, we investigate prototypical systems, including SPC/E water and aqueous urea mixtures.

III.1 Introduction

Kirkwood-Buff integrals (KBI) connect the microscopic structure of a liquid solution, via integrals of the radial distribution functions (RDF), and its thermodynamic properties, as obtained from fluctuations of the number of particles in subvolumes of the total system [26]. This connection between local structure and thermodynamics is particularly useful in computational soft-matter studies where KBI are widely used to evaluate isothermal compressibilities, partial molar volumes and derivatives of chemical potentials [7, 26, 25, 8]. In particular, applications of KBI include the investigation of the thermodynamics of complex molecular mixtures [39, 38, 16, 40, 44, 35, 11], solvation of macromolecules [25, 45, 57, 37, 42, 41, 59, 43], multicomponent diffusion in liquids [27, 33], protein self-assembly and aggregration [4, 20], Hofmeister ion chemistry [9], identification of nanostructures in water solutions of ionic liquids [31] and the parameterisation of atomistic [19, 15, 54, 34] and coarse-grained [17, 14] force fields. Recently, KBI have been applied to compute isothermal compressibilities of prototypical crystals [29, 36], showing unprecedented flexibility and range of applicability.

KBI are strictly defined in the grand canonical ensemble. Moreover, in practice, it is usual to take the thermodynamic limit (TL) to reduce their calculation to spherically symmetric real-space integrals of the radial distribution functions. In computer simulations, the TL is approximated by introducing periodic boundary conditions (PBC) for a system with a fixed number of particles N_0. Accordingly, finite subvolumes V with an average number of particles $\langle N \rangle$ are used to compute fluctuations of the number of particles and radial distribution functions [28, 58]. Periodicity, different thermodynamic ensembles and finite integration domains introduce artefacts in the resulting KBI.

Finite-size KBI have been proposed in the literature to bridge the existing gap between analytical expressions and numerical studies. The critical assumption is that fluctuations of the number of particles in subvolumes V inside the simulation box are equal to integrals of the corresponding closed-system radial distribution functions [53, 51, 47]. This equality provides two routes to obtain KBI in the TL. The first one, i.e. the spatial block analysis method (SBA), is based on calculating fluctuations of the number of particles in subdomains of volume V. By using arguments from thermodynamics of small systems [28], linear scaling relations are defined to extrapolate KBI in the TL [56, 55]. The second possibility is to correct the radial distribution functions for the differences in the thermodynamic ensemble, then integrate them using a kernel that takes into account finite-size domains [36, 18]. Naturally, the limit $V \to \infty$ gives the KBI in the TL.

Indeed, the two approaches are connected. In the limit $V > V_\zeta$, with V_ζ the volume defined by the correlation length of the system ζ, the integration of the radial distribution functions give an expression equivalent to the result obtained from linear scaling of fluctuations of the number of particles, including ensemble and finite integration domain effects [18]. Nevertheless, the local solvation structure information, provided by the short-range part of the RDF, is lost in this case. Moreover, the effect of periodic boundary conditions is not included in the final result.

In this work, we propose a method that connects the spatial block analysis method to the direct integration of exact, finite-size KBI. We evaluate the large r limit by introducing an asymptotic correction to the RDF. By defining the geometry of the subdomain, we write and solve KBI in Fourier space where the periodicity of the cell can also be incorporated, following the procedure proposed in Ref.
citech3Roman1999. We compute the $q \to 0$ limit by using the spatial block analysis method. We thus obtain KBI as a function of the volume of the subdomain and find excellent agreement with fluctuations of the number of particles for SPC/E water and aqueous urea mixtures for all values of V. The method is accurate, and its implementation is straightforward. It simply requires performing a spatial block analysis and calculating one-dimensional integrals of partial structure factors.

The paper is organised as follows: In Section III.2 we introduce the method, and in Section III.3 the computational details. We present the main results in Section III.4 and conclude in Section III.5.

III.2 Kirkwood-Buff integrals for finite-size systems

For a multicomponent fluid of species i, j, contained in a volume $V = L^3$, in thermal and chemical equilibrium with an infinite reservoir of particles, the Kirkwood-Buff integral (KBI) is defined as [26]

$$G_{ij} = V\left(\frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\langle N_i\rangle\langle N_j\rangle} - \frac{\delta_{ij}}{\langle N_i\rangle}\right) = \frac{1}{V}\int_V\int_V d\mathbf{r}_1 d\mathbf{r}_2\left[g_{ij}\left(\mathbf{r}\right) - 1\right], \tag{III.1}$$

where N_i is the number of particles of the i-species and the bracket $\langle\cdots\rangle$ denotes a thermal average, δ_{ij} is the Kronecker delta and g_{ij} is the pair correlation function defined in the grand canonical ensemble with $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$.
In computer simulations, we usually investigate systems with fixed number of particles N_0 with volume $V_0 = L_0^3$. Building on similar results for the Ornstein-Zernike equation [49], we define the

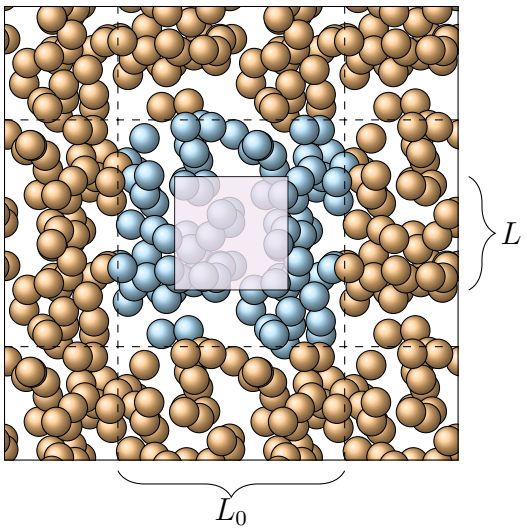

Figure III.1: Schematic representation of the spatial block analysis method. The N_0 blue particles represent the system with linear size L_0, and the red particles represent the periodic images. The purple box is a subvolume of linear size $L < L_0$ defined to compute fluctuations of the number of particles.

finite-size KBI as

$$\begin{aligned} G_{ij}(V;V_0) &= V\left(\frac{\langle N_iN_j\rangle' - \langle N_i\rangle'\langle N_j\rangle'}{\langle N_i\rangle'\langle N_j\rangle'} - \frac{\delta_{ij}}{\langle N_i\rangle'}\right) \\ &= \frac{1}{V}\int\int d\mathbf{r}_1 d\mathbf{r}_2\, R(\mathbf{r}_1)\, R(\mathbf{r}_2)\left[g_{ij}(\mathbf{r};V_0) - 1\right], \end{aligned} \tag{III.2}$$

where the average $\langle\cdots\rangle' \equiv \langle\cdots\rangle_{V,V_0}$ now explicitly depends on the subdomain and total volumes, V and V_0, respectively (See Fig. III.1). Here we focus on the integral term, that contains the radial distribution function of the closed system, $g_{ij}(\mathbf{r};V_0)$, and a step function $R(\mathbf{r})$ that defines the integration subdomain: it is one inside and zero outside the volume V. By defining Eq. (III.2), we connect explicitly density fluctuations and the integral of the pair correlation function for any subdomain V.
In the following, we focus on integrating the r.h.s. of Eq. (III.2). That is

$$G_{ij}(V;V_0) = \frac{1}{V}\int\int d\mathbf{r}_1 d\mathbf{r}_2\, R(\mathbf{r}_1)\, R(\mathbf{r}_2)\, h_{ij}(\mathbf{r};V_0)\,, \tag{III.3}$$

with $h_{ij}(\mathbf{r};V_0) = g_{ij}(\mathbf{r};V_0) - 1$.
To include the correction due to ensemble effects, we use the approximation proposed in Ref. [18]

$$g_{ij}(\mathbf{r};V_0) = g_{ij}(\mathbf{r}) - \frac{1}{V_0}\left(\frac{\delta_{ij}}{\rho_i} + G_{ij}^{\infty}\right), \tag{III.4}$$

based on the asymptotic limit $g_{ij}(r\to\infty;V_0) = 1 - (\delta_{ij}/\rho_i + G_{ij}^{\infty})/V_0$ discussed in Ref. citech3ben-naim. This implies that

$$h_{ij}(\mathbf{r};V_0) = h_{ij}(\mathbf{r}) - \frac{1}{V_0}\left(\frac{\delta_{ij}}{\rho_i} + G_{ij}^{\infty}\right), \tag{III.5}$$

thus, the finite-size KBI becomes

$$G_{ij}(V;V_0) = G_{ij}(V) - \frac{V}{V_0}\left(\frac{\delta_{ij}}{\rho_i} + G_{ij}^{\infty}\right) , \tag{III.6}$$

where the second term on the r.h.s. contains the correction due to ensemble effects [41, 18, 26, 25] and

$$G_{ij}(V) = \frac{1}{V}\int\int d\mathbf{r}_1 d\mathbf{r}_2\, R(\mathbf{r}_1)\, R(\mathbf{r}_2)\, h_{ij}(\mathbf{r})\,. \tag{III.7}$$

This expression can be easily written in Fourier space

$$G_{ij}(V) = \frac{1}{(2\pi)^3 V}\int d\mathbf{k}\, \tilde{R}(\mathbf{k})\, \tilde{R}(-\mathbf{k})\, \tilde{h}_{ij}\,(\mathbf{k})\ , \tag{III.8}$$

where $\tilde{h}_{ij}$ is the Fourier transform of h_{ij}. An additional advantage of integrating in reciprocal space is that periodic boundary conditions can be considered explicitly [49, 57]. It is enough to rewrite $\tilde{h}_{ij}(\mathbf{k})$ such that periodic copies of the system are included via a phase factor. That is, we include the complete contribution of the periodic boundary conditions into Eq. (III.8) as

$$G_{ij}(V) = \frac{1}{(2\pi)^3 V}\int d\mathbf{k}\, \tilde{R}(\mathbf{k})\, \tilde{R}(-\mathbf{k})\, \tilde{h}_{ij}^{\mathrm{PBC}}(\mathbf{k})\,, \tag{III.9}$$

where [49]

$$\tilde{h}_{ij}^{\mathrm{PBC}}(\mathbf{k}) = \sum_{n_x,n_y,n_z} e^{-\mathbf{k}\cdot\mathbf{s}_{n_x,n_y,n_z}}\, \tilde{h}_{ij}(\mathbf{k})\,, \tag{III.10}$$

with $\mathbf{s}_{n_x,n_y,n_z} = (n_x\, L_0, n_y\, L_0, n_z\, L_0)$ a vector specifying the system's periodic images such that $n_{x,y,z}$ takes integer values. In the following, we find that the choice $n_x = n_y = n_z = 1$ is sufficient to compute Eq. (VI.3) accurately.

We assume a homogeneous fluid such that $\tilde{h}_{ij}(\mathbf{k}) = \tilde{h}_{ij}(k)$ with $k = \sqrt{\mathbf{k}\cdot\mathbf{k}}$. Hence, in practice, we use the relation between $\tilde{h}_{ij}(k)$ and the partial structure factors S_{ij} [1, 21], namely

$$S_{ij}(k) = \delta_{ij} + \tilde{h}_{ij}(k)\,. \tag{III.11}$$

The partial structure factors are computed as

$$S_{ij}(\mathbf{k}) = \left\langle \frac{1}{\sqrt{N_i N_j}} \sum_{i'=1}^{N_i}\sum_{j'=1}^{N_j} \exp\left(-i\mathbf{k}\cdot(\mathbf{r}_{i'} - \mathbf{r}_{j'})\right)\right\rangle . \tag{III.12}$$

Consequently, the problem reduces to evaluate a single integral of the partial structure factors given by Eq. (VI.3).

In principle, Eq. (III.6) now includes all the finite size effects present in the simulation (finite boundary, periodicity of the box and ensemble). Before entering into applications, there are still two issues demanding our immediate attention. The first is that the asymptotic correction in Eq. (III.6) requires the value of G_{ij}^{∞}. The second concerns the evaluation of $\lim_{k\to 0} S_{ij}(k)$, that reduces, again, to evaluate G_{ij}^{∞}. Indeed, we have

$$\lim_{k\to 0} S_{ij}(k) = \delta_{ij} + \rho_i\, G_{ij}^{\infty}\,. \tag{III.13}$$

To obtain G_{ij}^{∞} we recall that, in the limit $V_\zeta < V < V_0$ (grand canonical ensemble), Eq. (III.7) can be approximated to $G_{ij}(V) \approx G_{ij}^{\infty} + \alpha_{ij}/V^{1/3}$ [53, 51, 56, 36] where α_{ij} is a constant. By including this approximation into Eq. (III.6), we recover the spatial block analysis (SBA) result consistent with the result reported in Ref. [18]

$$G_{ij}^{\mathrm{SBA}}(V;V_0) = G_{ij}^{\infty}\left(1-\frac{V}{V_0}\right) - \frac{V}{V_0}\frac{\delta_{ij}}{\rho_i} + \frac{\alpha_{ij}}{V^{1/3}}\,. \tag{III.14}$$

By evaluating density fluctuations for volumes $V \leq V_0$, as defined by the left hand side of Eq. (III.2), it is thus possible to extrapolate G_{ij}^{∞} [56, 18].

To summarise, the present method to evaluate KBI for finite systems requires information readily accessible from the simulation trajectory: density fluctuations for subvolumes $V \leq V_0$ and partial structure factors. Additional corrections to the RDF or finite-domain integration kernels are not required. Moreover, periodic boundary effects are trivially included in the calculation.

III.3 Computational details

To validate our approach, we first focus on liquid SPC/E water [5, 13, 62]. Molecular dynamics simulations have been carried out with GROMACS 4.5.1 [46] for systems containing 1000 and 8000 water molecules. We started with systems of initial density ≈ 26 waters/nm^3 (≈ 776 kg/m^3) that were optimised using steepest descent minimisation (50000 steps are sufficient). An equilibration run of 3.5 ns was carried out in the NPT ensemble at 1 bar. Next, we alternated 3.5 ns (time step = 1 fs) constant pressure (NPT) at P=1 bar and constant volume (NVT) simulations at T=300 K. For NPT simulations we used the Berendsen barostat [6], and for NVT simulations temperature was enforced by a velocity rescaling thermostat [10]. We continued with this protocol until we verified that in the NPT ensemble the density is 33.5 waters/nm^3 (1000 kg/m^3) and that in the NVT simulation pressure fluctuates around the 1 bar value. The last NVT trajectory obtained after this sequence of NPT–NVT equilibration runs was used for the spatial block analysis and for the calculation of the structure factor.

To test the method with a multicomponent case, we have re-used our simulation trajectories of aqueous urea solution [14, 25] using the Kirkwood-Buff derived force field [61] and SPC/E water [5, 13, 62] in GROMACS 4.5.1 [46] with a relatively small size of the simulation box ($L \sim 8$ nm). We have considered four more molar concentrations for a total of seven molar concentrations: 2.00, 3.06, 3.90, 5.07, 6.03, 7.10 and 8.03 M. Hence, we have alternated 3.5 ns (time step = 1 fs) constant pressure (NPT) at P=1 bar and constant volume (NVT) simulations at T=300 K. For NPT simulations we used a Berendsen barostat [6] to control the pressure, and for NVT simulations a velocity rescaling thermostat [10] to enforce the target temperature. We continued with this protocol (38 NPT–NVT cycles) until we verified that in the NVT simulation pressure fluctuates around 1 bar. Also in this case, the last NVT trajectory obtained after this NPT–NVT equilibration sequence was used for the spatial block analysis and for the calculation of the partial structure factors.

III.4 Results

III.4.1 Single component liquid: SPC/E water

For the single-component liquid, we focus on the Ornstein-Zernicke integral equation for finite size systems [7, 53, 45, 40]. For a closed system with fixed number of particles N_0 and volume V_0, including PBC. Similar to Eq. III.2, we define [49, 41]

$$\begin{aligned}\chi_T(V;V_0) &= \frac{\langle N^2\rangle' - \langle N\rangle'^2}{\langle N\rangle'} \\ &= 1 + \frac{\rho}{V}\int_V\int_V d\mathbf{r}_1 d\mathbf{r}_2 R(\mathbf{r}_1)R(\mathbf{r}_2)[g(\mathbf{r};V_0)-1]\,,\end{aligned} \tag{III.15}$$

and in this case we use the asymptotic correction to the RDF proposed in Refs [30, 45]

$$g(\mathbf{r};V_0) = g(\mathbf{r}) - \frac{\chi_T^\infty}{N_0}\,, \tag{III.16}$$

by neglecting $O(1/N_0^2)$ contributions. $\chi_T^\infty = \rho k_\mathrm{B} T \kappa_T$, κ_T being the isothermal compressibility of the bulk system. To solve the integral on the r.h.s. of Eq. VI.1, we use the same procedure as described in Section III.2. In cases where $V_\zeta < V < V_0$, we obtain the equivalent spatial block analysis expression [26, 25]

$$\chi_T^\mathrm{SBA}(V;V_0) = \chi_T^\infty\left(1-\frac{V}{V_0}\right) + \frac{\rho\alpha}{V^{1/3}}\,, \tag{III.17}$$

with α a constant.

First, we compute density fluctuations as defined on the l.h.s. of Eq. VI.1 for both $N_0 = 1000$ and 8000 systems. By defining $\lambda = (V/V_0)^{1/3}$, we plot $\lambda\chi_T$ as a function of λ (Fig. III.2). We extrapolate χ_T^∞ from the curve's slope in the region $\lambda < 0.3$ for the system with $N_0 = 8000$ water molecules. The choice of this linear region is motivated by the fact that $\lambda\chi_T$ as obtained from Eq. VI.1 has the maximum at $\lambda_\mathrm{max} = 0.63$. Thus, we estimated $\lambda = 0.3$ as the value where the curve starts deviating significantly from a straight line. We can also use $\lambda = 0.3$ to choose an appropriate size of the system for the spatial block analysis. By assuming that the correlation length of water is $\zeta = 1.5$ nm, we define $V_\zeta^{1/3} = 1.5 \times (4\pi/3)^{1/3}$ nm. The simulation box with $N = 8000$ water molecules has $V_0^{1/3} = 6.2$ nm, thus $\lambda = (V_\zeta/V_0)^{1/3} = 0.39$. This value is larger than $\lambda = 0.3$, still, it is sufficient to obtain a value of $\chi_T^\infty = 0.062$, in good agreement with the results reported in Ref. [25]. In practice, to select the size of the system one can start by estimating the correlation length from the radial distribution function and evaluating the linear size of the box such that $\lambda \approx 0.3$. This criteria can also be applied to binary mixtures.

We use the results of this linear fit to plot the SBA results, Eq. (III.17). Also for this system, we compute the structure factor, and correct for the $\lim_{k\to 0}$ by using the relation

$$\lim_{k\to 0} S(k) = \chi_T^\infty\,. \tag{III.18}$$

We thus compute an integral equivalent to Eq. (III.8) to obtain $\chi_T(V;V_0)$. The results for both systems are also presented in Fig. III.2. It is apparent that the agreement between density fluctuations and the integral method presented here is excellent. In contrast to the spatial block analysis

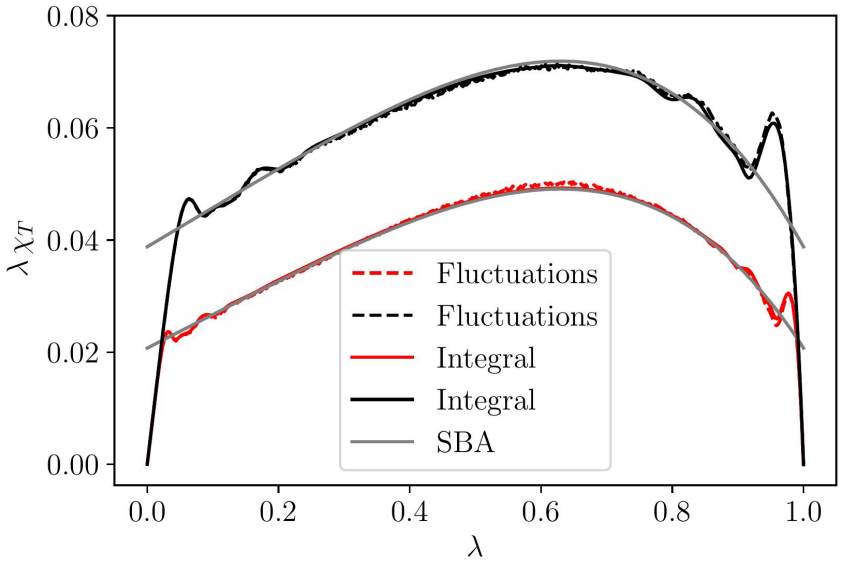

Figure III.2: Plots of the normalised finite-size isothermal compressibility $\lambda\chi_T$ as a function of $\lambda = (V/V_0)^{1/3}$ for systems with $N_0 = 1000$ (black) and 8000 (red) water molecules. Dashed lines correspond to density fluctuations and solid lines represent the method presented here. Solid grey lines correspond to the fitting of Eq. (III.17)

result (Eq. (III.17)), oscillations of $\lambda\chi_T$ at low values of λ, related to the local liquid structure, and at large values of λ, due to the periodicity of the simulation box, are consistently reproduced with our method. This is particularly interesting for the system with $N_0 = 1000$ water molecules, where oscillations are more pronounced. In this case, our integral method uses information from the system with $N_0 = 8000$ water molecules. The small box behaviour is reproduced *artificially* via the periodic images in Eq. (VI.2).

We now focus on the different finite-size effects present in the system with $N_0 = 1000$ water molecules. In Fig. III.3 we present four possibilities of evaluating the r.h.s. of Eq. (VI.1). (i) For the closed system, i.e. including the correction $\chi_T^\infty V/V_0$, with PBC, we observe that $\chi_T(\lambda = 1) = 0$, as expected. (ii) The closed system without PBC gives a limit $\chi_T(\lambda = 1) = \rho\alpha/V_0^{1/3}$ consistent with Eq. (III.17). (iii) An open system can be obtained by neglecting the correction $\chi_T^\infty V/V_0$. Moreover, by including PBC, we obtain $\chi_T(\lambda = 1) = \chi_T^\infty = 0.062$, precisely the thermodynamic limit value. (iv) For an open system without PBC $\chi_T(\lambda = 1) = \chi_T^\infty + \rho\alpha/V_0^{1/3}$, again, consistent with Eq. (III.17). These results thus highlight the role of PBC in enforcing the correct behaviour at the boundary of open and closed molecular systems.

III.4.2 Binary mixture: aqueous urea solution

We perform a similar analysis for the aqueous urea mixture case. First, we compute fluctuations of the number of particles as defined on the l.h.s. of Eq. (III.2). As for the single component case, we define $\lambda = (V/V_0)^3$ and plot λG_{ij} as a function of λ. We carried out this study for all concentrations. However, we only present the results for the case 8M in Fig. III.4 (dashed

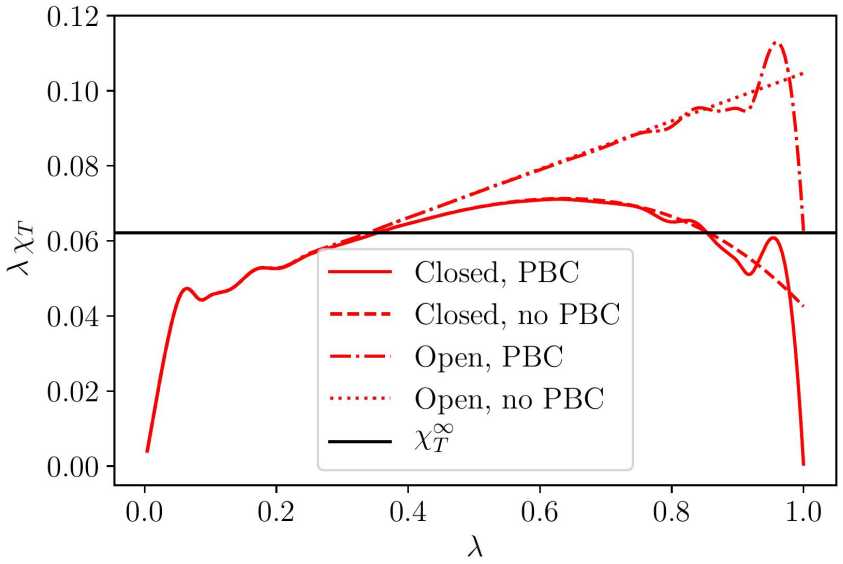

Figure III.3: Calculation of $\lambda\chi_T$ with our method, Eq. VI.3, for the system with $N_0 = 1000$ molecules. We present four cases: (solid) closed system with PBC, (dashed) closed system without PBC, (dash-dotted) open system with PBC, (dotted) open system without PBC. The black horizontal line corresponds to $\chi_T^\infty = 0.062$.

lines). Using the information from the linear region $\lambda < 0.3$, we extrapolate G_{ij}^∞ and obtain α_{ij}. In this case, we get $G_{\text{uu}}^\infty = -0.0867$,$G_{\text{uw}}^\infty = -0.0639$ and $G_{\text{ww}}^\infty = -0.0083$ nm^3 with uu, uw and ww corresponding to urea-urea, urea-water and water-water components, respectively. These values well reproduce derivatives of activity coefficients reported experimentally [18, 25] as well as excess chemical potentials trends with concentration obtained with different computational methods [28, 2]. We insert these values in Eq. (III.14), i.e. SBA, and plot this result as well (solid grey lines). Finally, we use the finite KBI introduced here (Eq. (III.6)) and use the Fourier integral, Eq. (VI.3) to compute $G_{ij}(V)$. We present both results with (solid lines) and without (dash-dotted lines) the correction to the ensemble finite-size effects, $G_{ij}^\infty\lambda^3$.

In this case as well, the results of our method accurately reproduce density fluctuations in the whole range $0 < \lambda < 1$, including both, local structure ($\lambda \ll 1$) and periodic boundary ($\lambda \approx 1$) features. As anticipated, it is also apparent that the SBA result does not reproduce these limiting cases. Nevertheless, in the limit $\lambda = 1$ the results from fluctuations, SBA and our integration (Closed, PBC) converge to $-\delta_{ij}/\rho_i$, the expected result of the KBI for a closed system [7]. As previously stated, we can separate finite-size contributions by focusing on the corresponding terms in Eqs (III.6) and (VI.3). In particular, for an open system (Open, PBC), i.e. $\lim_{V_0\to\infty}$, we verify that the KBI converge to G_{ij}^∞ when $\lambda = 1$.

We examine this in more detail in Fig. III.5 where the normalised KBI for urea-urea, λG_{uu}, is presented. In addition to the limiting cases discussed above, we also consider an open system without periodic boundary conditions (dotted line). It is apparent in the region $\lambda \approx 1$ (inner panel) that G_{uu} for an open system with PBC converges to the value in the thermodynamic limit G_{uu}^∞, whereas for the open system without PBC, G_{uu} is slightly larger than G_{uu}^∞ value by a factor

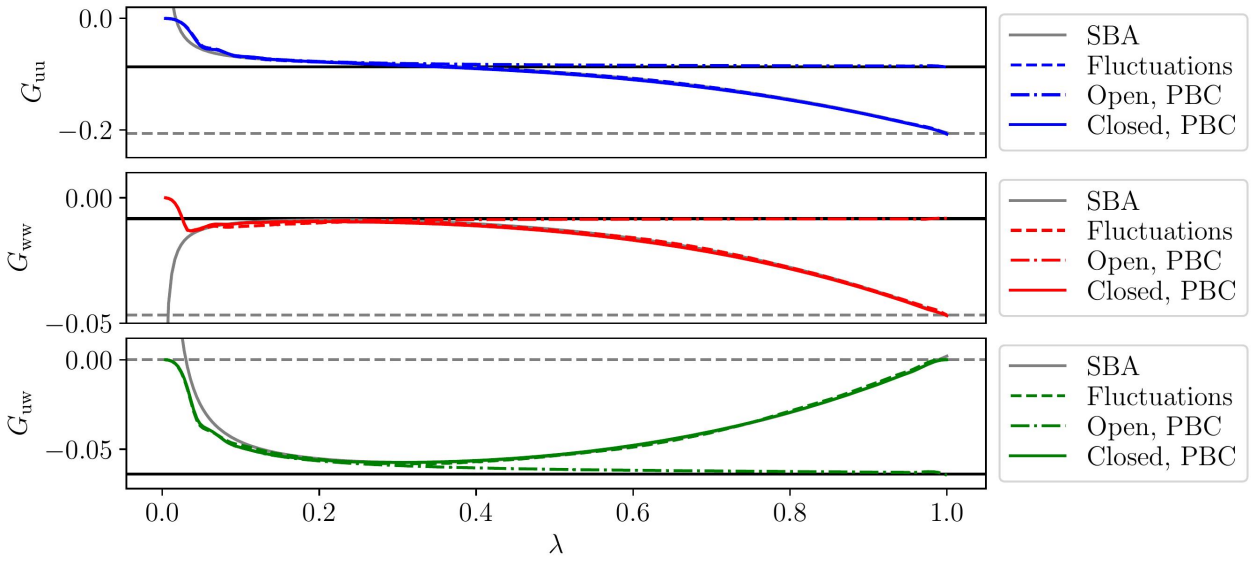

Figure III.4: KBI components G_{uu} (blue, top panel), G_{ww} (red, middle panel) and G_{uw} (green, bottom panel) for an 8M aqueous urea mixture as a function of $\lambda = (V/V_0)^3$. We present density fluctuations as obtained from the left hand side of Eq. (III.2) (dashed lines), the spatial block analysis approximation in Eq. (III.14) (grey lines) and from the finite KBI expression, Eq. (III.6) with $G_{ij}(V)$ given by the Fourier integral Eq. (VI.3) with (solid) and without (dash-dotted lines) the correction to ensemble effects given by $G_{ij}^{\infty}\lambda^3$. The solid black lines correspond to the KBI in the TL, G_{ij}^{∞}. The dashed grey lines indicate the asymptotic limit for the closed system, $-\delta_{ij}/\rho_i$.

$\alpha_{uu}/V_0^{1/3}$, as expected from the SBA expression, Eq. (III.14). As in the single component case, this result emphasises that PBC enforce the correct behaviour at the boundary of closed and open liquid mixtures. Finally, we also present the normalised running integral λG_{ij}^R (dashed red line) using

$$G_{ij}^R = 4\pi \int_0^R dr\, r^2 (g(r;V_0) - 1)\,, \tag{III.19}$$

(with $R > \zeta$), an expression frequently used in the literature. The major differences with the results presented in this work resulting from various finite-size effects highlight the apparent limitations of using such an expression to calculate KBI.

III.5 Concluding remarks

Finite Kirkwood-Buff integrals (KBI) enable us to sample the thermodynamic limit of liquid mixtures via relatively small computer simulations. The definition of finite KBI balance fluctuations of the number of particles in subdomains within the simulation box and integrals of the corresponding RDF. In this work, we underline this equality by reproducing density fluctuations as a function of the linear size of the subdomain via a simple integration strategy. In particular, we introduce a method to evaluate KBI via integrals of the partial structure factors in reciprocal space. A significant advantage of our approach corresponds to the direct inclusion of finite integration domains

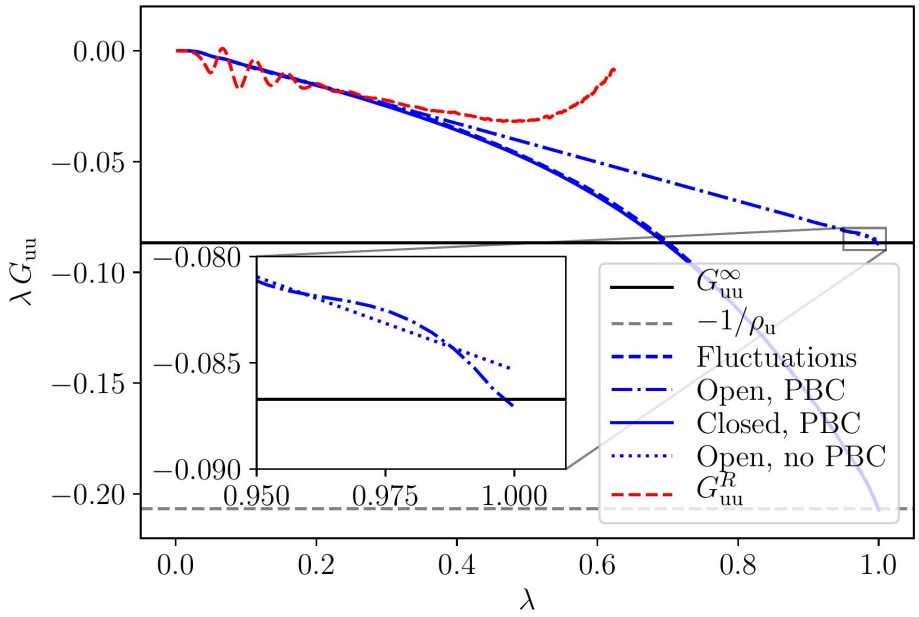

Figure III.5: Normalised KBI, λG_{uu}, as a function of λ obtained using various methods and conditions. Fluctuations - Fluctuations of the number of particles as obtained from the l.h.s. of Eq. (III.2) (dashed lines). Closed, PBC; Open, PBC - finite KBI expression, Eq. (III.6) with $G_{uu}(V)$ given by the Fourier integral Eq. (VI.3) with (solid) and without (dash-dotted lines) the correction to ensemble effects given by $G^{\infty}_{uu}\lambda^3$, both with PBC. Open, no PBC - finite KBI without the correction for the thermodynamic ensemble, and without PBC. The solid black line correspond to the KBI in the TL, G^{∞}_{uu}. The dashed grey line indicate the asymptotic limit for the closed system, $-1/\rho_u$. The red dashed line corresponds to the running integral Eq. (III.19). (Inset) Detail of the convergence to the TL.

and PBC contributions. Consequently, we can now identify and *remove* finite-size effects such that grand canonical and thermodynamic limit results become readily available from finite-size computer simulations. Moreover, we show that this scheme enables us to extrapolate our results to different sizes of the simulation box simply by modifying the periodicity factor in the integration procedure. The simplicity of the method is apparent since it only requires fluctuations of the number of particles calculated for different subdomains sizes and the partial structure factors. We foresee immediate applications in situations where PBC play a pivotal role, namely, the recently introduced KBI for crystalline materials [29, 36].

Acknowledgments

We are grateful to Kurt Kremer for insightful discussions. We also thank Atreyee Banerjee for her critical reading of the manuscript. R.C.-H. gratefully acknowledges funding from SFB-TRR146 of the German Research Foundation (DFG). Simulations have been performed on the THINC cluster at the Max Planck Institute for Polymer Research and on the COBRA cluster of the Max Planck Computing and Data Facility.

Bibliography

[1] N. W. Ashcroft and David C. Langreth. Structure of binary liquid mixtures. i. *Phys. Rev.*, 156:685–692, Apr 1967.

[2] Luis A. Baptista, Ravi Chandra Dutta, Mauricio Sevilla, Maziar Heidari, Raffaello Potestio, Kurt Kremer, and Robinson Cortes-Huerto. Density-functional-theory approach to the hamiltonian adaptive resolution simulation method. *Journal of Physics: Condensed Matter*, 33:184003, 2021.

[3] A. Ben-Naim. *Molecular Theory of Solutions*. Oxford University Press, 2006.

[4] Arieh Ben-Naim. Theoretical aspects of self-assembly of proteins: A kirkwood-buff-theory approach. *J. Chem. Phys.*, 138(22):224906, 2013.

[5] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[6] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. *J. Chem. Phys.*, 81:3684, 1984.

[7] K. Binder. Finite size scaling analysis of ising model block distribution functions. *Z. Phys. B*, 43(2):119–140, 1981.

[8] Vilde Bråten, Øivind Wilhelmsen, and Sondre Kvalvåg Schnell. Chemical potential differences in the macroscopic limit from fluctuations in small systems. *Journal of Chemical Information and Modeling*, 61(2):840–855, 2021. PMID: 33566592.

[9] Ellen E. Bruce, Halil I. Okur, Sina Stegmaier, Chad I. Drexler, Bradley A. Rogers, Nico F. A. van der Vegt, Sylvie Roke, and Paul S. Cremer. Molecular mechanism for the interactions of hofmeister cations with macromolecules in aqueous solution. *Journal of the American Chemical Society*, 142(45):19094–19100, 2020. PMID: 33124825.

[10] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[11] Alper T. Celebi, Noura Dawass, Othonas A. Moultos, and Thijs J. H. Vlugt. How sensitive are physical properties of choline chloride–urea mixtures to composition changes: Molecular dynamics simulations and kirkwood–buff theory. *The Journal of Chemical Physics*, 154(18):184502, 2021.

[12] R. Cortes-Huerto, K. Kremer, and R. Potestio. Communication: Kirkwood-buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *The Journal of Chemical Physics*, 145(14):141103, 2016.

[13] Liem X. Dang and B. Montgomery. Pettitt. Simple intramolecular model potentials for water. *J. Phys. Chem.*, 91(12):3349–3354, 1987.

[14] Tiago E. de Oliveira, Paulo A. Netz, Kurt Kremer, Christoph Junghans, and Debashish Mukherji. C–ibi: Targeting cumulative coordination within an iterative protocol to derive coarse-grained models of (multi-component) complex fluids. *The Journal of Chemical Physics*, 144(17):174106, 2016.

[15] M. Fyta. Structural and technical details of the kirkwood-buff integrals from the optimization of ionic force fields: focus on fluorides. *The European Physical Journal E*, 35(3):21, 2012.

[16] Aikaterini A. Galata, Stefanos D. Anogiannakis, and Doros N. Theodorou. Thermodynamic analysis of lennard-jones binary mixtures using kirkwood-buff theory. *Fluid Phase Equilibria*, 470:25–37, 2018. SI:John P O'Connell.

[17] Pritam Ganguly, Debashish Mukherji, Christoph Junghans, and Nico F. A. van der Vegt. Kirkwood–buff coarse-grained force fields for aqueous solutions. *Journal of Chemical Theory and Computation*, 8(5):1802–1807, 2012. PMID: 26593671.

[18] Pritam Ganguly and Nico F. A. van der Vegt. Convergence of sampling kirkwood-buff integrals of aqueous solutions with molecular dynamics simulations. *J. Chem. Theory Comput.*, 9(3):1347–1355, 2013.

[19] Moon Bae Gee, Nicholas R. Cox, Yuanfang Jiao, Nikolaos Bentenitis, Samantha Weerasinghe, and Paul E. Smith. A kirkwood-buff derived force field for aqueous alkali halides. *Journal of Chemical Theory and Computation*, 7(5):1369–1380, 2011. PMID: 21789033.

[20] Moon Bae Gee and Paul E. Smith. Kirkwood–buff theory of molecular and protein association, aggregation, and cellular crowding. *The Journal of Chemical Physics*, 131(16):165101, 2009.

[21] W. Härtl, C. Segschneider, H. Versmold, and P. Linse. On the structure factor of liquid-like ordered binary mixtures of colloidal suspensions. *Molecular Physics*, 73(3):541–552, 1991.

[22] M. Heidari, K. Kremer, R. Potestio, and R. Cortes-Huerto. Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations. *Molecular Physics*, 116(21-22):3301–3310, 2018.

[23] Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto. Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: Revisiting the spatial block analysis method. *Entropy*, 20(4), 2018.

[24] T. L. Hill. *Thermodynamics of Small Systems*. Dover, 1963.

[25] Myungshim Kang and Paul E. Smith. Preferential interaction parameters in biological systems by kirkwood–buff theory and computer simulation. *Fluid Phase Equilibria*, 256(1):14–19, 2007. 16th Symposium on Thermophysical Properties.

[26] John G. Kirkwood and Frank P. Buff. The statistical mechanical theory of solutions. i. *The Journal of Chemical Physics*, 19(6):774–777, 1951.

[27] Signe Kjelstrup, Sondre K Schnell, Thijs J H Vlugt, Jean-Marc Simon, Andre Bardow, Dick Bedeaux, and Thuat Trinh. Bridging scales with thermodynamics: from nano to macro. *Adv. Nat. Sci.: Nanosci. Nanotechnol.*, 5(2):023002, 2014.

[28] Hironori Kokubo, Jörg Rösgen, D. Wayne Bolen, and B. Montgomery Pettitt. Molecular basis of the apparent near ideality of urea solutions. *Biophysical Journal*, 93(10):3392 – 3407, 2007.

[29] Peter Krüger. Validity of the compressibility equation and kirkwood-buff theory for crystalline matter. *Phys. Rev. E*, 103:L061301, Jun 2021.

[30] Peter Krüger, Sondre K. Schnell, Dick Bedeaux, Signe Kjelstrup, Thijs J. H. Vlugt, and Jean-Marc Simon. Kirkwood-buff integrals for finite volumes. *J. Phys. Chem. Lett.*, 4(2):235–238, 2013.

[31] P. Kumari, V. V. S. Pillai, D. Gobbo, P. Ballone, and A. Benedetto. The transition from salt-in-water to water-in-salt nanostructures in water solutions of organic ionic liquids relevant for biological applications. *Phys. Chem. Chem. Phys.*, 23:944–959, 2021.

[32] J. L. Lebowitz and J. K. Percus. Long-range correlations in a closed system with applications to nonuniform fluids. *Phys. Rev.*, 122:1675–1691, Jun 1961.

[33] X. Liu, S. K. Schnell, J. M. Simon, P. Krüeger, D. Bedeaux, S. Kjelstrup, A. Bardow, and T. J. H. Vlugt. Diffusion coefficients from molecular dynamics simulations in binary and ternary mixtures. *Int. J. Thermophys.*, 34:1169, 2013.

[34] Philip Loche, Patrick Steinbrunner, Sean Friedowitz, Roland R. Netz, and Douwe Jan Bonthuis. Transferable ion force fields in water from a simultaneous optimization of ion solvation and ion–ion interaction. *The Journal of Physical Chemistry B*, 125(30):8581–8587, 2021. PMID: 34292738.

[35] Bernarda Lovrinčević, Adrien Bella, Isham Le Tenoux-Rachidi, Martina Požar, Franjo Sokolić, and Aurélien Perera. Methanol-ethanol "ideal" mixtures as a test ground for the computation of kirkwood-buff integrals. *Journal of Molecular Liquids*, 293:111447, 2019.

[36] Masafumi Miyaji, Bastien Radola, Jean-Marc Simon, and Peter Krüger. Extension of kirkwood–buff theory to solids and its application to the compressibility of fcc argon. *The Journal of Chemical Physics*, 154(16):164506, 2021.

[37] Debashish Mukherji and Kurt Kremer. Coil–globule–coil transition of pnipam in aqueous methanol: Coupling all-atom simulations to semi-grand canonical coarse-grained reservoir. *Macromolecules*, 46(22):9158–9163, 2013.

[38] Debashish Mukherji, Nico F. A. van der Vegt, and Kurt Kremer. Preferential solvation of triglycine in aqueous urea: An open boundary simulation approach. *Journal of Chemical Theory and Computation*, 8(10):3536–3541, 2012. PMID: 26593001.

[39] Debashish Mukherji, Nico F. A. van der Vegt, Kurt Kremer, and Luigi Delle Site. Kirkwood–buff analysis of liquid mixtures in an open boundary simulation. *Journal of Chemical Theory and Computation*, 8(2):375–379, 2012. PMID: 26596589.

[40] Anand Narayanan Krishnamoorthy, Christian Holm, and Jens Smiatek. Influence of cosolutes on chemical equilibrium: a kirkwood–buff theory for ion pair association–dissociation processes in ternary electrolyte solutions. *The Journal of Physical Chemistry C*, 122(19):10293–10302, 2018.

[41] Ewa Anna Oprzeska-Zingrebe, Miriam Kohagen, Johannes Kästner, and Jens Smiatek. Unfolding of dna by co-solutes: insights from kirkwood–buff integrals and transfer free energies. *The European Physical Journal Special Topics*, 227(14):1665–1679, 2019.

[42] Ewa Anna Oprzeska-Zingrebe and Jens Smiatek. Aqueous ionic liquids in comparison with standard co-solutes. *Biophysical Reviews*, 10(3):809–824, 2018.

[43] Ewa Anna Oprzeska-Zingrebe and Jens Smiatek. Interactions of a dna g-quadruplex with tmao and urea: a molecular dynamics study on co-solute compensation mechanisms. *Phys. Chem. Chem. Phys.*, 23:1254–1264, 2021.

[44] Panagiotis C. Petris, Stefanos D. Anogiannakis, Panagiotis-Nikolaos Tzounis, and Doros N. Theodorou. Thermodynamic analysis of n-hexane–ethanol binary mixtures using the kirkwood–buff theory. *The Journal of Physical Chemistry B*, 123(1):247–257, 2019.

[45] Veronica Pierce, Myungshim Kang, Mahalaxmi Aburi, Samantha Weerasinghe, and Paul E. Smith. Recent applications of kirkwood-buff theory to biological systems. *Cell Biochem. Biophys.*, 50(1):1–22, 2008.

[46] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.

[47] F L Román, J A White, and S Velasco. Fluctuations in an equilibrium hard-disk fluid: Explicit size effects. *J. Chem. Phys.*, 107:4635, 1997.

[48] F. L. Román, J. A. White, and S. Velasco. Block analysis method in off-lattice fluids. *EPL*, 42(4):371, 1998.

[49] F.L. Román, J.A. White, A. González, and S. Velasco. *Theory and Simulation of Hard-Sphere Fluids and Related Systems*, chapter Ensemble Effects in Small Systems, pages 343–381. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[50] F. L. Román, J. A. White, A. González, and S. Velasco. Fluctuations in a small hard-disk system: Implicit finite size effects. *The Journal of Chemical Physics*, 110(20):9821–9824, 1999.

[51] M Rovere, D W Heermann, and K Binder. The gas-liquid transition of the two-dimensional lennard-jones fluid. *J. Phys.: Condens. Matter*, 2(33):7009, 1990.

[52] M. Rovere, D. W. Hermann, and K. Binder. Block density distribution function analysis of two-dimensional lennard-jones fluids. *EPL*, 6(7):585, 1988.

[53] 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. *Phys. Rev. E*, 53:2382–2389, Mar 1996.

[54] Emanuel Schneck, Dominik Horinek, and Roland R. Netz. Insight into the molecular mechanisms of protein stabilizing osmolytes from global force-field variations. *The Journal of Physical Chemistry B*, 117(28):8310–8321, 2013. PMID: 23822090.

[55] Sondre K. Schnell, Xin Liu, Jean-Marc Simon, André Bardow, Dick Bedeaux, Thijs J. H. Vlugt, and Signe Kjelstrup. Calculating thermodynamic properties from fluctuations at small scales. *The Journal of Physical Chemistry B*, 115(37):10911–10918, 2011. PMID: 21838274.

[56] Sondre K. Schnell, Thijs J.H. Vlugt, Jean-Marc Simon, Dick Bedeaux, and Signe Kjelstrup. Thermodynamics of a small system in a μt reservoir. *Chem. Phys. Lett.*, 504(4–6):199 – 201, 2011.

[57] Seishi Shimizu, Richard Stenner, and Nobuyuki Matubayasi. Gastrophysics: Statistical thermodynamics of biomolecular denaturation and gelation from the kirkwood-buff theory towards the understanding of tofu. *Food Hydrocolloids*, 62:128–139, 2017.

[58] Luigi Delle Site, Giovanni Ciccotti, and Carsten Hartmann. Partitioning a macroscopic system into independent subsystems. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(8):083201, aug 2017.

[59] Madhusmita Tripathy, Swaminath Bharadwaj, Shadrack Jabes B., and Nico F. A. van der Vegt. Characterizing polymer hydration shell compressibilities with the small-system method. *Nanomaterials*, 10(8), 2020.

[60] Dario Villamaina and Emmanuel Trizac. Thinking outside the box: fluctuations and finite size effects. *Eur. J. Phys.*, 35(3):035011, 2014.

[61] Samantha Weerasinghe and Paul E. Smith. A kirkwood-buff derived force field for mixtures of urea and water. *The Journal of Physical Chemistry B*, 107(16):3891–3898, 2003.

[62] Yujie Wu, Harald L. Tepper, and Gregory A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.*, 124(2):024503, 2006.

IV. Finite-size excess-entropy scaling for simple liquid

Outline

Bibliographic Information

Mauricio Sevilla, Atreyee Banerjee, and Robinson Cortes-Huerto. Finite-size excess-entropy scaling for simple liquids. *The Journal of Chemical Physics*, 158(20):204502, 05 2023

Author contribution

- Mauricio Sevilla: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Software (lead); Writing – original draft (equal); Writing – review & editing (equal).
- Atreyee Banerjee: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal).
- Robinson Cortes-Huerto: Conceptualization (lead); Formal analysis (equal); Investigation (equal); Project administration (lead); Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal).

Abstract

Explicit and implicit size effects in computer simulations result from considering systems with a fixed number of particles and periodic boundary conditions, respectively. We investigate these effects in the relation $D^*(L) = A(L)\exp(\alpha s_2(L))$ between reduced self-diffusion coefficient $D^*(L)$ and two-body excess entropy $s_2(L)$ for prototypical simple-liquid systems of linear size L. To this aim, we introduce and validate a finite-size two-body excess entropy integral equation. Our analytical arguments and simulation results show that s2(L) exhibits a linear scaling with 1/L. Since D*(L) displays a similar behavior, we show that the parameters $A(L)$ and $\alpha(L)$ are also linearly proportional to $1/L$. By extrapolating to the thermodynamic limit, we report the coefficients $A^\infty = 0.048 \pm 0.001$ and $\alpha^\infty = 1.000 \pm 0.013$ that agree well with the universal values available in the literature [M. Dzugutov, Nature 381, 137–139 (1996)]. Finally, we find a power law relation between the scaling coefficients for $D^*(L)$ and $s_2(L)$, suggesting a constant viscosity-to-entropy ratio.

IV.1 Introduction

Excess entropy (s_{exc}), the difference between the entropy of a system and its ideal gas counterpart at the same temperature and density, is connected to the dynamical properties of simple liquids (See Ref. [19] for a recent review). This observation was first reported by Rosenfeld [51, 52], who showed that, for simple model liquids, reduced transport properties such as diffusivity, viscosity and thermal conductivity scale with the excess entropy as

$$X^* = A\exp\left(\alpha s_{\text{exc}}\right) , \tag{IV.1}$$

with X^* a dimensionless transport property and A and α parameters, independent of the thermodynamic state, determined by the interparticle potential.
Following similar physical arguments and assuming that the major contribution to s_{exc} comes from two-body terms, Dzugutov proposed a similar scaling relation between reduced self-diffusivity and a two-body approximation to the excess entropy s_2, namely [20]

$$D^* = A\exp(\alpha s_2) , \tag{IV.2}$$

with $D^* = D/\Gamma\sigma_r^2$ where D is the self-diffusion coefficient, σ_r measures the linear size of the particles and $\Gamma = 4\sigma_r^2 g(\sigma_r)\rho\sqrt{\pi k_{\text{B}}T/m}$ the collision frequency given by the Enskog theory [15] where $g(\sigma_r)$ is the value of the radial distribution function at a distance σ_r, ρ the density of the system, k_{B} the Boltzmann constant and m the mass of the particle. In this case, a large variety of simple liquids satisfy Eq. (IV.2) with the *universal* choice of parameters $A = 0.049$ and $\alpha = 1$ [20]. This excess entropy scaling has been widely validated for a large variety of simple [30, 14, 9, 39, 4, 6, 63] and molecular liquids[24, 41, 16, 23], including specially water [58, 2, 17, 1]. We also highlight that experimental studies have tested entropy scaling in somewhat challenging scenarios [40, 59], and the fact that Rosenfeld and Dzugutov relations are empirical but have been justified on theoretical grounds [55, 56, 43]. Furthermore, the structure–dynamics connection in Eq. (IV.2) has been proposed as a tool to investigate the relation between dynamical properties of computational models at different resolutions, [5], which is now routinely considered in the context of coarse-grained models [50, 31].
It has been extensively reported that transport properties exhibit explicit and implicit size effects

due to the finite size of the simulation box and the use of periodic boundary conditions (PBC), respectively [21, 34, 65]. In the particular case of the reduced self-diffusion coefficient D^*, given a cubic simulation box of linear size L, $D^* \equiv D^*(L)$ takes the form [21, 64, 32, 33, 12]:

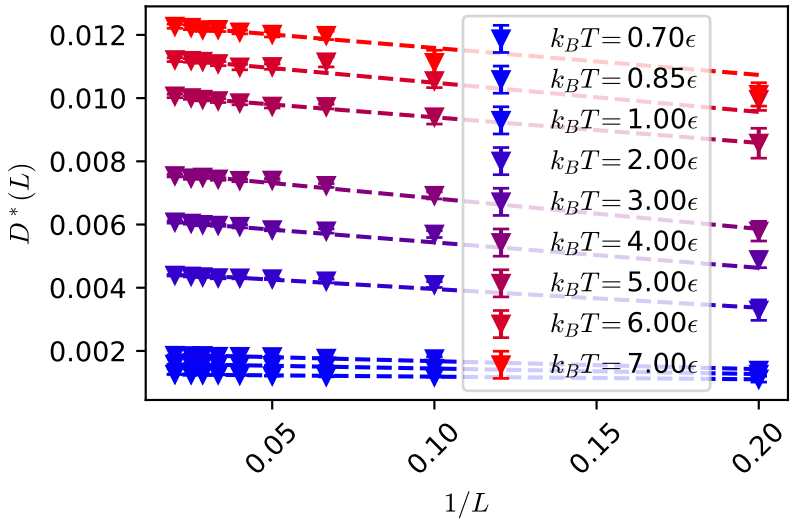

Figure IV.1: Reduced self-diffusion coefficient D^* as a function of the inverse of the box linear size $1/L$ for a Lennard-Jones liquid with density $\rho\sigma^3_{\mathrm{LJ}} = 0.864$ in the range of temperatures $0.7\epsilon \leq k_{\mathrm{B}}T \leq 7\epsilon$.

$$D^*(L) = D^{*\infty} - \frac{\delta}{L}\,, \tag{IV.3}$$

with $\delta = k_{\mathrm{B}}T\zeta/6\pi\eta\Gamma\sigma_r^2$ with $\zeta \approx 2.837297$, and T and η the temperature and viscosity of the system, respectively. Here, for completeness, we show our results of D^* vs $1/L$ in Figure IV.1 (See Section IV.2 for simulation details), which provide the basis of the present study. In the thermodynamic limit (TL), namely, in the limit $L \to \infty$, the self-diffusion coefficient takes the value $D^{*\infty}$.

Given the finite-size scaling of D^*, we expect that Eq. (IV.2) also depends on the size of the simulation box with a consistent $L \to \infty$ limit. Recent computational studies investigating entropy scaling for liquid water using ab initio molecular dynamics simulations [27] emphasize the relevance of this remark. In this case, the systems under consideration are rather small, and finite-size effects become increasingly important.

In this paper, we investigate the size scaling of Eq. (IV.2) by focusing on implicit and explicit finite-size effects present on the two-body excess entropy s_2. We find that s_2 obeys a scaling relation similar to D^*, which implies that the *universal* parameters A and α in Eq. (IV.2) also depend on the size of the simulation box. Finally, and perhaps more interestingly, our results indicate that a power law relates the scaling coefficients of D^* and s_2, suggesting a constant viscosity-to-entropy ratio [35, 3, 22, 29].

The paper is organized as follows: In Section IV.2, we present the model and computational details. In Section IV.3, we show that s_2 is ensemble invariant and that only implicit effects are relevant. In Section IV.4, we introduce and validate a finite-size version of s_2. In Section IV.5, we present

the scaling of the Dzugutov relation (Eq. (IV.2)) with the linear size of the simulation box. Finally, we conclude and provide our outlook in Section IV.6.

IV.2 Computational details

We investigate the excess entropy scaling for liquids whose potential energy is described by a 12–6 Lennard–Jones potential truncated, with cutoff radius $r_c/\sigma_{\mathrm{LJ}} = 2.5$, and shifted. The parameters ϵ, σ_{LJ} and m, define the energy, length and mass units, respectively. All the results are expressed in LJ units with time $\sigma_{\mathrm{LJ}}(m/\epsilon)^{1/2}$, temperature ϵ/k_{B} and pressure $\epsilon/\sigma_{\mathrm{LJ}}^3$. In the following, we identify σ_r of Eq. (IV.3) with σ_{LJ}. We consider cubic simulation boxes with linear sizes in the interval $5 \leq L/\sigma_{\mathrm{LJ}} \leq 50$, with fixed density $\rho\sigma_{\mathrm{LJ}}^3 = 0.864$. The systems are equilibrated at temperatures in the interval $0.7\epsilon \leq k_{\mathrm{B}}T \leq 7.0\epsilon$, enforced with a Langevin thermostat with damping coefficient $\gamma(\sigma(m/\epsilon)^{1/2}) = 1.0$. We equilibrate the samples for 10^7 molecular dynamics (MD) steps using a time step of $\delta t/(\sigma_{\mathrm{LJ}}(m/\epsilon)^{1/2}) = 10^{-3}$, followed by additional 10^7 MD steps on the NVE ensemble to verify that the temperature does not deviate substantially from the target value. Production runs span 10^7 MD steps. All the simulations have been performed with the LAMMPS simulation package [56].

IV.3 Explicit and implicit size effects

In this section, we identify the size effects affecting the calculation of excess entropy. We start with the definition of excess entropy for an N–particle system with respect to the ideal gas:

$$s_{\mathrm{exc}} = \frac{S - S_{\mathrm{IG}}}{Nk_{\mathrm{B}}} = \frac{S_2 + S_3 + \cdots}{Nk_{\mathrm{B}}} \,. \tag{IV.4}$$

In the following, we focus on two-body contributions, which mostly amount to 80–90% of the overall value of the excess entropy for simple liquids. [11, 7] In particular, we have in the grand canonical ensemble that [45, 42]

$$s_2 = -\frac{\rho}{2V} \int_V \int_V d\mathbf{r}_1 \, d\mathbf{r}_2 \, \left[g(\mathbf{r}) \ln g(\mathbf{r}) - (g(\mathbf{r}) - 1) \right] \,, \tag{IV.5}$$

with $s_2 = S_2/Nk_{\mathrm{B}}$ the two-body excess entropy per particle and $g(\mathbf{r})$ the radial distribution function (RDF) with $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. By taking the thermodynamic limit and assuming that the liquid is homogeneous and isotropic, we obtain

$$s_2^{\infty} = -2\pi\rho \int_0^{\infty} dr \, r^2 \left[g(r) \ln g(r) - (g(r) - 1) \right] \,, \tag{IV.6}$$

where we have replaced $g(\mathbf{r})$ with $g(r)$. When performing molecular dynamics simulations, we usually consider systems with a finite number of particles, typically not large enough to reach the thermodynamic limit. For such a reason, it is a common practice to truncate Eq. (IV.6) up to a cutoff radius R, namely

$$\begin{aligned} s_2^R = -2\pi\rho \int_0^R dr \, r^2 \, & \left[g(r; N_0) \ln g(r; N_0) \right. \\ & \left. - (g(r; N_0) - 1) \right] \,, \end{aligned} \tag{IV.7}$$

and evaluate the integral by using the RDF of the finite system, i.e. $g(r;N_0)$. This approach gives satisfactory results provided one uses relatively large simulation boxes in which $R \gg \zeta$ with ζ the correlation length of the system and $g(r;N_0) \to g(r)$.
In the following, we propose a different method. When evaluating the double integral in Eq. (IV.5) we must recognize that the volume V is finite. For such a reason, and following the strategy used to compute the compressibility equation [41, 26] and the Kirkwood-Buff integrals [36, 18, 25], we define a finite–size two–body excess entropy evaluated in a subvolume V of a system with a total number of particles N_0 (at a fixed volume V_0)

$$s_2(V;N_0) = -\frac{\rho}{2V}\int_V\int_V d\mathbf{r}_1\, d\mathbf{r}_2\, [g(\mathbf{r};N_0)\ln g(\mathbf{r};N_0) - (g(\mathbf{r};N_0)-1)]\,. \tag{IV.8}$$

The asymptotic correction to the finite-size RDF, given by the difference in the thermodynamic ensemble, gives [31, 30, 45, 47, 57, 46]

$$g(\mathbf{r};N_0) = g(\mathbf{r}) - \frac{\chi_T^\infty}{N_0}\,, \tag{IV.9}$$

with $\chi_T^\infty = \rho k_B T \kappa_T$, and κ_T being the bulk isothermal compressibility. We write the integrand in Eq. (IV.8) as

$$\begin{aligned} g(\mathbf{r};N_0)\ln g(\mathbf{r};N_0) &\approx g(\mathbf{r})\ln g(\mathbf{r}) \\ &\quad - \frac{\chi_T^\infty}{N_0}(1+\ln g(\mathbf{r})) \\ g(\mathbf{r};N_0) - 1 &= g(\mathbf{r}) - 1 - \frac{\chi_T^\infty}{N_0}\,, \end{aligned} \tag{IV.10}$$

where in the first line in the previous expression, we have neglected terms of the order $O\left(1/N_0^2\right)$. The two contributions χ_T^∞/N_0 cancel out exactly. The contribution $\chi_T^\infty \ln g(\mathbf{r})/N_0$ can be neglected by assuming a large number of particles (there is no V/V_0 contribution, only $1/V_0$. Hence, we can neglect it). This indicates that the two-body excess entropy is ensemble invariant, consistent with the result reported Ref. [62, 8]. We thus rewrite Eq. (IV.8) as

$$s_2(V) = -\frac{\rho}{2V}\int_V\int_V d\mathbf{r}_1\ d\mathbf{r}_2\, [g(\mathbf{r})\ln g(\mathbf{r}) - (g(\mathbf{r})-1)]\,. \tag{IV.11}$$

The volume V is finite and embedded into the volume V_0. The integration domains can be rearranged as $\int_V \int_V(\cdots) = \int_V \int_{V_0}(\cdots) - \int_V \int_{V_0-V}(\cdots)$. Using a similar argument as the one used to calculate the finite-size compressibility [53] and Kirkwood-Buff integrals [36], the term $\int_V \int_{V_0}(\cdots)$ gives s_2^∞ and the term $\int_V \int_{V_0-V}(\cdots)$ scales as $1/L$ with $L = V^{1/3}$ the linear size of the cubic simulation box. Thus, we obtain the following finite-size scaling for s_2

$$s_2(L) = s_2^\infty + \frac{\sigma}{L}\,, \tag{IV.12}$$

with σ a constant that depends on intensive thermodynamic quantities only. In the following section, we introduce a method to compute $s_2(L)$ and confirm its scaling behaviour with the linear

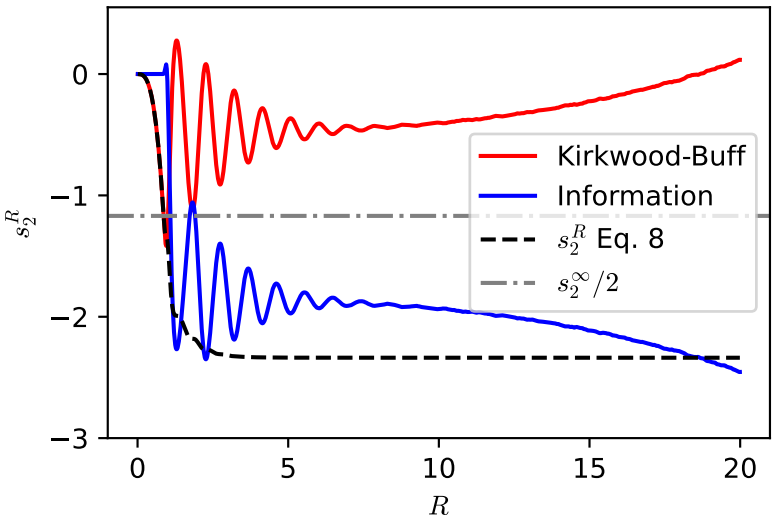

Figure IV.2: Plot of the two contributions, *Kirkwood-Buff* ($g(r;N_0)-1$) and *Information* ($g(r;N_0)\ln g(r;N_0)$), to the truncated integral s_2^R for a system of linear size $L/\sigma_{\mathrm{LJ}}=35$ at $k_{\mathrm{B}}T=2.0\epsilon$. It is apparent that the two terms oscillate out-of-phase for small values of R, and their sum converges to s_2^∞ when $R\to\infty$.

size of the simulation box.

To finish this section, we use the truncated integral, Eq. (IV.7), to verify numerically that ensemble finite-size contributions cancel out almost exactly [63]. For a system of size $L/\sigma_{\mathrm{LJ}}=35$ at $k_{\mathrm{B}}T=2.0\epsilon$, we separate the $g(r;N_0)-1$, *Kirkwood-Buff*, and the $g(r;N_0)\ln g(r;N_0)$, *Information*, contributions and plot them as a function of the truncation radius R (See Figure IV.2). Both integrals diverge for large values of R, Kirkwood-Buff to infinity and Information to minus infinity, which signals a clear ensemble finite-size effect. However, these two finite-size contributions balance each other (See the dashed-dotted line in the figure), and the sum of the two integrals converges to s_2^∞ for large values of R. Due to this error cancellation, the truncation Eq. (IV.7) gives a reasonable approximation to s_2^∞, provided a large value of R, and its finite-size dependence has been commonly overlooked in the literature.

IV.4 Finite-volume excess entropy

Based on previous work aiming at computing the isothermal compressibility [49] and Kirkwood-Buff integrals [57], we define a finite-volume two-body excess entropy as

$$s_2(V)=-\frac{\rho}{2V}\int\int d\mathbf{r}_1\,d\mathbf{r}_2\,R(\mathbf{r}_1)\,R(\mathbf{r}_2)\,h(\mathbf{r})\,, \tag{IV.13}$$

with $R(\mathbf{r})$ a step function that defines the finite integration subdomain, being equal to one inside and to zero outside the volume V [49]. The function $h(\mathbf{r})$ is defined as

$$h(\mathbf{r})=g(\mathbf{r})\ln g(\mathbf{r})-(g(\mathbf{r})-1)\,. \tag{IV.14}$$

We write the double integral of $s_2(V)$ in Fourier space and include the periodicity of the simulation of the box in $h(\mathbf{r})$ explicitly. Thus

$$s_2(V) = -\frac{\rho}{2(2\pi)^3 V} \int d\mathbf{k}\, \tilde{R}(\mathbf{k})\, \tilde{R}(-\mathbf{k})\, \tilde{h}^{\mathrm{PBC}}(\mathbf{k})\,, \tag{IV.15}$$

where [49]

$$\tilde{h}^{\mathrm{PBC}}(\mathbf{k}) = \sum_{n_x,n_y,n_z} e^{-\mathbf{k}\cdot\mathbf{s}_{n_x,n_y,n_z}} \tilde{h}(\mathbf{k})\,, \tag{IV.16}$$

with $\tilde{h}(\mathbf{k})$ the Fourier transform of $h(\mathbf{r})$ and $\mathbf{s}_{n_x,n_y,n_z} = (n_x L_x, n_y L_y, n_z L_x)$ a vector specifying the system's periodic images such that $n_{x,y,z}$ takes integer values. In the following, we consider a cubic simulation box with $L_x = L_y = L_z = L$. As before [57], we choose $|n_x| \leq 1$, $|n_y| \leq 1$ and $|n_z| \leq 1$ to compute Eq. (VI.2). Finally, we assume a homogeneous and isotropic fluid such that $\tilde{h}(\mathbf{k}) = \tilde{h}(k)$ with $k = \sqrt{\mathbf{k}\cdot\mathbf{k}}$.

To validate our approach, we verify that Eqs (IV.15) and (IV.7) converge to the same value

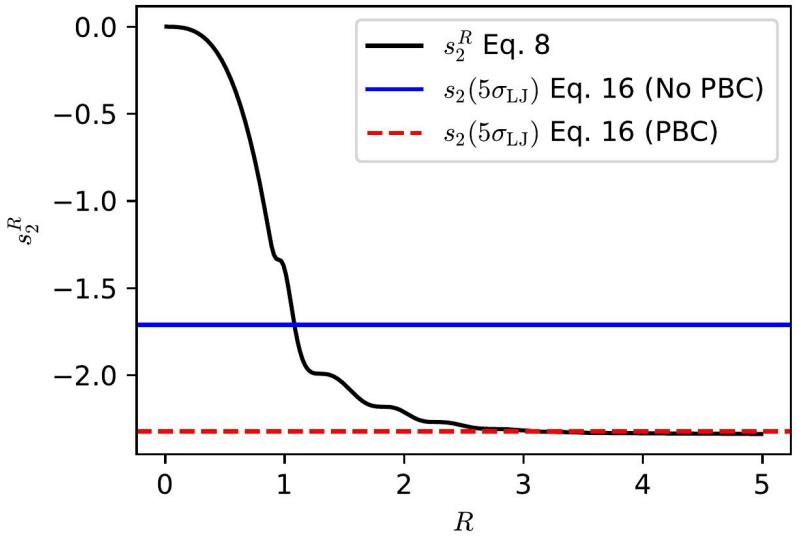

Figure IV.3: Running s_2 as a function of the ratio R/L for the case $L/\sigma_{\mathrm{LJ}} = 5$ at $k_{\mathrm{B}}T = 2.0\epsilon$. The black line corresponds to the truncation Eq. IV.7, and the red and blue curves are the result of Eq. (IV.15) including ($|n_x| \leq 1$, $|n_y| \leq 1$ and $|n_z| \leq 1$) and not including ($|n_x| = |n_y| = |n_z| = 0$) PBC, respectively. By including PBC, the integral Eq. (IV.15) converges to the thermodynamic limit.

in the thermodynamic limit. To this aim, we consider a system with linear size $L/\sigma_{\mathrm{LJ}} = 50$ at $k_{\mathrm{B}}T = 2.0\epsilon$, compute the RDF and evaluate the truncated integral Eq. (IV.7). According to Eq. (IV.12), implicit finite-size effects are the most relevant in this case. Hence, by considering a sufficiently large simulation box, the large R limit of Eq. (IV.7) converges to the TL value. We present this result in Fig. IV.3 (black solid curve). To evaluate Eq. (IV.15), we take the RDF from the simulation box with linear size $L/\sigma_{\mathrm{LJ}} = 20$ and perform the Fourier transform procedure described above to obtain $\tilde{h}(\mathbf{k})$. It is apparent, as expected, that with explicit PBC, the finite-size s_2 gives the TL value (red dashed curve). Instead, by removing PBC, there is a significant deviation from the TL value that we attribute to the $1/L$ dependence in Eq. (IV.12) (blue solid

curved).

We verify this $1/L$ dependence in the finite-size s_2. In Figure IV.4, we plot the result of s_2^R,

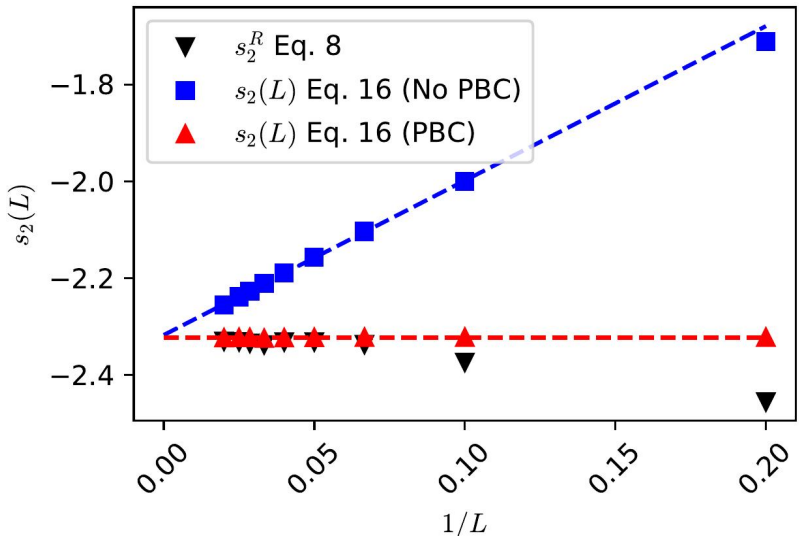

Figure IV.4: s_2 as a function of the inverse of the simulation box size L for systems at $k_BT = 2.0\epsilon$. The black triangles are calculated with the truncated integral (Eq. (IV.7)), the red triangles and blue squares were calculated with the double integral (Eq. (IV.15)) including and excluding PBC, respectively.

Eq. (IV.7), as a function of $1/L$ (black inverted triangles). There, it is apparent that the integral converges when the linear size of the system is $L/\sigma_{LJ} = 15$. The result of using $s_2(V)$, Eq. (IV.15), with explicit PBC, always converges to the TL value (red triangles), regardless of the linear size of the system. More interestingly, by removing PBC from Eq. (IV.15), we observe a clear linear dependence with $1/L$ (blue squares). Furthermore, by extrapolating this behaviour (blue dashed line) to the axis $1/L = 0$, we obtain a linear extrapolation to s_2^∞ (dashed red line). This result completes the validation of both, Eqs (IV.12) and (IV.15).

IV.5 Size scaling of the Dzugutov relation

In this section, we combine the reduced self-diffusion coefficient D^* data, presented in Figure IV.1, with the calculation of the finite-volume s_2 to investigate the size scaling of the Dzugutov relation, Eq. (IV.2). To this aim, we verify that the scaling of s_2 with $1/L$ is valid in the temperature range $0.7\epsilon \leq k_BT \leq 7.0\epsilon$ for our LJ system at $\rho\sigma_{LJ}^3 = 0.864$. We present these results in Figure IV.5, where it is apparent that s_2 scales with $1/L$ in a wide temperature range.

We now collect all our data presented in Figures IV.1 and IV.5 to investigate the scaling of Eq. (IV.2) with the simulation box size. The result is presented in Figure IV.6 where the diffusion constant D^* is plotted against $-s_2$. A clear trend with system size emerges, indicating that Eq. (IV.2) remains valid even for the smallest simulation boxes considered and showing that the parameters A and α are also size dependent. By extrapolating D^* and $-s_2$ to the limit $1/L \to 0$, we obtain the TL values given by the black empty triangles that well agree with the reference scaling provided by Eq. (IV.2) (black dashed line).

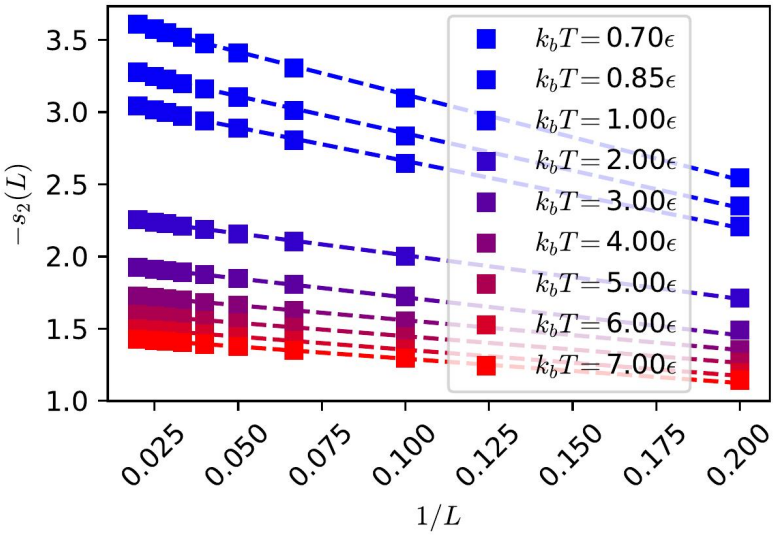

Figure IV.5: -s_2 as a function of $1/L$ for a LJ system at $\rho\sigma^3_{\mathrm{LJ}} = 0.864$ and different temperatures. All data points were obtained with the RDF for the system of linear size $L/\sigma_{\mathrm{LJ}} = 20$ and using Eq. (IV.15) without PBC.

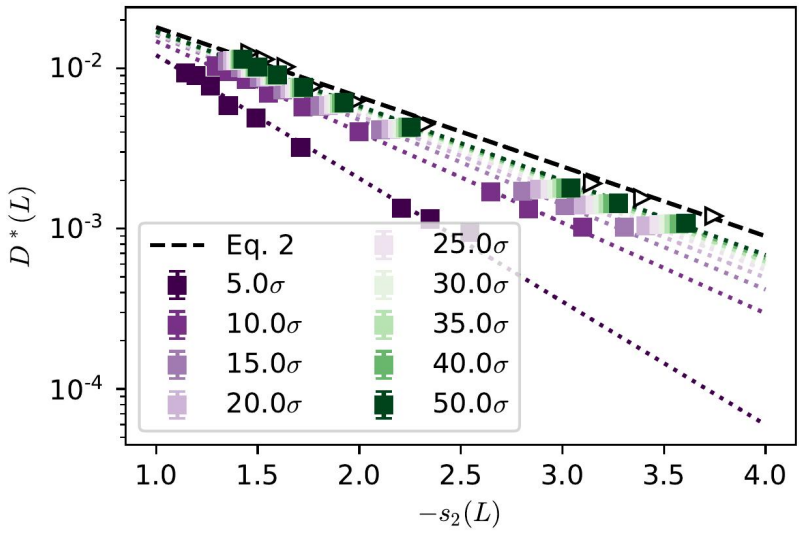

Figure IV.6: Reduced self-diffusion coefficient D^* as a function of $-s_2$ for different system sizes and temperatures. The empty black triangles show the thermodynamic limit values for $-s_2$ and D^*.

We investigate the size dependence of the parameters A and α and present the result in Figure IV.7. It is apparent that A and α display a linear scaling with $1/L$. Perhaps more importantly , we report the values in the limit $L \to \infty$ (TL) $A^\infty = 0.048 \pm 0.001$ and $\alpha^\infty = 1.000 \pm 0.013$, in good agreement with Ref. [20]. Hence, these results show that the size scaling of the reduced self-diffusion coefficient in Eq. (IV.2) result in a similar scaling for s_2, A and α.
Finally, we investigate the relation between the coefficients δ and σ of the finite-size scaling of D^* and s^2, respectively. In Figure IV.8, we plot σ as a function of δ and observe a power law relation

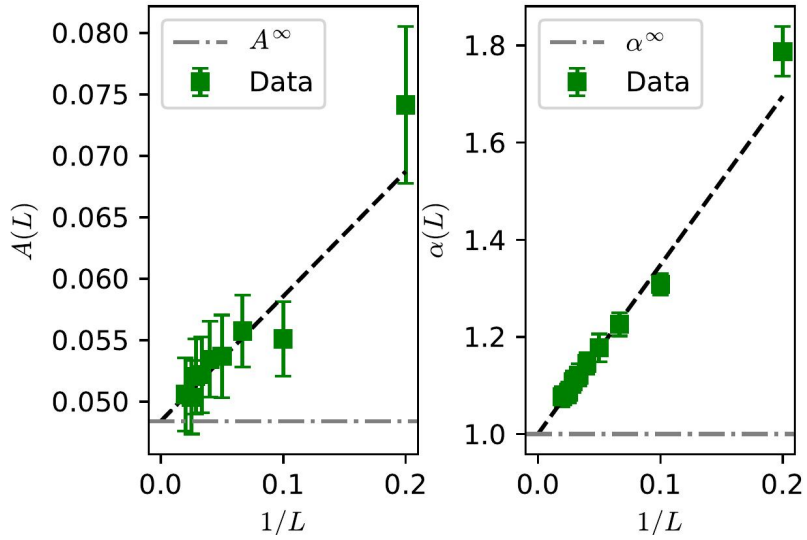

Figure IV.7: Parameters A and α in Eq. (IV.2) as a function of $1/L$. In both cases, a linear scaling can be appreciated and the limit $L \to \infty$ gives the universal values $A^\infty = 0.048$ and $\alpha^\infty = 1$.

of the form $\sigma = a\delta^b$ with $a = 1.256 \pm 0.118$ and $b = -0.513 \pm 0.020$.

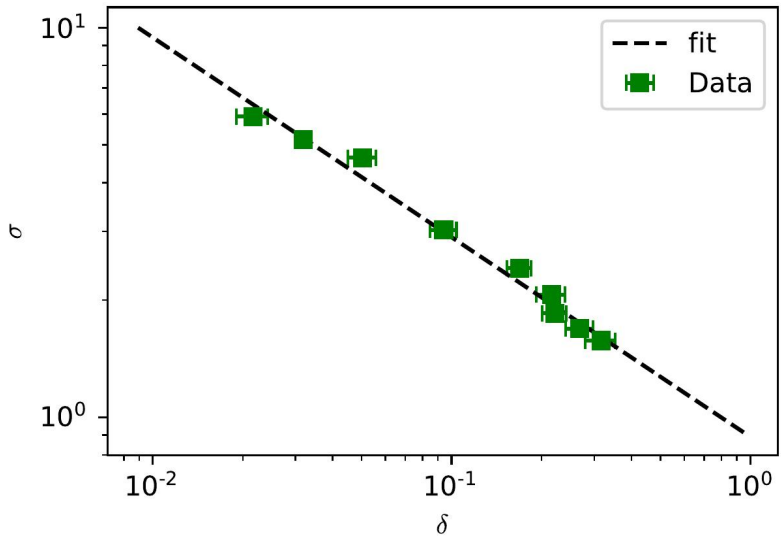

Figure IV.8: Coefficients $\sigma(T)$ as a function of $\delta(T)$ for all the temperatures considered here. We used a power law $\sigma = a\delta^b$ to fit the data.

IV.6 Summary and outlook

We define a finite-size two-body excess entropy $s_2(L)$ integral equation with L the linear size of the simulation box. Using analytical arguments and simulations of a prototypical Lennard-Jones liquid at different densities and temperatures, we show that $s_2(L) = s_2^\infty + \sigma/L$ with σ a constant that depends on intensive thermodynamic quantities. Given the well-know finite-size scaling of the

self-diffusivity, $D^*(L) = D^{*\infty} - \delta/L$, we show that the universal scaling relation between entropy and diffusion $D^* = A \exp(\alpha s_2)$ also exhibits a finite-size dependence and, by extrapolating to the TL, report $A = 0.048 \pm 0.001$ and $\alpha = 1.000 \pm 0.013$, in good agreement with values reported in the literature. Finally, and perhaps more interestingly, we show that the scaling coefficients σ and δ of s_2 and D^*, respectively, are related by a somewhat simple power law $\sigma = a\delta^b$ with $a = 1.256 \pm 0.118$ and $b = -0.513 \pm 0.020$.

The finite-size scaling of s_2 can be rationalised in terms of the thermodynamics of small systems [28, 44]. In particular, the statistical mechanics of a few model small systems in confinement has been derived recently [13]. The authors have shown that given the high surface area–to–volume ratio of small systems, thermodynamic properties include surface contributions. In the case of entropy, these contributions include $1/L$ terms with L, the linear size of the system. Since other transport properties, including thermal conductivity and viscosity, exhibit similar size effects and are related to s_2 via an exponential relation, we feel that the finite-size entropy scaling investigated here might play a role in understanding the non-equilibrium thermodynamics of confined, small systems [10].

The power law relation between the scaling coefficients of self-diffusion and two-body excess entropy is somewhat intriguing. On the one hand, the size scaling in the self-diffusion appears as a consequence of the conservation of linear momentum [64]. On the other hand, the finite-size scaling in the two-body entropy results from a surface contribution due to the confinement of the system [13]. Admittedly, we do not have a satisfactory explanation for this connection.

Nevertheless, we point out that the ratio $\delta^b/\sigma = 1/a$ might be related to a constant viscosity-to-entropy ratio. Indeed, δ is inversely proportional to the system's viscosity, and a simple dimensional analysis tells us that σ has units of entropy times length. Interestingly, string theory methods have been used to conjecture that, for fluids in equilibrium, the viscosity to entropy density ratio has a lower bound at $\hbar/4\pi k_\mathrm{B}$ [35] with $\hbar$ the reduced Planck constant. This relation, tested for various fluid systems [3, 22, 29], has been originally derived by considering that the entropy density of a black hole is proportional to the surface to volume ratio of its event horizon. We find this connection fascinating, and, in our opinion, it deserves further investigation.

Acknowledgments

We thank Kurt Kremer and Burkhard Dünweg for their insightful discussions. We also thank Denis Andrienko for his critical reading of the manuscript. R.C.-H. gratefully acknowledges funding from SFB-TRR146 of the German Research Foundation (DFG). Simulations have been performed on the THINC cluster at the Max Planck Institute for Polymer Research and the COBRA cluster at the Max Planck Computing and Data Facility.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Bibliography

[1] Manish Agarwal, Mohammad Parvez Alam, and Charusita Chakravarty. Thermodynamic, diffusional, and structural anomalies in rigid-body water models. *The Journal of Physical Chemistry B*, 115(21):6935–6945, 06 2011.

[2] Manish Agarwal, Murari Singh, Ruchi Sharma, Mohammad Parvez Alam, and Charusita Chakravarty. Relationship between structure, entropy, and diffusivity in water and water-like liquids. *The Journal of Physical Chemistry B*, 114(20):6995–7001, 05 2010.

[3] G.G.N. Angilella, N.H. March, F.M.D. Pellegrino, and R. Pucci. Proposed lower bound for the shear viscosity to entropy density ratio in some dense liquids. *Physics Letters A*, 373(10):992–998, 2009.

[4] J. A. Armstrong and P. Ballone. Computational verification of two universal relations for simple ionic liquids. kinetic properties of a model 2:1 molten salt. *The Journal of Physical Chemistry B*, 115(17):4927–4938, 05 2011.

[5] J. A. Armstrong, C. Chakravarty, and P. Ballone. Statistical mechanics of coarse graining: Estimating dynamical speedups from excess entropies. *The Journal of Chemical Physics*, 136(12):124503, 2012.

[6] A. Banerjee, M. K. Nandi, and S. M. Bhattacharyya. Validity of the rosenfeld relationship: A comparative study of the network forming ntw model and other simple liquids. *Journal of Chemical Sciences*, 129(7):793–800, 2017.

[7] Atreyee Banerjee, Shiladitya Sengupta, Srikanth Sastry, and Sarika Maitra Bhattacharyya. Role of structure and entropy in determining differences in dynamics for glass formers with different interaction potentials. *Physical review letters*, 113(22):225701, 2014.

[8] Andras Baranyai and Denis J. Evans. Direct entropy calculation from computer simulation of liquids. *Phys. Rev. A*, 40:3817–3822, Oct 1989.

[9] Sorin Bastea. Transport properties of dense fluid argon. *Phys. Rev. E*, 68:031204, Sep 2003.

[10] D. Bedeaux, S. Kjelstrup, and S. K. Schnell. *Nanothermodynamics. General Theory*. PoreLab Publisher, 2020.

[11] István Borzsák and András Baranyai. On the convergence of green's entropy expansion. *Chemical physics*, 165(2-3):227–230, 1992.

[12] Alexandru Botan, Virginie Marry, and Benjamin Rotenberg. Diffusion in bulk liquids: finite-size effects in anisotropic systems. *Molecular Physics*, 113(17-18):2674–2679, 2015.

[13] Vilde Bråten, Dick Bedeaux, Øivind Wilhelmsen, and Sondre Kvalvåg Schnell. Small size effects in open and closed systems: What can we learn from ideal gases about systems with interacting particles? *The Journal of Chemical Physics*, 155(24):244504, 2021.

[14] Jean-Louis Bretonnet. Self-diffusion coefficient of dense fluids from the pair correlation function. *The Journal of Chemical Physics*, 117(20):9370–9373, 2002.

[15] S. Chapman, T.G. Cowling, D. Burnett, and C. Cercignani. *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases*. Cambridge Mathematical Library. Cambridge University Press, 1990.

[16] Ravi Chopra, Thomas M. Truskett, and Jeffrey R. Errington. Excess entropy scaling of dynamic quantities for fluids of dumbbell-shaped particles. *The Journal of Chemical Physics*, 133(10):104506, 2010.

[17] Ravi Chopra, Thomas M Truskett, and Jeffrey R Errington. On the use of excess entropy scaling to describe the dynamic properties of water. *The Journal of Physical Chemistry B*, 114(32):10558–10566, 2010.

[18] R. Cortes-Huerto, K. Kremer, and R. Potestio. Communication: Kirkwood-buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *The Journal of Chemical Physics*, 145(14):141103, 2016.

[19] Jeppe C Dyre. Perspective: Excess-entropy scaling. *The Journal of chemical physics*, 149(21):210901, 2018.

[20] Mikhail Dzugutov. A universal scaling law for atomic diffusion in condensed matter. *Nature*, 381:137–139, 1996.

[21] Burkhard Dünweg and Kurt Kremer. Molecular dynamics simulation of a polymer chain in solution. *The Journal of Chemical Physics*, 99(9):6983–6997, 1993.

[22] G. Faussurier, S.B. Libby, and P.L. Silvestrelli. The viscosity to entropy ratio: From string theory motivated bounds to warm dense matter transport. *High Energy Density Physics*, 12:21–26, 2014.

[23] Guillaume Galliero, Christian Boned, and Josefa Fernández. Scaling of the viscosity of the lennard-jones chain fluid model, argon, and some normal alkanes. *The Journal of Chemical Physics*, 134(6):064505, 2011.

[24] Teena Goel, Chandra Nath Patra, Tulsi Mukherjee, and Charusita Chakravarty. Excess entropy scaling of transport properties of lennard-jones chains. *The Journal of Chemical Physics*, 129(16):164904, 2008.

[25] M. Heidari, K. Kremer, R. Potestio, and R. Cortes-Huerto. Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations. *Molecular Physics*, 116(21-22):3301–3310, 2018.

[26] Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto. Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: Revisiting the spatial block analysis method. *Entropy*, 20(4), 2018.

[27] Cecilia Herrero, Michela Pauletti, Gabriele Tocci, Marcella Iannuzzi, and Laurent Joly. Connection between water's dynamical and structural properties: Insights from ab initio simulations. *Proceedings of the National Academy of Sciences*, 119(21):e2121641119, 2022.

[28] T. L. Hill. *Thermodynamics of Small Systems*. Dover, 1963.

[29] U. Hohm. On the ratio of the shear viscosity to the density of entropy of the rare gases and H2,N2,CH4, and CF4. *Chemical Physics*, 444:39–42, 2014.

[30] J. J. Hoyt, Mark Asta, and Babak Sadigh. Test of the universal scaling law for the diffusion coefficient in liquid metals. *Phys. Rev. Lett.*, 85:594–597, Jul 2000.

[31] Jaehyeok Jin, Kenneth S Schweizer, and Gregory A Voth. Understanding dynamics in coarse-grained models: I. universal excess entropy scaling relationship. *arXiv preprint arXiv:2208.00078*, 2022.

[32] Gota Kikugawa, Shotaro Ando, Jo Suzuki, Yoichi Naruke, Takeo Nakano, and Taku Ohara. Effect of the computational domain size and shape on the self-diffusion coefficient in a Lennard-Jones liquid. *The Journal of Chemical Physics*, 142(2):024503, 2015.

[33] Gota Kikugawa, Takeo Nakano, and Taku Ohara. Hydrodynamic consideration of the finite size effect on the self-diffusion coefficient in a periodic rectangular parallelepiped system. *The Journal of Chemical Physics*, 143(2), 2015.

[34] Pascal Kordt, Thomas Speck, and Denis Andrienko. Finite-size scaling of charge carrier mobility in disordered organic semiconductors. *Phys. Rev. B*, 94:014208, Jul 2016.

[35] P. K. Kovtun, D. T. Son, and A. O. Starinets. Viscosity in Strongly Interacting Quantum Field Theories from Black Hole Physics. *Physical Review Letters*, 94(11):111601, 2005.

[36] Peter Krüger, Sondre K. Schnell, Dick Bedeaux, Signe Kjelstrup, Thijs J. H. Vlugt, and Jean-Marc Simon. Kirkwood-buff integrals for finite volumes. *J. Phys. Chem. Lett.*, 4(2):235–238, 2013.

[37] J. L. Lebowitz and J. K. Percus. Long-range correlations in a closed system with applications to nonuniform fluids. *Phys. Rev.*, 122:1675–1691, Jun 1961.

[38] J. L. Lebowitz and J. K. Percus. Thermodynamic properties of small systems. *Phys. Rev.*, 124:1673–1681, Dec 1961.

[39] G. X. Li, C. S. Liu, and Z. G. Zhu. Scaling law for diffusion coefficients in simple melts. *Phys. Rev. B*, 71:094209, Mar 2005.

[40] Xiaoguang Ma, Wei Chen, Ziren Wang, Yuan Peng, Yilong Han, and Penger Tong. Test of the universal scaling law of diffusion in colloidal monolayers. *Phys. Rev. Lett.*, 110:078302, Feb 2013.

[41] Marco Malvaldi and Cinzia Chiappe. Excess entropy scaling of diffusion in room-temperature ionic liquids. *The Journal of Chemical Physics*, 132(24):244502, 2010.

[42] Raymond D. Mountain and Harold J. Raveché. Entropy and molecular correlation functions in open systems. ii two- and three-body correlations. *The Journal of Chemical Physics*, 55(5):2250–2255, 1971.

[43] Manoj Kumar Nandi, Atreyee Banerjee, Shiladitya Sengupta, Srikanth Sastry, and Sarika Maitra Bhattacharyya. Unraveling the success and failure of mode coupling theory from consideration of entropy. *The Journal of Chemical Physics*, 143(17):174504, 2015.

[44] Andrea Puglisi, Alessandro Sarracino, and Angelo Vulpiani. Thermodynamics and Statistical Mechanics of Small Systems. *Entropy*, 20(6):392, 2018.

[45] Harold J. Raveché. Entropy and molecular correlation functions in open systems. i. derivation. *The Journal of Chemical Physics*, 55(5):2242–2250, 1971.

[46] F. L. Román, A. González, J. A. White, and S. Velasco. Fluctuations in the number of particles of the ideal gas: A simple example of explicit finite-size effects. *Am. J. Phys.*, 67:1149, 1999.

[47] F L Román, J A White, and S Velasco. Fluctuations in an equilibrium hard-disk fluid: Explicit size effects. *J. Chem. Phys.*, 107:4635, 1997.

[48] F.L. Román, J.A. White, A. González, and S. Velasco. *Theory and Simulation of Hard-Sphere Fluids and Related Systems*, chapter Ensemble Effects in Small Systems, pages 343–381. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[49] F. L. Román, J. A. White, A. González, and S. Velasco. Fluctuations in a small hard-disk system: Implicit finite size effects. *The Journal of Chemical Physics*, 110(20):9821–9824, 1999.

[50] Gustavo G Rondina, Michael C Böhm, and Florian Müller-Plathe. Predicting the mobility increase of coarse-grained polymer models from excess entropy differences. *Journal of Chemical Theory and Computation*, 16(3):1431–1447, 2020.

[51] Yaakov Rosenfeld. Relation between the transport coefficients and the internal entropy of simple liquids. *Phys. Rev. A*, 15(6):2545–2549, 1977.

[52] Yaakov Rosenfeld. A quasi-universal scaling law for atomic transport in simple fluids. *Journal of Physics: Condensed Matter*, 11(28):5415, 1999.

[53] M. Rovere, D. W. Hermann, and K. Binder. Block density distribution function analysis of two-dimensional lennard-jones fluids. *EPL*, 6(7):585, 1988.

[54] 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. *Phys. Rev. E*, 53:2382–2389, Mar 1996.

[55] Alok Samanta, Sk. Musharaf Ali, and Swapan K. Ghosh. Universal scaling laws of diffusion in a binary fluid mixture. *Phys. Rev. Lett.*, 87:245901, Nov 2001.

[56] Kazuhiko Seki and Biman Bagchi. Relationship between entropy and diffusion: A statistical mechanical derivation of rosenfeld expression for a rugged energy landscape. *The Journal of chemical physics*, 143(19):194110, 2015.

[57] Mauricio Sevilla and Robinson Cortes-Huerto. Connecting density fluctuations and kirkwood–buff integrals for finite-size systems. *The Journal of Chemical Physics*, 156(4):044502, 2022.

[58] Ruchi Sharma, Manish Agarwal, and Charusita Chakravarty. Estimating the entropy of liquids from atom–atom radial distribution functions: silica, beryllium fluoride and water. *Molecular Physics*, 106(15):1925–1938, 2008.

[59] Florian Spieckermann, Daniel Şopu, Viktor Soprunyuk, Michael B Kerber, Jozef Bednarčík, Alexander Schökel, Amir Rezvan, Sergey Ketov, Baran Sarac, Erhard Schafler, et al. Structure-dynamics relationships in cryogenically deformed bulk metallic glass. *Nature communications*, 13(1):1–9, 2022.

[60] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.*, 271:108171, 2022.

[61] Dario Villamaina and Emmanuel Trizac. Thinking outside the box: fluctuations and finite size effects. *Eur. J. Phys.*, 35(3):035011, 2014.

[62] Duane C. Wallace. On the role of density fluctuations in the entropy of a fluid. *The Journal of Chemical Physics*, 87(4):2282–2284, 1987.

[63] Michael Widom and Michael Gao. First principles calculation of the entropy of liquid aluminum. *Entropy*, 21(2):131, 2019.

[64] In-Chul Yeh and Gerhard Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *The Journal of Physical Chemistry B*, 108(40):15873–15879, 2004.

[65] Hayat Zaoui, Pier Luca Palla, Fabrizio Cleri, and Evelyne Lampin. Length dependence of thermal conductivity by approach-to-equilibrium molecular dynamics. *Phys. Rev. B*, 94:054304, Aug 2016.

V. Chilling alcohol on the computer: isothermal compressibility and the nucleation of hydrogen-bond clusters in liquid propan-1-o

Outline

Bibliographic Information

Luis A. Baptista, Mauricio Sevilla, Manfred Wagner, Kurt Kremer, and Robinson Cortes-Huerto. Chilling alcohol on the computer: isothermal compressibility and the formation of hydrogen-bond clusters in liquid propan-1-ol. *The European Physical Journal E*, 46(117), 11 2023

Author Contributions

Luis A. Baptista and Mauricio Sevilla have contributed equally to this work.

- Luis A. Baptista: Methodology; Formal analysis; Investigation; Writing – original draft.
- Mauricio Sevilla: Methodology; Formal analysis; Investigation; Writing – original draft;
- Manfred Wagner: Methodology; Formal analysis; Investigation;
- Robinson Cortes-Huerto: Conceptualization; Methodology; Formal analysis; Investigation; Funding acquisition; Writing – original draft; Writing – review & editing; Supervision.

Abstract

Molecular dynamics simulations have been performed to compute the isothermal compressibility κ_T of liquid propan-1-ol in the temperature range $200 \leq T \leq 300$ K. A change in behaviour, from normal (high T) to anomalous (low T), has been identified for κ_T at $210 < T < 230$ K. The average number of hydrogen bonds (H–bond) per molecule turns to saturation in the same temperature interval, suggesting the formation of a relatively rigid network. Indeed, simulation results show a strong tendency to nucleate H–bond clusters with distinct boundaries, with the average largest size and width of the size distribution growing upon decreasing temperature, in agreement with previous theoretical and experimental studies. These results also emphasise a connection between the behaviour of κ_T and the nucleation of nanometric structures.

V.1 Introduction

Hydrogen bonds (H–bonds) play a fundamental role in a wide range of phenomena in soft matter systems [39, 37], from the anomalous thermodynamic properties of water [32, 38], the solubility of (macro-)molecular synthetic and biological systems[6, 25] to the stabilisation and replication of DNA [48]. Collectively they are of significant relevance in supramolecular chemistry, where H–bonds are employed as one key concept to form (meta-)stable, functional polymer structures [1, 10]. The strength of H–bonds (4–8 $k_{\mathrm{B}}T$), with k_{B} being the Boltzmann constant and $T = 300$ K, lies between the strong covalent bonds (80 $k_{\mathrm{B}}T$ for a typical carbon-carbon bond) and the weak van der Waals interactions (less than $k_{\mathrm{B}}T$). Consequently, individual H–bonds can rapidly form and break, with characteristic lifetime $\tau \sim 10^{-11}$ s [34]. As a result of that intermediate (H)bond strength, directionality and rapid dynamics, many systems show a marked tendency to create networks that are highly temperature dependent [44, 45]. It is thus clear that, among others, the temperature dependence of H–bond interactions in molecular liquids is critical for achieving a fundamental understanding and developing engineered applications.

Indeed, thermoresponsive materials are now the focus of significant research efforts, given their encouraging range of applicability in various fields, including biology, energy, environment and materials science. A prominent example is given by smart, responsive polymers where even slight changes in the environment trigger substantial structural, stability and functional changes that can be tuned for advanced materials applications [29, 30] and pave the way for designing equivalent biocompatible materials [50]. Stimuli-responsive supramolecular assemblies also represent a case where a fine calibration of intermolecular interactions drives structure-bioactivity relationships [8]. Also, in the biological context, competing residue-osmolyte interactions are fundamental to understanding the denaturation of proteins [5]. Finally, thermoresponsive ionic liquid/water mixtures have been recently proposed as promising systems for applications in forward osmosis desalination, low-enthalpy thermal storage [14], and water harvesting from the atmosphere [15]. Common to all these phenomena and applications are their delicate corresponding temperature-dependant H–bond interactions.

To better understand the combined effect of temperature and H–bond interactions, it is advantageous to investigate relatively simple systems containing the essential chemical characteristics and structure common to more complex organic fluids. Monohydroxy alcohols represent a prototypical system to investigate H–bond networks due to tunable hydroxyl (OH-) group density resulting from increasing or decreasing the alkyl chain length. In these systems, two molecules form an H–bond

when the hydrogen in the OH group of one participates as the donor and the electronegative oxygen of the other as the acceptor. This indicates that a single OH group potentially forms up to three H–bonds, which results in molecular clusters exhibiting a wide variety of geometries [18, 44, 45, 2]. The dynamics properties of these H–bond structures have been extensively investigated in the literature. Dielectric response measurements in low-temperature propan-1-ol [20, 45, 46, 17, 49, 33, 2], for example, reveal that in addition to the α relaxation process present in other liquids and related to structural rearrangements of the alkyl chain, there is a lower frequency relaxation process (Debye). Based on NMR measurements, the latter is believed to be induced by the H–bond network [18]. Its characteristic relaxation time τ_D varies strongly between $\tau_D \approx 10^{-10}s$ close to $T = 330$ K and $\tau_d \approx 10^{0}s$ at about $T = 125$ K. However, to the best of our knowledge, no computational studies in the literature address the thermodynamic effects, isothermal compressibility in particular, resulting from nucleating H–bond clusters in the system as a function of the temperature. A possible reason for this gap might be that, at low temperatures, one expects to see nanometric-size structures that imply considerably large system sizes and an appropriate description of finite-size effects. An approach to overcome these problems for computing bulk thermodynamic quantities from simulations of molecular liquids has been just recently developed by some of us [23, 22, 43]. This approach we now apply to this problem.

Hence, in this work, we compute the isothermal compressibility κ_T for samples of 20000 propan-1-ol molecules in the liquid state in the temperature range $200 < T < 300$ K by molecular dynamics simulations. Following Ref. [18] τ_D, the longest structural relaxation time in the system is expected to vary between 10^{-7} s ($T = 200$ K) and about 3×10^{-10} s ($T = 300$ K). To account for that, all the the samples were equilibrated for 100 ns. Moreover, our relatively large samples (simulation boxes of linear size $L \sim 13$ nm) minimise finite-size effects. Nevertheless, the spatial block analysis (SBA) method [23, 22], that takes into account explicit and implicit size effects, had to be used to accurately compute κ_T in the thermodynamic limit. A change in the temperature dependence of κ_T, from normal (high T) to anomalous (low T) liquid behaviour, is apparent in the region $210 < T < 230$ K. Results from a hierarchical clustering method of H-bond networks in propan-1-ol reveal that the change in κ_T with decreasing temperature is related to the nucleation and growth of clusters of molecules connected via H–bonds.

V.2 Computational Details

Simulations of liquid propan-1-ol in the NPT ensemble have been performed using the OPLS-AA force field [26], which includes bonded and non-bonded interactions. The former account for stretching, bending and torsion contributions from covalent bonds, whereas the latter include dispersion, via Lennard-Jones potentials, and electrostatic interactions. A typical simulation snapshot, indicating four propan-1-ol molecules connected via hydrogen bonds (H–bonds), is presented in Fig. V.1. The sample occupies a cubic volume with applied periodic boundary conditions. Long-range electrostatic forces are treated via the smooth particle mesh Ewald algorithm [12].

Constant temperature and pressure conditions have been enforced using the velocity rescale thermostat [7] and the Parrinello-Rahman-Andersen barostat [3, 35], respectively. Simulations have been performed using the GROMACS 2019.6 [28] molecular dynamics simulation package with a time step of 2 fs.

Samples of 20000 propan-1-ol molecules have been quenched from $T = 300$ K down to $T = 240$

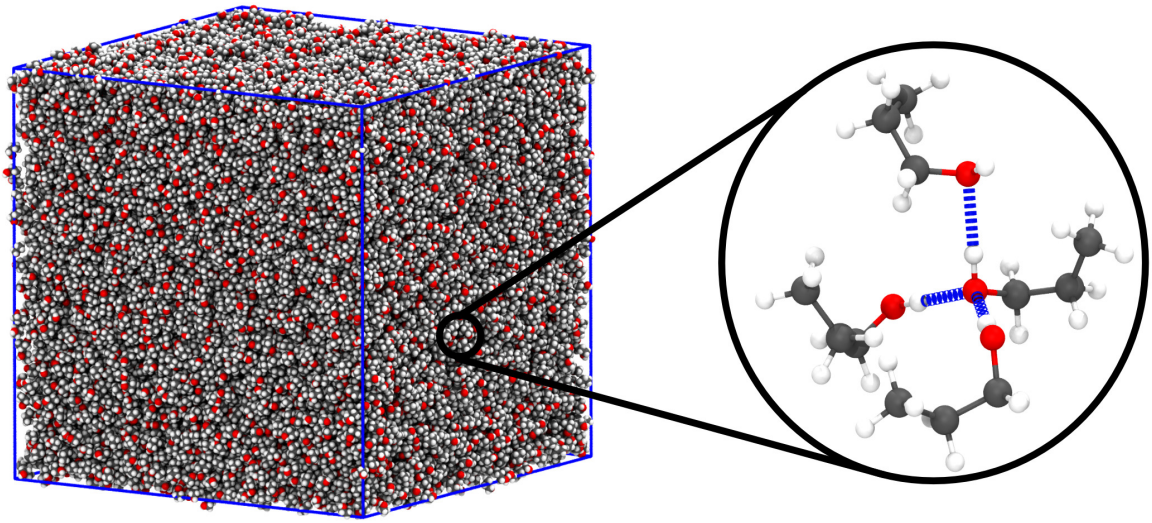

Figure V.1: Simulation snapshot of the liquid propan-1-ol at $T = 300$ K. A system containing 20000 molecules occupies a volume with linear size $L \sim 13$ nm. The detail shows four propan-1-ol molecules connected via H–bonds, represented as dashed blue lines. The colour code for the atoms is as follows: H, white; C, grey and O, red.

K in steps of $\Delta T = 20$ K, then from $T = 240$ K to $T = 200$ K in steps of $\Delta T = 10$ K. At each temperature, the system is equilibrated at NPT conditions for 100 ns, and statistics are accumulated, also at NPT conditions, for further 50 ns. The equilibration of the simulated samples has been controlled by monitoring the behaviour of the fluctuations of the number of particles for subdomains within the simulation box (See Right panel in Fig. V.3). The system's pressure is set to $P = 1$ bar for all temperatures.

V.3 Results and Discussion

The investigation of structural properties starts with calculating the oxygen–oxygen radial distribution function, $g(r)$, in the range of temperatures $200 \leq T \leq 300$ K. As expected, the first and second peaks of $g(r)$ weakly increase with decreasing temperature and the first minimum gets deeper with a marked shift towards smaller distances. See Fig. V.2 (Left panel). Still, in the whole temperature range, these $g(r)$ correspond to a liquid system, which is confirmed by the linear relation between average potential energy and temperature (result not shown).

Fluctuations in the long-range tail of the RDF encode information related to the thermodynamic properties of the liquid. This information becomes accessible from the simulation trajectory by computing the static structure factor, given by the expression:

$$S(k) = 1 + 4\pi\rho \int_0^\infty dr\, r^2 \frac{\sin(kr)}{kr} (g(r) - 1)\,, \tag{V.1}$$

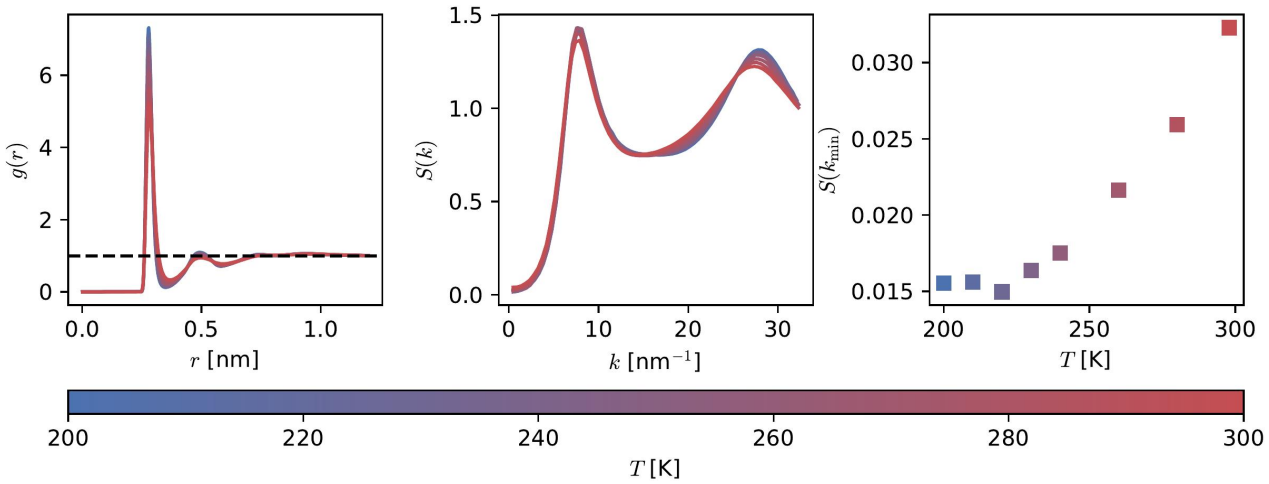

Figure V.2: (Left panel) Temperature dependence of the radial distribution function $g(r)$ and (Middle panel) structure factor $S(k)$. (Right panel) $S(k_{\min})$ vs T with $k_{\min} \approx 0.48$ nm^{-1} the minimum $k-$value in the reciprocal space discretisation.

with k being the norm of a reciprocal-lattice vector, and ρ the system's number density. To avoid numerical instabilities at the low-k region [42], typically the expression

$$S(\mathbf{k}) = \left\langle \frac{1}{N} \sum_{i'=1}^{N} \sum_{j'=1}^{N} \exp\left(-i\mathbf{k} \cdot (\mathbf{r}_{i'} - \mathbf{r}_{j'})\right) \right\rangle \tag{V.2}$$

is used, where the indices i, j refer to propan-1-ol oxygen atoms. The average runs over values of $k = |\mathbf{k}|$ and the statistical ensemble. To compute $S(k)$, we discretise the reciprocal space by considering vectors $\mathbf{k} = 2\pi(n_x, n_y, n_z)$ with n integer and a maximum value of $n_{\max} = 70$. The results are presented in the middle panel of Fig. V.2. The low height of the first peak in $S(k)$ confirms that the system remains liquid in the temperature range considered [21, 16]. It can be seen that upon decreasing temperature, the first (~ 7 nm^{-1}) and second (~ 27 nm^{-1}) peaks become weakly more pronounced. More interestingly, in the range $k < 2$ nm^{-1} (Right panel Fig. V.2), a change of behaviour becomes apparent with $\lim_{k\to 0} S(k)$ going from monotonically decreasing to a sudden increase starting at $T < 220$ K. The overall shape of $S(k)$, including the *anomalous* $k \to 0$ limit, is consistent with neutron diffraction measurements and related monte carlo simulations [45]. We recall that the limit $k \to 0$ of $S(k)$ is related to the isothermal compressibility κ_T via the expression

$$\lim_{k\to 0} S(k) = \rho k_{\mathrm{B}} T \kappa_T \,, \tag{V.3}$$

with k_{B} the Boltzmann constant. In practice, one evaluates $S(k_{\min})$ with $k_{\min}$ the minimum $k-$value in the reciprocal space discretisation, and the previous expression remains valid provided $k_{\min} \ll 2\pi/\xi$ with ξ being the system's correlations length. We present $S(k_{\min})/k_{\mathrm{B}}T\rho$ with $k_{\min} \approx 0.48$ nm^{-1}. Results can be seen in Fig. V.3 (Left panel).
Furthermore, we use the spatial block analysis (SBA) method to compute the isothermal compressibility [40, 41, 23, 22]. Inside a simulation box of volume V_0, fluctuations of the number of

molecules are computed for a subvolume V, using the expression

$$\chi_T(V;V_0) = \left.\frac{\langle N^2\rangle - \langle N\rangle^2}{\langle N\rangle}\right|_{V;V_0} . \tag{V.4}$$

The extrapolation of the isothermal compressibility in the thermodynamic limit κ_T is obtained by systematically varying the size of the subdomain in the interval $0 < V \leq V_0$, and applying the formula [23, 22]

$$\lambda\chi_T(\lambda) = \chi_T^\infty \left(1-\lambda^3\right)\lambda + \frac{\rho\alpha}{V_0^{1/3}} , \tag{V.5}$$

with $\lambda \equiv (V/V_0)^{1/3}$, $V < V_0$, α being an intensive proportionality constant and $\chi_T(\lambda) \equiv \chi_T(V;V_0)$. When $\lambda \ll 1$, $\lambda^3 \approx 0$ and $\chi_T^\infty = \rho k_\mathrm{B} T \kappa_T$ can be extrapolated from a linear fit to the simulation data.

An example of the use of the SBA method, Eq. (V.5), is given on the middle panel in Fig. V.3

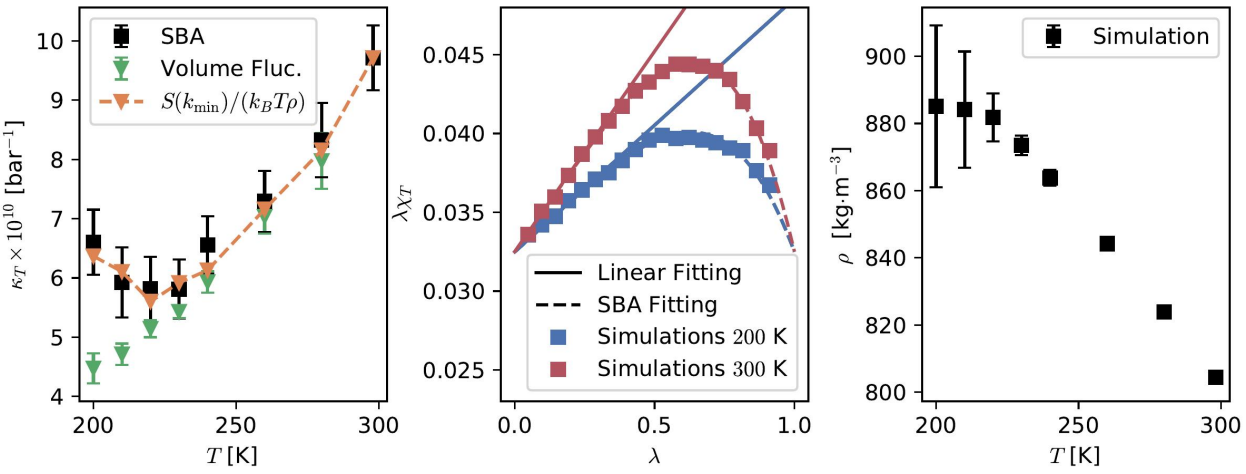

Figure V.3: (Left panel) Isothermal compressibility as a function of temperature. The change from normal to anomalous behaviour is apparent in the interval $210 < T < 230$ K when computing κ_T with the SBA method, Eq. (V.5) (Particle Fluctuations), in agreement with the behaviour exhibited by the $\lim_{k\to 0} S(k)$. By contrast, a direct calculation with Eq. (V.6) (Volume Fluctuations) fails to capture the correct behaviour at low temperatures. (Middle panel) Fluctuations of the number of particles $\chi_T(\lambda)$ as a function of the size of the subdomain $\lambda = (V/V_0)^{1/3}$ for two representative temperatures, $T = 200$ and 300 K. The slope in the linear fitting, Eq. (V.5) with $\lambda \ll 1$, gives the bulk isothermal compressibility $\chi_T^\infty = \rho k_\mathrm{B} T \kappa_T$. Using the obtained values for χ_T^∞ and α in Eq. (V.5), it is clear that the SBA method fits the whole simulation data set (SBA fitting). (Right panel) Density as a function of temperature.

for two representative temperatures $T = 200$ and 300 K. The overlap between the simulation data and the SBA fitting from Eq. (V.5) is, in our experience, a good indicator of the simulated samples being equilibrated [23]. The temperature dependence of κ_T is presented on the left panel in Fig. V.3. Reassuring the tendency observed with the $\lim_{k\to 0} S(k)$ (Fig. V.2), κ_T exhibits a deviation from the normal fluid behaviour visible in the interval $210 < T < 230$ K. The density as a function of temperature also indicates a change in the system's thermodynamics upon cooling

(See right panel in Fig. V.3). The linear monotonic behaviour observed at high temperature saturates at temperatures below $T = 230$ K. Interestingly, error bars increase systematically with decreasing temperature, signalling large fluctuations in the volume, compatible with an increase in κ_T.

We note in passing that a common practice in the literature corresponds to computing κ_T via the formula

$$\kappa_T = \frac{1}{k_B T \langle V_0 \rangle} \left(\langle V_0^2 \rangle - \langle V_0 \rangle^2 \right)_{\mathrm{NPT}} , \tag{V.6}$$

by evaluating volume fluctuations in simulations performed in the NPT ensemble. It has been reported that, to ensure convergence, this procedure requires significantly longer simulation times, especially at low temperatures [24]. For this reason, smaller systems (1000 propan-1-ol molecules, $L \sim 5$ nm) were simulated for ~ 1 μs to maximise the statistical sample size. The results presented on the left panel in Fig. V.3 indicate that using Eq. (V.6) gives markedly different results below $T = 230$ K. We interpret this discrepancy by noting that the volume fluctuations approach does not include any corrections regarding finite-size effects, which become more relevant at low temperatures. Consistent with this statement, $S(k)$ for these smaller samples do not exhibit the anomalous behaviour reported in Fig. V.2 (Data not shown).

In the temperature range considered here, there is no apparent divergence of κ_T, thus there is no indication of a possible liquid-liquid phase transition [47, 27]. However, the nucleation of long-range nanometric domains has been widely proposed as the reason behind the anomaly in the low-k region of $S(k)$ exhibited in various low-temperature liquids [13, 36, 11]. In a previous publication, some of us have found a similar anomaly in the structure factor in a supercooled binary mixture, identified with long-range concentration fluctuations [4]. However, for a single-component molecular system, the microscopic origin of the anomaly is not obvious. Therefore, in the following, this mechanism is discussed in more detail.

It has been proposed that, upon supercooling, monohydroxy alcohols form transient *polymeric* chains, connected via H–bonds, responsible for the intense dielectric response at low frequencies (Debye peak), observed in these systems [18, 44]. A simple count of H–bonds as a function of temperature, as seen in Fig. V.4, shows that the average number of H–bonds per molecule (n_H) increases monotonically upon decreasing temperature, tending to saturation at $T \sim 220$ K. This result emphasises the crucial role of H–bonds in dictating the structural properties observed. Indeed, at low temperatures $n_H \to 1$, consistent with the transient polymer model discussed in the literature [18, 44]. NMR chemical shift as a function of temperature (see right y-axis in Fig. V.4 and supporting information) indicates that hydrogen atoms in the hydroxyl group become more localised and increase their tendency to make directional bonds (with oxygen in a neighbouring -OH group) with decreasing temperature, in reasonable agreement with the simulation results. At this stage, it is still open whether the system nucleates nanostructures connected via H–bonds that increase their average size as temperature decreases.

To investigate this possibility, propan-1-ol molecules were classified into disjoint clusters with elements connected via H–bonds and a statistical analysis was performed based on the clusters' size distribution. We use a geometric definition of H–bonds, with donor–acceptor distance ≤ 0.35 nm and H–donor–acceptor angle ≤ 30 degrees. Donor and acceptor refer to O in propan-1-ol. The average number of H–bonds per molecule belonging to a cluster is ~ 2 at all temperatures. Snapshots of the ten largest clusters instantaneously present in a simulation are shown in Fig. V.5 at $T = 200$ K (Left panel) and $T = 300$ K (Right panel), which resemble the polymer chains expected

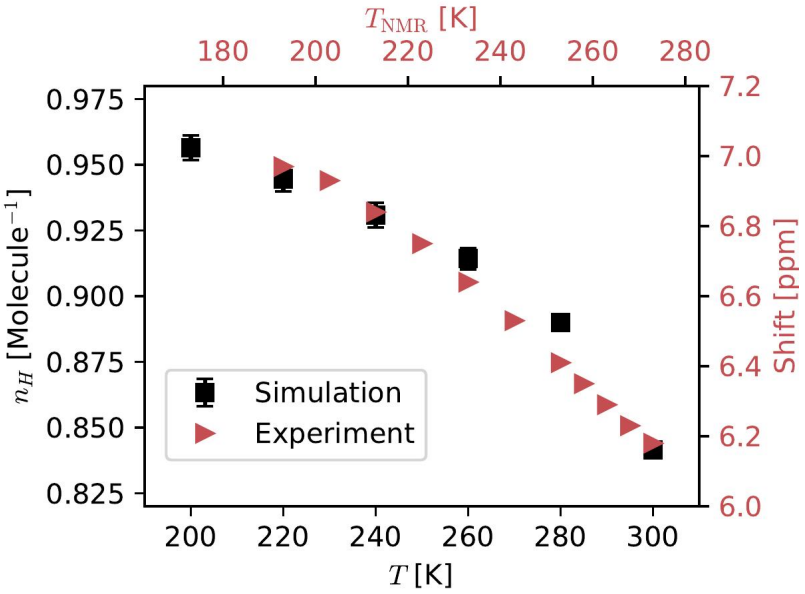

Figure V.4: $n_H = \langle N_H \rangle / N$ and NMR chemical shift vs temperature. $\langle N_H \rangle$ is the average total number of H–bonds in the system and N is the total number of molecules. The value of n_H is equivalent to counting all incoming H–bonds associated with O atoms and dividing it by twice the number of molecules. Simulation and experimental data show a rapid initial increase, followed by a turn to saturation at lower temperatures.

from experimental [18] and theoretical [44] studies. It is clear that at lower temperatures, clusters reach nanometric sizes ($\sim$ 300 molecules), larger than the clusters visible at room temperature ($\sim$ 80 molecules). Moreover, large fluctuations and the deviation of the monotonous behaviour of the density with temperature (right panel Fig. V.3) indicate that the density of the clusters is not the same as the density of the bulk liquid. These results are consistent with the transient-chain model [18], and agree well with statistical models and neutron diffraction experiments supporting the presence of clusters [44, 45, 46]. More interestingly, this cluster size dependence on temperature is systematic, with the mean largest-cluster size increasing upon decreasing temperature, as seen on the left panel in Fig. V.6. The cluster size distribution shown in the middle panel in Fig. V.6 agrees with this picture, with larger clusters appearing at lower temperatures. The cluster population is analysed using a hierarchical clustering method [31, 9, 19]. In particular, we compute the modularity of the H–bond network. Namely, given a proposed network division into *communities* (clusters), it measures the quality of the division quantified in terms of a large number of edges (H–bonds) within communities and only a few between them. The degree k_i of a vertex i in the system, i.e. the ith-oxygen atom, is defined as the number of incident H-bonds, that is

$$k_i = \sum_j A_{ij} \,, \tag{V.7}$$

with the matrix

$$A_{ij} = \begin{cases} 1 \text{ if atoms } (i,j) \text{ form a H–bond} \,, \\ 0 \text{ otherwise} \,. \end{cases} \tag{V.8}$$

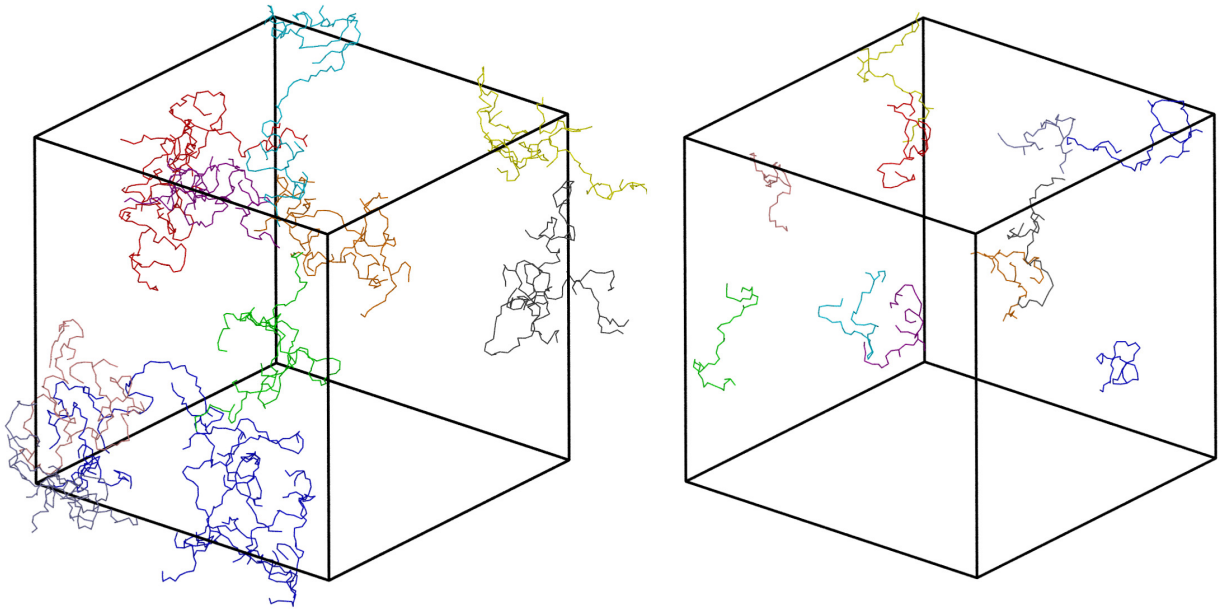

Figure V.5: Ten largest H–bond clusters instantaneously present in a simulation snapshot at $T = 200$ K (Left panel) and $T = 300$ K (Right panel). The clusters are highlighted using the same colour for propan-1-ol molecules belonging to a geometrically defined H–bond network. The linear size of the simulation box is $L \sim 13$ nm.

Finally, the modularity Q is defined as

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta_{c_i,c_j} \,, \tag{V.9}$$

with $m = \sum_{ij} A_{ij}/2$ a normalisation factor equal to the total number of edges, and δ_{c_i,c_j} a function with value one if the molecules (i, j) belong to the same cluster and zero otherwise. According to this definition, $Q = 0$ corresponds to a homogeneous, randomised network. In contrast, values in the range $0.3 < Q \leq 1.0$ represent the tendency for the network to show a large community structure [31, 9], namely, the presence of clusters of highly-interconnected molecules with well-defined boundaries. The result of Q as a function of temperature is shown on the right panel in Fig. V.6. It can be appreciated that even at high temperatures, the system has a marked tendency to form H–bond clusters with definite boundaries ($Q > 0.8$), with substantially growing size upon decreasing temperature.

V.4 Conclusions

In contrast to other organic liquids, low-temperature monohydroxy alcohols exhibit a simple dielectric behaviour identified with a Debye process, which results from the formation of stable H–bond structures. A vast body of theoretical and experimental literature has investigated these structures. However, to the best of our knowledge, a systematic investigation of temperature-dependent thermodynamic properties is somewhat lacking. Indeed, the nucleation of H–bond structures should

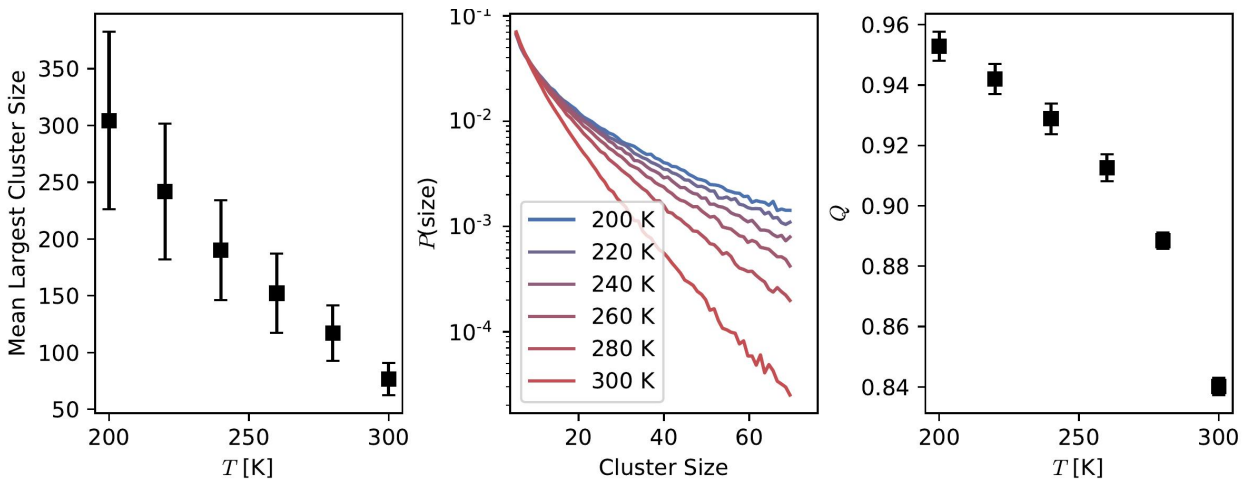

Figure V.6: (Left) Temperature dependence of the mean largest cluster size. (Middle) Density distribution –Log scale– of H–bond clusters as a function of their size. (Right) Modularity Q of the H–bond network vs T.

be reflected by the system's thermodynamics. To contribute to filling this gap, we compute the isothermal compressibility of propan-1-ol in the range of temperature $200 \leq T \leq 300$ K. Our results show that the system transition from a normal (high T) to an anomalous liquid (low T) at $210 < T < 230$ K, which can also be observed in the high-frequency region of the structure factor $S(k)$. We find that the average number of H–bonds saturate starting approximately in the same temperature range, suggesting that the network increases its rigidity at this temperature. By classifying the propan-1-ol molecules into disjoint H–bond clusters, we investigate their size distribution using the hierarchical clustering method. These results reveal that clusters with well-defined boundaries nucleate and increase their size upon decreasing temperature. The presence of these clusters is compatible with the behaviour of the isothermal compressibility κ_T. Indeed, κ_T is related to the limit $S(k \to 0)$ that indicates the presence of large scattering domains with increasing sizes upon decreasing temperature. These scattering domains have been identified with H–bond polymeric chains, consistent with previous experimental and theoretical results [18, 44].

acknowledgments

The authors thank various colleagues at the MPI-P for stimulating discussions, particularly Burkhard Dünweg, Kostas Daoulas, Nancy C. Forero-Martinez, Atreyee Banerjee, Katrin Amann-Winkel and Werner Steffen, and Robert Graf for his help with NMR measurements. R.C.-H. also thank Pietro Ballone for his insightful comments. L.A.B., K.K. and R.C.-H. thankfully acknowledge funding from SFB-TRR146 of the German Research Foundation (DFG). Simulations have been performed on the THINC cluster at the Max Planck Institute for Polymer Research and the COBRA cluster at the Max Planck Computing and Data Facility. We dedicate this work to Phil Pincus, who has been an inspiring colleague, mentor and friend for K.K. for more than 40 years.

Data Availability Statement

Data will be made available upon reasonable request.

Bibliography

[1] T. Aida, E. W. Meijer, and S. I. Stupp. Functional supramolecular polymers. *Science*, 335(6070):813–817, 2012.

[2] Antonela Ananiadou, George Papamokos, Martin Steinhart, and George Floudas. Effect of confinement on the dynamics of 1-propanol and other monohydroxy alcohols. *The Journal of Chemical Physics*, 155(18), 11 2021. 184504.

[3] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 07 2008.

[4] Atreyee Banerjee, Mauricio Sevilla, Joseph F. Rudzinski, and Robinson Cortes-Huerto. Finite-size scaling and thermodynamics of model supercooled liquids: long-range concentration fluctuations and the role of attractive interactions. *Soft Matter*, 18:2373–2382, 2022.

[5] Luis A. Baptista, Yani Zhao, Kurt Kremer, Debashish Mukherji, and Robinson Cortes-Huerto. Stabilizing -helicity of a polypeptide in aqueous urea: Dipole orientation or hydrogen bonding? *ACS Macro Letters*, 12(0):841–847, 0. PMID: 37318585.

[6] S Bekiranov, R Bruinsma, and P Pincus. Solution behavior of polyethylene oxide in water as a function of temperature and pressure. *Physical Review E*, 55(1):577–585, 1997.

[7] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

[8] Sarah Chagri, David Y. W. Ng, and Tanja Weil. Designing bioresponsive nanomaterials for intracellular self-assembly. *Nature Reviews Chemistry*, 6(5):320–338, May 2022.

[9] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70(6):066111, 2004.

[10] Tom F. A. De Greef, Maarten M. J. Smulders, Martin Wolffs, Albert P. H. J. Schenning, Rint P. Sijbesma, and E. W. Meijer. Supramolecular polymerization. *Chemical Reviews*, 109(11):5687–5754, 2009. PMID: 19769364.

[11] E. Eckstein, J. Qian, R. Hentschke, T. Thurn-Albrecht, W. Steffen, and E. W. Fischer. X-ray scattering study and molecular simulation of glass forming liquids: Propylene carbonate and salol. *The Journal of Chemical Physics*, 113(11):4751–4762, 2000.

[12] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 11 1995.

[13] E.W. Fischer. Light scattering and dielectric studies on glass forming liquids. *Physica A: Statistical Mechanics and its Applications*, 201(1):183–206, 1993.

[14] Nancy C. Forero-Martinez, Robinson Cortes-Huerto, Antonio Benedetto, and Pietro Ballone. Thermoresponsive ionic liquid/water mixtures: From nanostructuring to phase separation. *Molecules*, 27(5), 2022.

[15] Nancy C. Forero-Martinez, Robinson Cortes-Huerto, Lainey Ward, and Pietro Ballone. Water harvesting by thermoresponsive ionic liquids: A molecular dynamics study of the water absorption kinetics and of the role of nanostructuring. *The Journal of Physical Chemistry B*, 127(24):5494–5508, 2023. PMID: 37267503.

[16] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, volume 1 of *Computational Science Series*. Academic Press, San Diego, second edition, 2002.

[17] Jan Gabriel, Florian Pabst, and Thomas Blochowicz. Debye process and -relaxation in 1-propanol probed by dielectric spectroscopy and depolarized dynamic light scattering. *The Journal of Physical Chemistry B*, 121(37):8847–8853, 2017. PMID: 28872311.

[18] C. Gainaru, R. Meier, S. Schildmann, C. Lederle, W. Hiller, E. A. Rössler, and R. Böhmer. Nuclear-magnetic-resonance measurements reveal the origin of the debye process in monohydroxy alcohols. *Phys. Rev. Lett.*, 105:258303, Dec 2010.

[19] Yitian Gao, Hongwei Fang, and Ke Ni. A hierarchical clustering method of hydrogen bond networks in liquid water undergoing shear flow. *Scientific Reports*, 11(1):9542, 2021.

[20] C. Hansen, F. Stickel, T. Berger, R. Richert, and E. W. Fischer. Dynamics of glass-forming liquids. III. Comparing the dielectric - and -relaxation of 1-propanol and o-terphenyl. *The Journal of Chemical Physics*, 107(4):1086–1093, 07 1997.

[21] J. P. Hansen and I. McDonald. *Theory of Simple Liquids*. Academic, London, 1990.

[22] M. Heidari, K. Kremer, R. Potestio, and R. Cortes-Huerto. Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations. *Molecular Physics*, 116(21-22):3301–3310, 2018.

[23] Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto. Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: Revisiting the spatial block analysis method. *Entropy*, 20(4), 2018.

[24] C. Huang, K. T. Wikfeldt, T. Tokushima, D. Nordlund, Y. Harada, U. Bergmann, M. Niebuhr, T. M. Weiss, Y. Horikawa, M. Leetmaa, M. P. Ljungberg, O. Takahashi, A. Lenz, L. Ojamäe, A. P. Lyubartsev, S. Shin, L. G. M. Pettersson, and A. Nilsson. The inhomogeneous structure of water at ambient conditions. *Proceedings of the National Academy of Sciences*, 106(36):15214–15218, 2009.

[25] C Jeppesen and K Kremer. Single-chain collapse as a first-order transition: model for PEO in water. *Europhysics Letters*, 34(8):563, 1996.

[26] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[27] Kyung Hwan Kim, Alexander Späh, Harshad Pathak, Fivos Perakis, Daniel Mariedahl, Katrin Amann-Winkel, Jonas A. Sellberg, Jae Hyuk Lee, Sangsoo Kim, Jaehyun Park, Ki Hyun Nam, Tetsuo Katayama, and Anders Nilsson. Maxima in the thermodynamic response and correlation functions of deeply supercooled water. *Science*, 358(6370):1589–1593, 2017.

[28] Lindahl, Abraham, Hess, and van der Spoel. Gromacs 2019.6 source code, February 2020.

[29] Debashish Mukherji and Kurt Kremer. Smart polymers for soft materials: From solution processing to organic solids. *Polymers*, 15(15), 2023.

[30] Debashish Mukherji, Carlos M. Marques, and Kurt Kremer. Smart responsive polymers: Fundamentals and design principles. *Annual Review of Condensed Matter Physics*, 11(1):271–299, 2020.

[31] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.

[32] Anders Nilsson and Lars G. M. Pettersson. The structural origin of anomalous properties of liquid water. *Nature Communications*, 6(1):8998, Dec 2015.

[33] Florian Pabst, Jan Philipp Gabriel, Till Böhmer, Peter Weigl, Andreas Helbling, Timo Richter, Parvaneh Zourchang, Thomas Walther, and Thomas Blochowicz. Generic structural relaxation in supercooled liquids. *The Journal of Physical Chemistry Letters*, 12(14):3685–3690, 2021. PMID: 33829796.

[34] J.A. Padró, L. Saiz, and E. Guàrdia. Hydrogen bonding in liquid alcohols: a computer simulation study. *Journal of Molecular Structure*, 416(1):243–248, 1997. Hydrogen Bonding - Experimental and Theoretical Studies.

[35] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.

[36] A. Patkowski, Th. Thurn-Albrecht, E. Banachowicz, W. Steffen, P. Bösecke, T. Narayanan, and E. W. Fischer. Long-range density fluctuations in orthoterphenyl as studied by means of ultrasmall-angle x-ray scattering. *Phys. Rev. E*, 61:6909–6913, Jun 2000.

[37] Christine Peter and Kurt Kremer. Multiscale simulation of soft matter systems. *Faraday Discuss.*, 144:9–24, 2010.

[38] Lars Gunnar Moody Pettersson, Richard Humfry Henchman, and Anders Nilsson. Water—the most anomalous liquid. *Chemical Reviews*, 116(13):7459–7462, 2016. PMID: 27405667.

[39] P. Pincus. A survey of soft matter. *AIP Conference Proceedings*, 1420(1):115–127, 02 2012.

[40] F. L. Román, J. A. White, and S. Velasco. Block analysis method in off-lattice fluids. *EPL*, 42(4):371, 1998.

[41] F.L. Román, J.A. White, A. González, and S. Velasco. *Theory and Simulation of Hard-Sphere Fluids and Related Systems*, chapter Ensemble Effects in Small Systems, pages 343–381. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[42] Felix Sedlmeier, Dominik Horinek, and Roland R Netz. Spatial correlations of density and structural fluctuations in liquid water: A comparative simulation study. *J. Am. Chem. Soc,*, 133(5):1391–1398, 2011.

[43] Mauricio Sevilla and Robinson Cortes-Huerto. Connecting density fluctuations and Kirkwood–Buff integrals for finite-size systems. *The Journal of Chemical Physics*, 156(4):044502, 01 2022.

[44] Per Sillrén, Johan Bielecki, Johan Mattsson, Lars Börjesson, and Aleksandar Matic. A statistical model of hydrogen bond networks in liquid alcohols. *The Journal of Chemical Physics*, 136(9):094514, 03 2012.

[45] Per Sillrén, Jan Swenson, Johan Mattsson, Daniel Bowron, and Aleksandar Matic. The temperature dependent structure of liquid 1-propanol as studied by neutron diffraction and EPSR simulations. *The Journal of Chemical Physics*, 138(21):214501, 06 2013.

[46] P. Sillrén, A. Matic, M. Karlsson, M. Koza, M. Maccarini, P. Fouquet, M. Götz, Th. Bauer, R. Gulich, P. Lunkenheimer, A. Loidl, J. Mattsson, C. Gainaru, E. Vynokur, S. Schildmann, S. Bauer, and R. Böhmer. Liquid 1-propanol studied by neutron scattering, near-infrared, and dielectric spectroscopy. *The Journal of Chemical Physics*, 140(12):124501, 2014.

[47] RJ Speedy and CA Angell. Isothermal compressibility of supercooled water and evidence for a thermodynamic singularity at 45°C. *The Journal of Chemical Physics*, 65(3):851–858, 1976.

[48] J.L. Tymoczko, J.M. Berg, and L. Stryer. *Biochemistry: A Short Course*. W. H. Freeman, 2011.

[49] Peter Weigl, Daniel Koestel, Florian Pabst, Jan Philipp Gabriel, Thomas Walther, and Thomas Blochowicz. Local dielectric response in 1-propanol: -relaxation versus relaxation of mesoscale structures. *Physical Chemistry Chemical Physics*, 21(44):24778–24786, 2019.

[50] Y. Zhao, M. K. Singh, K. Kremer, R. Cortes-Huerto, and D. Mukherji. Why do elastin-like polypeptides possibly have different solvation behaviors in water-ethanol and water-urea mixtures? 53(6):2101–2110, 2020.

VI. Density Fluctuations, Solvation Thermodynamics and Coexistence Curves in Grand Canonical Molecular Dynamics Simulations

Outline

Bibliographic Information

Mauricio Sevilla, Luis A. Baptista, Kurt Kremer, and Robinson Cortes-Huerto. Density Fluctuations, Solvation Thermodynamics and Coexistence Curves in Grand Canonical Molecular Dynamics Simulations. *The Journal of Chemical Physics*, 162 (8): 080901, 02 2025

Author Contributions

- Mauricio Sevilla: Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal).
- Luis A. Baptista: Data curation (equal); Formal analysis (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal).
- Kurt Kremer: Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal).

§ Robinson Cortes-Huerto: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Software (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

Abstract

Fluid transport across nanometric channels induced by electric, pressure, and concentration gradients is ubiquitous in biological systems and fosters various applications. In this context, computer simulation setups with well-defined open-boundary equilibrium starting states are essential in understanding and assisting experimental studies. However, open-boundary computational methods are scarce and do not typically satisfy all the equilibrium conditions imposed by reality. Namely, in the absence of external gradients, (1) the system of interest (SoI) must be at thermodynamic and chemical equilibrium with an infinite reservoir of particles; (2) the fluctuations of the SoI in equilibrium should sample the grand canonical ensemble; (3) the local solvation thermodynamics, which is extremely sensitive to finite-size effects due to solvent depletion, should be correctly described. This point is particularly relevant for out-of-equilibrium systems; and (4) finally, the method should be robust enough to deal with phase transitions and coexistence conditions in the SoI. In this study, we demonstrate with prototypical liquid systems embedded into a reservoir of ideal gas particles that the adaptive resolution simulation (AdResS) method, coupled with particle insertion/deletion steps (AdResS+PI), satisfies all these requirements. Therefore, the AdResS+PI setup is suitable for performing grand canonical and stationary non-equilibrium simulations of open systems.

VI.1 Introduction

The study of fluid flow at the nanometric scale, i.e. nanofluidics [15], has recently gained considerable attention due to the prevalent role in biological systems [60, 19] and a broad range of applications [5], ranging from energy harvesting [17] to ionic computing [42]. Despite the highly developed experimental techniques, the interplay of confinement, non-equilibrium conditions, surface-specific interactions and quantum effects [27, 62] imposes significant challenges to the complete theoretical understanding of nanofluidics phenomena[47, 8, 28]. In this context, open-boundary molecular dynamics (MD) approaches [12, 13, 36, 35, 20, 61, 32, 24] open the possibility to accompany, design and interpret experimental efforts.

Nevertheless, the stationary non-equilibrium, open-boundary conditions present in experiments cannot be readily imposed on standard computational methods. Even the modest simulation of the corresponding equilibrium state must sample the grand canonical ensemble. The vast family of grand canonical Monte Carlo (MC) [1, 2, 34, 38, 49, 52, 50, 51, 18, 53, 16] and hybrid MD/MC [14, 33, 9, 54, 37, 10, 44, 6, 29] techniques, which rely upon particle insertion steps that become cumbersome when increasing the system's density and molecular complexity, are not mean to impose gradients on open-boundaries. Thus, developing and testing open-boundary computational techniques that sample the grand canonical ensemble in equilibrium is paramount. To this aim, we identify four essential conditions that open-boundary MD methods aiming at investigating non-equilibrium phenomena should fulfil:

Constant chemical potential - The molecular species in the system of interest (SoI) should be in thermal and chemical equilibrium with an infinite reservoir of particles. To the best of our knowledge, among the open-boundary MD methods available in the literature, only the adaptive resolution [3, 22, 24, 4] and the constant μ methods [39, 26] partially satisfy this condition since

the total size of the system limits the size of the reservoir under consideration. The next three requirements illustrate the relevance of the infinite size of the reservoir for open-boundary simulations.

Grand canonical ensemble - The fluctuations of the number of particles, $\Delta^2 N/\langle N\rangle$, in the SoI should sample the grand canonical ensemble. More specifically, given a single-component SoI of volume $V_0 \gg V_\zeta$, with V_ζ the volume define by the system's correlation length ζ, embedded into an infinite reservoir of particles, the quantity $\Delta^2 N/\langle N\rangle$ should satisfy the finite-size isothermal compressibility equation for every subvolume V in V_0 [43, 48]

$$\begin{aligned}\Delta^2 N/\langle N\rangle &= \chi_T(V;V_0)\\ &= 1+\frac{\rho}{(2\pi)^3 V}\int d\mathbf{k}\,\tilde{R}(\mathbf{k})\,\tilde{R}(-\mathbf{k})\,\tilde{h}^{\mathrm{PBC}}(\mathbf{k};V_0)\,.\end{aligned} \tag{VI.1}$$

In this expression, $\Delta^2 N \equiv (\langle N^2\rangle - \langle N\rangle^2)$ and $\langle N\rangle$ represents the average number of particles in the subvolume $V \leq V_0$ such that the average density is $\rho = \langle N\rangle/V$. In the asymptotic limit, $|\mathbf{r}| > \zeta$, the radial distribution function of the open system, $g(\mathbf{r})$, is connected to the one computed in the closed system, $g(\mathbf{r};N_{\mathrm{Tot}})$, by the relation $g(\mathbf{r};N_{\mathrm{Tot}}) = g(\mathbf{r}) - \rho k_{\mathrm{B}} T\kappa_T/N_{\mathrm{Tot}}$ with κ_T being the bulk isothermal compressibility [31, 30, 45, 46, 57, 11, 25, 21]. Hence, finite-size effects due to using different statistical ensembles can be easily identified by either considering a system with a finite total number of particles N_{Tot} (canonical) or including an infinite reservoir, i.e. $N_{\mathrm{Tot}} \to \infty$, (grand canonical). For simplicity, $h(\mathbf{r}) = g(\mathbf{r}) - 1$ is defined, and ensemble finite-size effects are corrected when introducing explicitly the asymptotic limit of the radial distribution function. The integral on the right-hand side of the equation is evaluated in Fourier space so that $\tilde{h}$ is the Fourier transform of h. The step function $R(\mathbf{r})$ accounts for the finite integration subdomain, and $\tilde{R}$ corresponds to its Fourier transform.

An additional advantage of evaluating the integral in Fourier space is that explicit finite-size effects due to the use of periodic boundary conditions (PBC) can be trivially introduced via

$$\tilde{h}^{\mathrm{PBC}}(\mathbf{k};V_0) = \sum_{n_x,n_y,n_z} \exp(-\mathbf{k}\cdot\mathbf{s}_{n_x,n_y,n_z})\tilde{h}(\mathbf{k})\,, \tag{VI.2}$$

with $\mathbf{s}_{n_x,n_y,n_z} = (n_x\,L_{0x}, n_y\,L_{0y}, n_z\,L_{0x})$ a vector specifying the system's periodic images such that $n_{x,y,z}$ takes integer values and $V_0 = L_{0x}\times L_{0y}\times L_{0z}$, where an orthorhombic simulation box is assumed. The choice $|n_x| \leq 1$, $|n_y| \leq 1$ and $|n_z| \leq 1$ is sufficient to compute Eq. (VI.1) accurately. To sum up, Eq. VI.1 is computed for a reference simulation and used as a benchmark. Fluctuations of the number of particles $\Delta^2 N/\langle N\rangle$ are computed for the SoI, and if the two results coincide for every V in V_0, the SoI samples the grand canonical ensemble. Moreover, in the limit $V \to V_0$, $\Delta^2 N/\langle N\rangle$ should converge to $\rho k_{\mathrm{B}} T\kappa_T$ + constant $\times\, \rho/\mathrm{V}_0^{1/3}$ in the grand canonical ensemble. By contrast, in the canonical ensemble $\Delta^2 N \to$ constant $\times\, \rho/\mathrm{V}_0^{1/3}$ [11, 25, 21].

Solvation thermodynamics - The simulation should be free of depletion effects that result from exhausting the number of particles in the reservoir [35]. In practice, these effects can be monitored by evaluating Kirkwood-Buff integrals (KBI). Let us consider a binary mixture of a and b particles

in a volume $V_0 \gg V_\zeta$, embedded into an infinite reservoir. For every subvolume $V \leq V_0$, the finite-size KBI, $G_{ij}(V;V_0)$ satisfies [48]

$$\begin{aligned} V\left(\frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\langle N_i\rangle\langle N_j\rangle} - \frac{\delta_{ij}}{\langle N_i\rangle}\right) &= G_{ij}(V;V_0) \\ &= \frac{1}{(2\pi)^3 V}\int d\mathbf{k}\,\tilde{R}(\mathbf{k})\,\tilde{R}(-\mathbf{k})\,\tilde{h}_{ij}^{\mathrm{PBC}}(\mathbf{k};V_0)\,, \end{aligned} \tag{VI.3}$$

with δ_{ij} the Kronecker delta, $\langle N_i\rangle$ the average number of particles in the volume V and the indices i,j indicate the type of particle a,b. As for the single component case, the grand canonical ensemble can be characterised in this equation by taking the limit for the total number of particles $N_{\mathrm{Tot}} \to \infty$ in the two-component radial distribution function [7].

In practice, a benchmark simulation allows the evaluation of the integral in Eq. (VI.3), which reduces to integrating the system's partial structure factors. Hence, a simulation samples the grand canonical ensemble if the fluctuations of the number of particles on the left-hand side of Eq. (VI.3) follows the solution of the integral on the right-hand side of the same equation evaluated as a benchmark. Furthermore, both results should converge to $G_{ij} + \mathrm{constant}/\mathrm{V}_0^{1/3}$ with G_{ij} being the KBI in the thermodynamic limit. Thus, the excess coordination number N_{ij}, i.e. the change in the number of i particles resulting from removing a j particle from the volume V_0, should be $\approx \rho_i G_{ij}$. Instead, the corresponding limit for a closed system is $\approx -\delta_{ij}$.

Coexistence and critical conditions - To investigate phase transformations in open molecular systems, it is crucial that the particle reservoir could be easily controlled to impose different thermodynamic conditions while maintaining thermal and chemical equilibrium with the SoI. We emphasise that near-critical conditions and second-order phase transitions have a limited description by any computational approach due to severe finite-size effects. Hence, let us explore the liquid-vapour coexistence region in the $T-\rho$ diagram of a single-component liquid. That means that the average density of the SoI should fluctuate between high-density (liquid) and low-density (vapour) values while keeping a constant chemical potential with the reservoir. From theoretical and practical perspectives, this point is crucial when simulating non-equilibrium conditions.

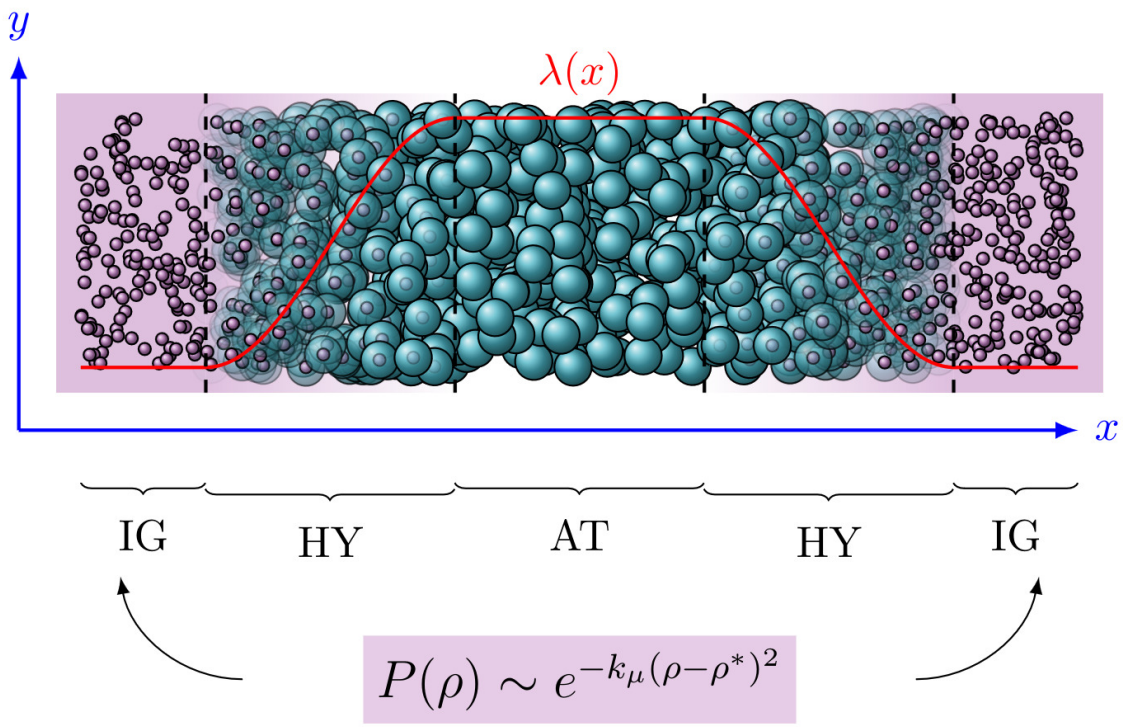

Figure VI.1: Schematic representation of the simulation setup. A slab geometry oriented along the x-axis with periodic boundary conditions applied in all directions. The switching field $\lambda(x)$ defines atomistic (AT), hybrid (HY) and ideal gas (IG) regions by taking values 1, $1 > \lambda(x) > 0$ and 0, respectively. A uniform density profile ensures the SoI has a constant chemical potential with the IG reservoir. The particle insertion/deletion algorithm is applied to the IG region, which is effectively in contact with an infinite particle reservoir. Periodic Boundary Conditions (PBC) are applied along x–, y– and z–directions.

In this paper, we show that the Adaptive Resolution Simulation Method (AdResS) with particle insertion/deletion steps (AdResS+PI) fulfils the conditions outlined above. To this aim, we simulate prototypical single-component and mixtures of Lennard-Jones systems embedded into infinite reservoirs of ideal gas particles that exhibit enough complexity (finite-size effects, non-trivial solvation behaviour, critical points, and a coexistence region) to test the AdResS+PI setup extensively.

VI.2 Model and Methods

The AdResS Hamiltonian [41, 40, 23] for a fluid composed of N molecules containing N_a atoms is

$$H_{[\lambda]}(r,p) = \mathcal{K} + V^{\text{intra}} + \sum_{\alpha=1}^{N} \{\lambda_\alpha V_\alpha + V^{\text{ext}}(\lambda_\alpha)\}\,, \tag{VI.4}$$

with (r,p) positions and momenta and $\mathcal{K} = \sum_{i=1}^{N_a} \mathbf{p}_i^2/2m_i$ being the total kinetic energy of the system. Latin indices run over atoms, and Greek indices over molecules. The potential

$V^{\text{intra}} = \sum_{\alpha=1}^{N} \sum_{i \neq j \in \alpha} V^{\text{intra}}(r_{ij})$ accounts for intra-molecular interactions, with r_{ij} being the separation between atoms i and j, which belong to the same molecule α. Intermolecular interactions are included in the potential term $V_\alpha = \frac{1}{2} \sum_{\beta \neq \alpha} \sum_{i \neq j} V(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|)$ with $\mathbf{r}_{\alpha i}$ being the position of the atom i in the molecule α. The molecules' resolution is determined by the switching field $\lambda_\alpha \equiv \lambda(\mathbf{R}_\alpha)$ with $\mathbf{R}_\alpha$ being the position of the centre of mass of the molecule α. When $\lambda = 0$ the Hamiltonian describes an homogeneous ideal gas system provided $V^{\text{ext}}(0) = \text{constant}$. In particular, we set $V^{\text{ext}}(0) = 0$. For $0 < \lambda \leq 1$, the Hamiltonian describes an inhomogeneous system, namely, an interacting system in the presence of an external field (Figure VI.1).

In the following, the SoI is defined as the region in the simulation box where $\lambda = 1$, i.e. the AT region in Fig. VI.1. By assuming that the SoI is in the grand canonical ensemble, namely, it is thermal and chemical equilibrium with an infinite reservoir of chemical potential μ^{id}, we have recently shown that the grand potential can be written as a functional of the density [4], i.e.

$$\begin{aligned} &\Omega_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})] = \\ &F_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})] + \int d\mathbf{r}\, \rho^{[\lambda]}(\mathbf{r}) \left(V^{\text{ext}}(\lambda(\mathbf{r})) - \mu(\lambda(\mathbf{r}))\right), \end{aligned} \tag{VI.5}$$

with the functional $F_{[\lambda]}[\rho^{[\lambda]}]$ being the intrinsic Helmholtz free energy corresponding to the Hamiltonian (VI.4), independent of the external potential. The condition that the SoI, i.e. the AT region in Fig. VI.1, is in thermal and chemical equilibrium with the reservoir can be imposed explicitly as $F_{[\lambda]}[\rho^{[\lambda]}] = F^{\text{id}}[\rho(\mathbf{r})]$ with F^{id} the Helmholtz free energy of the ideal gas. With this condition, the minimisation of the functional $\Omega_{[\lambda]}[\rho^{[\lambda]}(\mathbf{r})]$ implies

$$\rho(\mathbf{r}) = \rho^{\text{id}} \exp\left(-\beta\{V^{\text{ext}}(\lambda(\mathbf{r})) - \mu^{\text{exc}}(\lambda(\mathbf{r}))\}\right), \tag{VI.6}$$

with ρ^{id} being the density of the ideal gas. A useful condition, that guarantees that $\rho(\mathbf{r}) = \rho^{\text{id}}$ constant, is given by $V^{\text{ext}}(\lambda(\mathbf{r})) = \mu^{\text{exc}}(\lambda(\mathbf{r}))$. Conversely, by classical density functional theory, if a thermodynamic force is applied to the molecules instantaneously present in the HY region such that the average density is constant across the simulation box, minus the integral of the force is equal to the excess chemical potential of the SoI with respect to the ideal gas at the same density and temperature. These results show that the AdResS method satisfies the first requirement concerning a constant chemical potential between the SoI and the reservoir. Moreover, unlike standard grand canonical Monte Carlo, the AdResS method does not require the system's chemical potential μ as an input parameter. This is an advantageous feature since μ is typically a difficult quantity to compute.

The infinite reservoir conditions could be approximated by increasing the size of the IG region as much as possible. However, it is exact and computationally more efficient to introduce Metropolis Monte Carlo particle insertion/deletion steps in the ideal gas (AdResS+PI) such that the probability that the present density ρ changes by a quantity $\delta\rho$, is given by [24]

$$\text{acc}(\rho \to \rho \pm \delta\rho) = \min[1, \exp(-k_\mu \delta\rho(\delta\rho \pm 2(\rho - \rho^{\text{id}})))], \tag{VI.7}$$

with k_μ a free parameter related to the width of the distribution of possible values of ρ, with ρ^{id} being the reference density enforced in the IG region.

The remainder of the paper is devoted to using the AdResS+PI approach to investigate the three requirements described above to simulate open systems with a well-defined equilibrium state.

VI.3 Results

VI.3.1 Grand canonical ensemble

To prove that the simulation setup samples the grand canonical ensemble, it is sufficient to show that the fluctuations of the number of particles in the SoI follow Eq. VI.1. We compute the benchmark integral in Eq. (VI.1) for a liquid system whose potential energy is described by a 12-6, truncated and shifted, Lennard-Jones (LJ) potential with cutoff radius $r_c/\sigma = 2.5$. The parameters ϵ, σ and m define the units of energy, length and mass, respectively. The results are expressed in LJ units with time $\tau = \sigma(m/\epsilon)^{1/2}$, temperature ϵ/k_B and pressure ϵ/σ^3. A cubic box of linear size 30σ is considered with density fixed at $\rho\sigma^3 = 0.864$. The system is equilibrated at $k_BT/\epsilon = 2$, enforced by a Langevin thermostat with damping coefficient $\gamma(\sigma(m/\epsilon)^{1/2}) = 1$ for 15×10^7 steps using a time step of $\delta t/(\sigma(m/\epsilon)^{1/2}) = 0.01$. Production runs span 15×10^7 MD steps. All simulations were performed with an in-house version of the LAMMPS [56] simulation package.

By calculating the structure factor, it is thus possible to evaluate the integral on the right-hand side of Eq. (VI.1), correct the radial distribution function for ensemble effects and use the PBC term to extrapolate the result corresponding to the simulation box geometry of Fig. VI.1 with the volume V_0 defined as the volume of the AT region. We consider two orthorombic simulation boxes defined by the vectors (15σ,15σ,50σ) and (30σ,30σ,85σ) and with the AT region of dimensions given by (15σ,15σ,15σ) and (30σ,30σ,30σ), respectively. The results of the integration for the two simulation boxes as a function of the ratio $\nu = (V/V_0)^{1/3}$, i.e., fluctuations of the number of particles in the grand canonical ensemble, are presented as the solid blue curves in Fig. VI.2. The solid black curves represent the fluctuations of the number of particles in the canonical ensemble cases for the same geometries, which are shown for comparison. The important difference between the two curves corresponds to the region $\nu \approx 1$ or $V \to V_0$, where the two limiting cases discussed in the introduction are apparent.The thermodynamic limit $\chi_T^\infty = \rho k_B T \kappa_T$ is shown in Fig. VI.2 (dashed-dotted black line) to illustrate that $\lim_{V=V_0 \longrightarrow \infty} \chi(V; V_0) = \chi_T^\infty$

With this benchmark at hand, we compute the fluctuations of the number of particles in the AT region within the AdResS+PI framework described above. Starting from fully atomistic samples, hybrid (HY) and ideal gas (IG) regions are defined with linear sizes 15σ and σ, respectively. The HY region is uniformly discretized in slabs of size 0.5σ. To equilibrate the samples, the thermodynamic and drift forces needed to flatten the density profile across the simulation are updated every 10^4 steps during 7×10^7 steps. After 4×10^7 steps, the density profile is considered converged, and the integral of the final thermodynamic and drift forces gives the excess chemical potential of the LJ liquid at these conditions, $\mu_{exc}/\epsilon = 8.866 \pm 0.061$. At this stage, the particle insertion/deletion algorithm is applied every 1000 steps, and the simulation runs for further 10^7

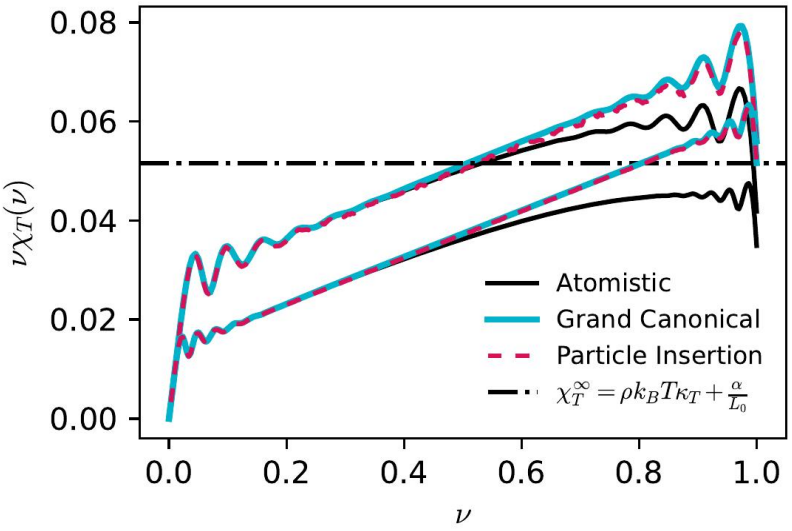

Figure VI.2: Fluctuations of the number of particles for two different simulation boxes defined by the vectors (15σ,15σ,50σ) and (30σ,30σ,80σ). The solid blue curves correspond to the solution of the grand canonical finite-size integral in Eq. (VI.1) for the small (upper) and the large (lower) samples. The solid black curves are the corresponding solution in the canonical ensemble. The horizontal dashed-dotted black line represents the thermodynamic limit $\rho k_{\mathrm{B}} T \kappa_T$. The fluctuations of the number of particles in the AT region obtained within the AdResS+PIframework are presented as dashed red curves.

steps for production results.

The result of computing $\Delta^2 N \equiv (\langle N^2 \rangle - \langle N \rangle^2)$ in the AT region for subvolumes $V \leq V_0$ is presented as the dashed red curve in Fig. VI.2 for the two simulation boxes considered. The AdResS+PIsimulation reproduces the benchmark solid blue curve, thus demonstrating that the AT region samples the grand canonical ensemble. As expected from the considerations in the introduction, the linear regime at intermediate values of ν converges to $\rho k_{\mathrm{B}} T \kappa_T$, represented by the horizontal dashed-dotted black line, plus a contribution $\propto 1/V_0^{1/3}$, which shifts down the curve for the larger system, approaching the thermodynamic limit value.

VI.3.2 Solvation thermodynamics

To investigate solvation thermodynamics, we focus on investigating a binary mixture (a, b) of a fluid whose potential energy is given by the purely repulsive truncated and shifted 12-6 LJ potential with cutoff radius $2^{1/6}\sigma$. The potential parameters are $\sigma_{aa} = \sigma_{bb} = \sigma_{ab} = \sigma$ and $\epsilon_{aa} = 1.2\epsilon$, $\epsilon_{bb} = 1.0\epsilon$ and $\epsilon_{ab} = (\epsilon_{aa} + \epsilon_{bb})/2$. As for the single-component case, the results are expressed in LJ units with energy ϵ, length σ, and mass $m_a = m_b = m$. A constant temperature of $k_{\mathrm{B}} T/\epsilon = 1.2$ is enforced by the Langevin thermostat, as before.

The benchmark has been computed for a system with a total number of atoms $N_{\mathrm{Tot}} = 24000$, and the range of mole fractions of a-molecules $x_a = 0.2, \cdots, 0.8$. The pressure is fixed at $P\sigma^3/\epsilon = 9.8$. After equilibration by alternating NVT-NPT runs, the final NVT production run corresponds to

a cubic simulation box with volume V_0 being the average volume in the last NPT run such that pressure does not deviate significantly from the target value. With the resulting partial structure factors, the integral on the right-hand side of Eq. (VI.3) can be evaluated for arbitrary geometries of the simulation box, giving the behaviour of the local solvation environment for every subdomain $V \leq V_0$. Furthermore, the thermodynamic limit value of the KBI, G_{ij} is obtained using the spatial block analysis method [11, 25, 21].

We compute the fluctuations of the number of particles in the AT region using the AdResS+PIsetup.

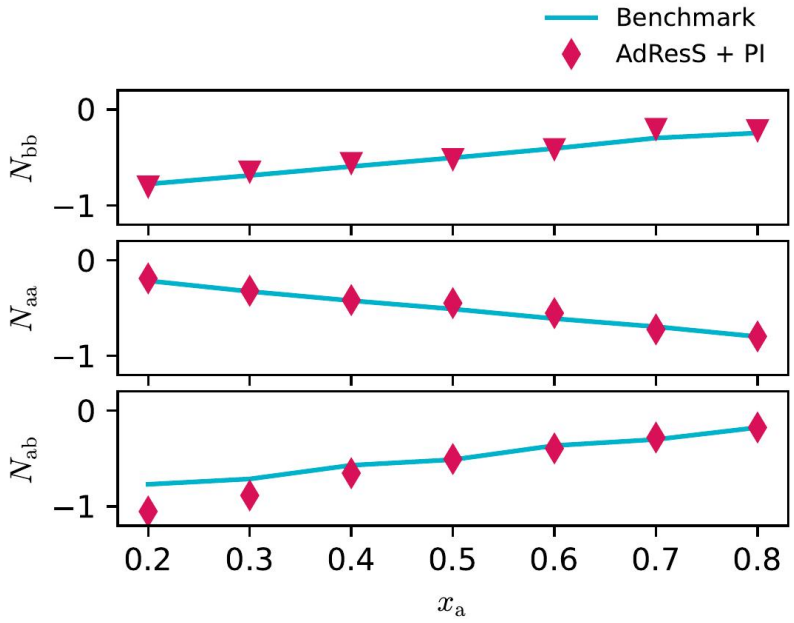

Figure VI.3: Excess coordination numbers $N_{ij} = \rho_i G_{ij}$ as a function of the concentration of a-molecules, x_a. The solid blue curves give the benchmark result. The red diamonds correspond to the result obtained by extrapolating the fluctuations of the number of particles in the AdResS+PIsimulation to evaluate G_{ij}.

Orthorhombic fully atomistic simulation boxes were equilibrated using the NPT-NVT cycles described above. Using the final sizes for these systems, the benchmark result of computing the integral in Eq. (VI.4) is obtained. AT and HY regions of linear sizes 20σ and 10σ, respectively, were defined. The remaining simulation parameters were taken from the previous single-component LJ liquid computations. The simulation converges when the values of the excess chemical potential for the a, b-molecules reproduce within the error bars the values reported in the literature ($\mu^{a,b}_{\text{exc}}/\epsilon = (15.627 \pm 0.124, 15.374 \pm 0.122)$ for $x_1 = 0.5$, for example). Following this step, the particle insertion/deletion algorithm is applied every 1000 steps, and the production run corresponds to further 2×10^8 steps.

In this case, it was also verified that the fluctuations of the number of particles as computed by the left-hand side of Eq. (VI.3) reproduce the benchmark result obtained by evaluating the right-hand side of the same equation (Result not shown), indicating that the AdResS+PIsimulation reproduces the local and long-range solvation behaviour expected in the grand canonical ensem-

ble. From the linear regime at intermediate values of ν, the value of G_{ij} is extrapolated. We present (Fig. VI.3) the excess coordination number N_{ij} as calculated from the benchmark result (solid blue) and the AdResS+PI(red diamonds) simulation. It is also apparent in this case that the two results agree well, indicating that the AdResS+PIframework is robust enough to describe the system's solvation thermodynamics accurately.

VI.3.3 Coexisting conditions

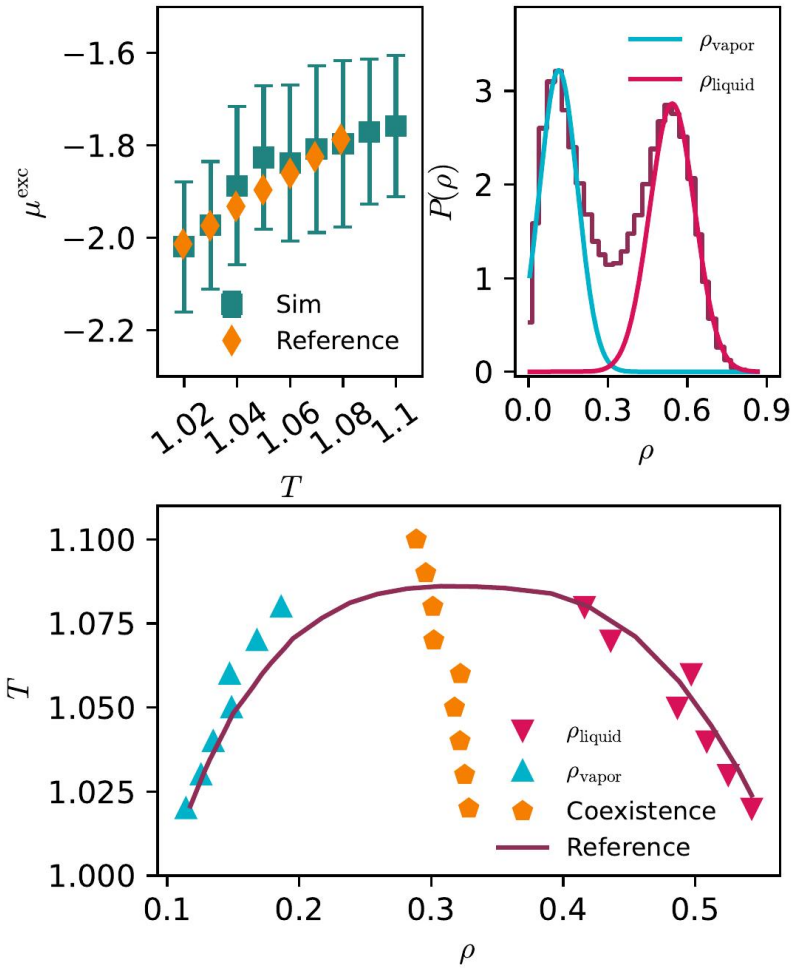

Figure VI.4: Critical point and coexistence conditions for the truncated and shifted LJ liquid. (a) Excess chemical potential μ_{exc} as a function of temperature: our simulation data (red diamonds) and rescaled reference data from Ref. [58, 59]. (b) Probability distributions for the density of the coexisting liquid (red) and vapor (blue) systems at $k_{\mathrm{B}}T/\epsilon = 1.02$. (c) Peak densities for the coexistent liquid (red) and vapor (blue) systems. The orange circles indicate the coexistence diameter. These data compare well with the results reported for the truncated and shifted LJ liquid [55].

Finally, we investigate the critical point transition and liquid-vapor coexistence conditions for the LJ system discussed in the first part of the manuscript. To this aim, we quenched the system

at the critical density $\rho_c\sigma^3 = 0.319$ in the range of temperatures $k_BT/\epsilon = 1.02, \cdots, 1.10$, which includes the critical temperature $k_BT_c/\epsilon = 1.086$ [55].

Results are presented in Fig. VI.4. First, the excess chemical potential as a function temperature across the coexistence line is calculated and compared with the values obtained with grand canonical Monte Carlo in Ref. [58, 59]. Even though the error bars in our calculation are close to 5%, the overall trend displayed by the reference data is well reproduced by the AdResS simulation. Here, we emphasize that, in contrast to the grand canonical Monte Carlo method, the chemical potential is an output of the AdResS method, resulting from imposing an initial reference density.

We have computed density fluctuations in the AT region (See Fig. VI.4(b)). Below the critical point, the probability distribution shows the well-known bimodal distribution obtained with grand canonical Monte Carlo simulations, showing the system's tendency to fluctuate between liquid and vapor states. For the temperature range considered, we identified the maxima of these distributions and plotted the result as a function of the system's average density. The result is presented in Fig. VI.4 (c) and compared with the grand canonical Monte Carlo results for the same LJ liquid as reported in Ref. [55]. It can be appreciated that the AdResS+PI results reproduce the reference data reasonably well.

VI.4 Conclusions

Confinement, surface-specific interactions, and quantum effects make nanofluidics an ideal framework to challenge our understanding of non-equilibrium statistical mechanics. Computational methods are crucial to investigating these systems, aiming to interpret and drive experimental studies. In this context, we argue that any open-boundary molecular dynamics simulation under the effect of external electric, density or concentration gradients should reduce to a well-defined equilibrium state when the external perturbation is not present. This state corresponds to the system of interest (SoI) being in chemical and thermal equilibrium with an infinite reservoir of particles. Hence, in equilibrium, the simulation setup must sample the grand canonical ensemble.

Specifically, for every subvolume V within the volume V_0 of the SoI, the fluctuations of the number of particles should reproduce the corresponding finite-size integral equations evaluated under grand canonical conditions. Moreover, in the case of multicomponent systems, the solvation structure for every $V \leq V_0$ should correspond to the finite-size Kirkwood-Buff integrals in the grand canonical ensemble. This condition ensures that the SoI is free from artefacts due to the depletion of particles in the reservoir. Finally, the simulation method should be robust enough to investigate the complexities of the phase diagram, such as near-critical conditions and coexistence curves.

We have simulated Lennard-Jones (LJ) liquids and mixtures in thermal and chemical equilibrium with an infinite reservoir of ideal gas particles using the AdResS method with particle insertion/deletion steps (AdResS+PI). In this setup, we have demonstrated that the fluctuations of the number of particles of the LJ systems sample the grand canonical ensemble for every subvolume $V \leq V_0$. Furthermore, we have investigated the near critical point and liquid-vapor coexistence conditions, and the results presented compare well with existing data. Therefore, we conclude

that this AdResS+PIframework is suitable for investigating equilibrium open molecular liquids and mixtures. The simplicity of the ideal gas reservoir becomes a tremendous advantage when imposing different conditions, such as pressure and concentration gradients and external potentials, that can be applied simultaneously on the SoI. These studies will be the subject of forthcoming papers.

acknowledgments

L.A.B., K.K. and R.C.-H. thankfully acknowledge funding from SFB-TRR146 of the German Research Foundation (DFG). K.K. and R.C.-H. thank Maziar Heidari and Raffaello Potestio for a fruitful collaboration that constitutes the basis of the present work. R.C.-H. also thank Debashish Mukherji, Nancy C. Forero-Martinez, Kostas Daoulas and Pietro Ballone for many stimulating discussions. Simulations have been performed on the THINC cluster at the Max Planck Institute for Polymer Research. **Data Availability Statement**

Data will be made available upon reasonable request.

Bibliography

[1] D.J. Adams. Chemical potential of hard-sphere fluids by monte carlo methods. *Molecular Physics*, 28(5):1241–1252, 1974.

[2] D.J. Adams. Grand canonical ensemble monte carlo for a lennard-jones fluid. *Molecular Physics*, 29(1):307–311, 1975.

[3] Animesh Agarwal, Han Wang, Christof Schütte, and Luigi Delle Site. Chemical potential of liquids and mixtures via adaptive resolution simulation. *J. Chem. Phys.*, 141(3):034102, 2014.

[4] L A Baptista, R C Dutta, M Sevilla, M Heidari, R Potestio, K Kremer, and R Cortes-Huerto. Density-functional-theory approach to the hamiltonian adaptive resolution simulation method. *Journal of Physics: Condensed Matter*, 33(18):184003, apr 2021.

[5] Nicholas AW Bell, Christian R Engst, Marc Ablay, Giorgio Divitini, Caterina Ducati, Tim Liedl, and Ulrich F Keyser. Dna origami nanopores. *Nano letters*, 12(1):512–517, 2012.

[6] Luc Belloni. Non-equilibrium hybrid insertion/extraction through the 4th dimension in grand-canonical simulation. *The Journal of Chemical Physics*, 151(2):021101, July 2019. _eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5110478/16756655/021101_1_online.pdf.

[7] A. Ben-Naim. *Molecular Theory of Solutions*. Oxford University Press, 2006.

[8] Lydéric Bocquet and Elisabeth Charlaix. Nanofluidics, from bulk to interfaces. *Chemical Society Reviews*, 39(3):1073–1095, 2010.

[9] Sharada Boinepalli and Phil Attard. Grand canonical molecular dynamics. *The Journal of chemical physics*, 119(24):12769–12775, 2003.

[10] Yunjie Chen and Benoît Roux. Efficient hybrid non-equilibrium molecular dynamics-monte carlo simulations with symmetric momentum reversal. *The Journal of Chemical Physics*, 141(11):114107, 2014.

[11] R. Cortes-Huerto, K. Kremer, and R. Potestio. Communication: Kirkwood-buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *The Journal of Chemical Physics*, 145(14):141103, 2016.

[12] Colin Denniston and Mark O. Robbins. Molecular and continuum boundary conditions for a miscible binary fluid. *Phys. Rev. Lett.*, 87:178302, Oct 2001.

[13] Colin Denniston and Mark O Robbins. General continuum boundary conditions for miscible binary fluids from molecular dynamics simulations. *The Journal of Chemical Physics*, 125(21), 2006.

[14] Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

[15] Theo Emmerich, Nathan Ronceray, Kumar Varoon Agrawal, Slaven Garaj, Manish Kumar, Aleksandr Noy, and Aleksandra Radenovic. Nanofluidics. *Nature Reviews Methods Primers*, 4(1):69, September 2024.

[16] Arman Fathizadeh and Ron Elber. A mixed alchemical and equilibrium dynamics to simulate heterogeneous dense fluids: Illustrations for lennard-jones mixtures and phospholipid membranes. *The Journal of Chemical Physics*, 149(7):072325, 2018.

[17] Jiandong Feng, Michael Graf, Ke Liu, Dmitry Ovchinnikov, Dumitru Dumcenco, Mohammad Heiranian, Vishal Nandigana, Narayana R Aluru, Andras Kis, and Aleksandra Radenovic. Single-layer mos2 nanopores as nanopower generators. *Nature*, 536(7615):197–200, 2016.

[18] Giovanni Garberoglio. Boltzmann bias grand canonical monte carlo. *The Journal of chemical physics*, 128(13):134109, 2008.

[19] Eric Gouaux and Roderick MacKinnon. Principles of selective ion transport in channels and pumps. *science*, 310(5753):1461–1465, 2005.

[20] Lin Guo, Shiyi Chen, and Mark O Robbins. Effective slip boundary conditions for sinusoidally corrugated surfaces. *Physical Review Fluids*, 1(7), 2016.

[21] M. Heidari, K. Kremer, R. Potestio, and R. Cortes-Huerto. Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations. *Molecular Physics*, 116(21-22):3301–3310, 2018.

[22] Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Spatially resolved thermodynamic integration: An efficient method to compute chemical potentials of dense fluids. *Journal of Chemical Theory and Computation*, 14(7):3409–3417, 2018. PMID: 29874069.

[23] Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Spatially resolved thermodynamic integration: An efficient method to compute chemical potentials of dense fluids. *Journal of Chemical Theory and Computation*, 14(7):3409–3417, 2018. PMID: 29874069.

[24] Maziar Heidari, Kurt Kremer, Ramin Golestanian, Raffaello Potestio, and Robinson Cortes-Huerto. Open-boundary Hamiltonian adaptive resolution. From grand canonical to non-equilibrium molecular dynamics simulations. *The Journal of Chemical Physics*, 152(19):194104, 05 2020.

[25] Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto. Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: Revisiting the spatial block analysis method. *Entropy*, 20(4), 2018.

[26] Tarak Karmakar, Aaron R. Finney, Matteo Salvalaglio, A. Ozgur Yazaydin, and Claudio Perego. Non-Equilibrium Modeling of Concentration-Driven processes with Constant Chemical Potential Molecular Dynamics Simulations. *Accounts of Chemical Research*, 56(10):1156–1167, May 2023. Publisher: American Chemical Society.

[27] Nikita Kavokine, Marie-Laure Bocquet, and Lydéric Bocquet. Fluctuation-induced quantum friction in nanoscale water flows. *Nature*, 602(7895):84–90, 2022.

[28] Nikita Kavokine, Roland R Netz, and Lydéric Bocquet. Fluids at the nanoscale: from continuum to subcontinuum transport. *Annual Review of Fluid Mechanics*, 53(1):377–410, 2021.

[29] Jeongmin Kim, Luc Belloni, and Benjamin Rotenberg. Grand-canonical molecular dynamics simulations powered by a hybrid 4D nonequilibrium MD/MC method: Implementation in LAMMPS and applications to electrolyte solutions. *The Journal of Chemical Physics*, 159(14):144802, 10 2023.

[30] J. L. Lebowitz and J. K. Percus. Long-range correlations in a closed system with applications to nonuniform fluids. *Phys. Rev.*, 122:1675–1691, Jun 1961.

[31] J. L. Lebowitz and J. K. Percus. Thermodynamic properties of small systems. *Phys. Rev.*, 124:1673–1681, Dec 1961.

[32] Yawei Liu, Raman Ganti, Hugh G. A. Burton, Xianren Zhang, Wenchuan Wang, and Daan Frenkel. Microscopic marangoni flows cannot be predicted on the basis of pressure gradients. *Phys. Rev. Lett.*, 119:224502, Nov 2017.

[33] B Mehlig, DW Heermann, and BM Forrest. Hybrid monte carlo method for condensed-matter systems. *Physical Review B*, 45(2):679, 1992.

[34] Mihaly Mezei. A cavity-biased (t, v, μ) monte carlo method for the computer simulation of fluids. *Molecular Physics*, 40(4):901–906, 1980.

[35] Debashish Mukherji and Kurt Kremer. Coil-Globule-Coil transition of PNIPAm in aqueous methanol: Coupling All-Atom simulations to Semi-Grand canonical Coarse-Grained reservoir. *Macromolecules*, 46:9158, 2013.

[36] Debashish Mukherji, Nico F. A. van der Vegt, and Kurt Kremer. Preferential solvation of triglycine in aqueous urea: An open boundary simulation approach. *J. Chem. Theory Comput.*, 8(10):3536–3541, 2012.

[37] Jerome P Nilmeier, Gavin E Crooks, David DL Minh, and John D Chodera. Nonequilibrium candidate monte carlo is an efficient tool for equilibrium simulation. *Proceedings of the National Academy of Sciences*, 108(45):E1009–E1018, 2011.

[38] AZ Panagiotopoulos. Exact calculations of fluid-phase equilibria by monte carlo simulation in a new statistical ensemble. *International Journal of Thermophysics*, 10:447–457, 1989.

[39] C. Perego, M. Salvalaglio, and M. Parrinello. Molecular dynamics simulations of solutions at constant chemical potential. *The Journal of Chemical Physics*, 142(14):144113, 04 2015.

[40] Raffaello Potestio, Pep Español, Rafael Delgado-Buscalioni, Ralf Everaers, Kurt Kremer, and Davide Donadio. Monte carlo adaptive resolution simulation of multicomponent molecular liquids. *Phys. Rev. Lett.*, 111:060601, Aug 2013.

[41] Raffaello Potestio, Sebastian Fritsch, Pep Español, Rafael Delgado-Buscalioni, Kurt Kremer, Ralf Everaers, and Davide Donadio. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.*, 110:108301, Mar 2013.

[42] P Robin, T Emmerich, A Ismail, A Niguès, Y You, G-H Nam, A Keerthi, A Siria, AK Geim, B Radha, et al. Long-term memory and synapse-like dynamics in two-dimensional nanofluidic channels. *Science*, 379(6628):161–167, 2023.

[43] F. L. Román, J. A. White, and S. Velasco. Block analysis method in off-lattice fluids. *Europhysics Letters*, 42(4):371, may 1998.

[44] Gregory A Ross, Ariën S Rustenburg, Patrick B Grinaway, Josh Fass, and John D Chodera. Biomolecular simulations under realistic macroscopic salt conditions. *The Journal of Physical Chemistry B*, 122(21):5466–5486, 2018.

[45] 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. *Phys. Rev. E*, 53:2382–2389, Mar 1996.

[46] J.J. Salacuse. Particle fluctuations within sub-regions of an n-particle, three-dimensional fluid: Finite-size effects and compressibility. *Physica A*, 387(13):3073 – 3083, 2008.

[47] Reto B Schoch, Jongyoon Han, and Philippe Renaud. Transport phenomena in nanofluidics. *Reviews of modern physics*, 80(3):839–883, 2008.

[48] Mauricio Sevilla and Robinson Cortes-Huerto. Connecting density fluctuations and kirkwood–buff integrals for finite-size systems. *The Journal of Chemical Physics*, 156(4):044502, 2022.

[49] JC Shelley and GN Patey. A configuration bias monte carlo method for ionic solutions. *The Journal of chemical physics*, 100(11):8265–8270, 1994.

[50] JC Shelley and GN Patey. A configuration bias monte carlo method for water. *The Journal of chemical physics*, 102(19):7656–7663, 1995.

[51] Wei Shi and Edward J Maginn. Continuous fractional component monte carlo: an adaptive biasing method for open system atomistic simulations. *Journal of chemical theory and computation*, 3(4):1451–1463, 2007.

[52] Berend Smit. Grand canonical monte carlo simulations of chain molecules: adsorption isotherms of alkanes in zeolites. *Molecular Physics*, 85(1):153–172, 1995.

[53] Mohammad Soroush Barhaghi, Korosh Torabi, Younes Nejahi, Loren Schwiebert, and Jeffrey J Potoff. Molecular exchange monte carlo: A generalized method for identity exchanges in grand canonical monte carlo simulations. *The Journal of chemical physics*, 149(7):072318, 2018.

[54] Harry A Stern. Molecular simulation with variable protonation states at constant ph. *The Journal of chemical physics*, 126(16):164112, 2007.

[55] Monika Thol, Gabor Rutkai, Roland Span, Jadran Vrabec, and Rolf Lustig. Equation of State for the Lennard-Jones Truncated and Shifted Model Fluid. *International Journal of Thermophysics*, 36(1):25–43, January 2015.

[56] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.*, 271:108171, 2022.

[57] Dario Villamaina and Emmanuel Trizac. Thinking outside the box: fluctuations and finite size effects. *Eur. J. Phys.*, 35(3):035011, 2014.

[58] Nigel B. Wilding. Critical-point and coexistence-curve properties of the lennard-jones fluid: A finite-size scaling study. *Phys. Rev. E*, 52:602–611, Jul 1995.

[59] Nigel B. Wilding. Computer simulation of fluid phase transitions. *American Journal of Physics*, 69(11):1147–1155, 11 2001.

[60] Gary Yellen. The voltage-gated potassium channels and their relatives. *nature*, 419(6902):35–42, 2002.

[61] Hiroaki Yoshida, Sophie Marbach, and Lydéric Bocquet. Osmotic and diffusio-osmotic flow generation at high solute concentration. II. Molecular dynamics simulations. *The Journal of Chemical Physics*, 146(19):194702, 05 2017.

[62] Xiaoqing Yu, Alessandro Principi, Klaas-Jan Tielrooij, Mischa Bonn, and Nikita Kavokine. Electron cooling in graphene enhanced by plasmon–hydron resonance. *Nature Nanotechnology*, 18(8):898–904, 2023.

Conclusions

Conclusions

Computer simulations are full of finite-size effects due to the unavoidable limitations in both time and length scales accessible. Regardless of the fact that advances in hardware and software have been increasing rapidly in the last decades, the differences between scales of theory and experiments with simulations are still of several orders of magnitude. Principally, the limitations come from the fact that thermodynamic experiments are in scales reaching the thermodynamic limit (TL), meaning that the number of particles and volume increase ($N \to \infty$ and $V \to \infty$) by keeping a ratio between them constant ($N/V = \rho$). Hence, it is necessary to develop tools, methods and strategies to extrapolate computer simulations to the thermodynamic limit, where all results should meet. Depending on the scale of interest, different methods can be used in order to provide meaningful results. As expected, to reach larger scales, a smaller resolution is required to balance out the cost/resources of the simulations. In this thesis, we focused on simulations of liquids where microscopic detail is relevant to extracting thermodynamic properties. This kind of simulation points immediately to several technical limitations: First, the number of atoms to be simulated is reduced (up to $\sim 10^5$). Second, the reachable ensembles are limited and normally different to experimental conditions and theoretical considerations, i.e. the TL. Third, when studying bulk properties, simulations have to be modified in order to eliminate boundary effects and reproduce the real bulk attributes. The way to solve this issue is by constructing simulations in a torus-like geometry, the so-called Periodic Boundary Conditions (PBCs), so as to technically have infinite simulations as artificial copies of the system are next to the boundaries.

This work combines three main pillars in order to connect computer simulations to the thermodynamic limit: Firstly, theoretical development and correct description of the mechanical statistics of small systems. Secondly, numerical methods and analysis tools to corroborate our theoretical results. And lastly, the applications to ideal and realistic systems for single-component liquids and binary mixtures. In the rest of this section, we will go through the different projects developed in the last years and presented in this document, highlighting their role in each one of the pillars previously mentioned.

Let us start with the two fisrt points, theoretical development and analysis tools. In the theory of liquids, the computation of thermodynamic quantities is done by understanding their local structures and correlations. We started with the Kirkwood-Buff Integrals (KBI), as they connect the microscopic structure with the macroscopic properties of liquids. It has been well established that the KBIs can be directly computed in the thermodynamic limit by extrapolating the local density fluctuations via the Spatial Block Analysis method. Supported by these results in chapter III, we developed an explicit expression combined with a numerical implementation to compute the KBIs for finite systems, allowing us to independently add or subtract all possible size effects, making it possible to provide a ground basis for further comparison with simulations with a specific

combination of size effects (Specifically, grand canonical simulations with contributions given by PBCs which will be discussed later). This was achieved by taking the corrections to $g(r)$ given by the ensemble differences as proposed by Cortes-Huerto et al. [http://dx.doi.org/10.1063/1.4964779] and going to the Fourier space to be able to include both, the specificities of the geometry of the boxes and the contributions given by the PBCs. Although this is not a method to compute the KBIs in the TL, it does allow us to describe the finite-size contributions coinciding with the measured fluctuations coming from the SBA. More importantly, it allowed us to artificially change the sizes and geometries of the systems studied while choosing which size effects to include or ignore. This made it possible for us to claim when an MD computer simulation correctly sampled the grand canonical ensemble.

The development and understanding of the correct finite size computation of the KBI allowed us to compute some other quantities that can also be obtained from integral equations. In particular, in different fields as in theory of glasses, the two-body excess-entropy s_2 is of great interest, hence in chapter IV we studied its computation and size effects. Opposed to the KBI, s_2 came to be free of any kind of ensemble effect. This is of special importance, as s_2 contains a term proportional to the KBI, which displays strong ensemble finite-size effects. Nevertheless, the direct implication of this lack of ensemble effect was that, to our knowledge, a finite-size version of s_2 was never reported before this work. After finding a size dependence of $1/L_0$ with the size of the simulation box, we tested the well-established empirical relation by Dzugutov, where a dimensionless diffusion coefficient D^* extrapolated to the TL relates with s_2 via a simple exponential scaling. By using the finite size version of both s_2 and D^*, we found that the Dzugutov coefficients also scale with size, such that after extrapolating to the thermodynamic limit, they coincide with those reported in the literature. Interestingly, the scaling coefficients for s_2 and D^* follow a power-law, suggesting an unexpected viscosity-to-entropy ratio.

On the side of applications to the thermodynamics of different systems, we have three projects involving ideal mixtures and a realistic set of simulations. First, we focused on understanding whether relatively small differences in the tail of the radial distribution functions result in contrasting thermodynamic properties of model supercooled liquids in chapter II. The connection between dynamics, which changes substantially upon supercooling, and structure, which seems to remain the same, is vital in understanding the supercooled state. We consider two different simple mixtures, a first one with Lennard-Jones (LJ) interactions and a second one including purely repulsive interactions or Weeks-Chandler-Andersen (WCA). The choice of these particular pairs of supercooled liquids comes as they have been reported to have somehow similar structures regarding the radial distribution functions but still considerably different dynamics. Our results contradict this statement since they show structural and thermodynamic information strongly depend on the tail of the correlation functions, especially under supercooling conditions, and they differ significantly in these two systems. In particular, we took two different density cases and different thermodynamic quantities, such as radial distribution functions, isothermal compressibilities, chemical potentials and partial structure factors and their limit of $k \to 0$. On the one hand, for the low density case ($\rho = 1.2\sigma^3$), computations of the density-density and concentration-concentration Bathia-Thornton partial structure factors, with their corresponding limit $k \to 0$ computed independently via the finite-size Kirkwood-Buff analysis, and the excess coordination number; allowed us to verify that there are significant differences in the structure, as the attractive interactions on the LJ model tend to nucleate long-range density domains, while the case of purely repulsive interactions WCA behaves like a normal liquid with monotonic behaviour of all

the quantities mentioned before. On the other hand, in the case of high density ($\rho = 1.6\sigma^3$) we found no considerable difference in the range of temperatures studied. This can be understood by the fact that at higher densities, the typical distances are smaller, making excluded volume-like interaction effects more prominent. Continuing with the computation of thermodynamic quantities, we moved to a realistic system in chapter V. In this work, we studied the isothermal compressibility of a pure propan-1-ol liquid in the range of $200 < T < 300$ K, finding a change in behaviour around $T \sim 220$ K. This change between a normal (high T) and a anomalous (low T) liquid, correspond to a similar behaviour to the one presented in chapter II where we identified nucleation of density/concentration domains in the mixtures studied. Yet, in this case, the nature of the domains required further investigation and analysis. First of all, we used different methods for the computation of the bulk isothermal compressibility, finding that the popular expression based on the volume fluctuations of NPT simulations requires long and rather large simulations in order to ensure convergence. Instead, the KBI calculation, based on the fluctuations in the number of particles, alongside the limiting case of the structure factor $S(\lim_{k\to 0} k)$, provides a reliable result for the computation of the bulk isothermal compressibility with the need of much less statistics provided the system is properly equilibrated. By studying the average number of hydrogen bonds (H-bond), we found that there is also a difference in behaviour around the same temperatures, suggesting the creation of structures driven by H-bond linking in our simulations. These results, consistent with the existing literature, were supported by NMR experiments. Lastly, we found that the anomalous behaviour comes from the fact that under cooling, propan-1-ol creates well-defined polymer-like clusters of molecules connected via H-bonds.

Lastly, we combine all the methods and results previously developed to demonstrate that the Adaptive Resolution with Particle Insertion method (`AdResS+ PI`) is able to sample the grand canonical ensemble. To get there, we start with the work developed in chapter I, where we consider a typical set-up of a `H-AdResS` simulation such that two different representations (Atomistic and Ideal Gas) coexist within the same simulation box allowing particles to change resolution *on-the-fly*. In order to guarantee a constant density profile, an additional force has to be imposed on the system. After assuming the grand canonical statistics, we showed that the Hamiltonian used in the `H-AdResS` is well defined and suitable to do statistical mechanics. This comes from a connection between adaptive resolution and classical density functional theory (DFT) resulting from identifying the system with an inhomogeneous liquid under an external potential. After considering the one-to-one correspondence between external potential and equilibrium density from DFT, the approach presented for `H-AdResS` comes naturally, where the thermodynamic force is included to ensure a constant density profile along the whole simulation box. Yet, a last question remains: Do our `AdResS+ PI` simulations properly describe the Grand Canonical (GC) ensemble? This question brings us to the last chapter of this thesis VI.

We showed that to ensure the GC description of our simulations is not enough to apply the `H-AdResS`, as the convergence of the fluctuation of the number of particles for large values of the subvolumes ($\nu \to 1$) require long and rather large simulations. Instead, providing a correct description of the GC statistics directly on the IG region, meaning, a particle insertion algorithm directly on the reservoir, provides a faster and more stable outcome. This takes advantage of the construction of the `H-AdResS`, as the particle insertion algorithm depends strongly of the particular system of study, but using an ideal gas representation of the system, immediately implies a specific statistics that can be deduced analytically, making even more general and robust the construction of the simulations. In other words, the choice of having an ideal gas representation fixes the

probability distribution for the particle insertion algorithm, and that remains the same even for different systems, as we always refer to the ideal gas.

With this result, we conclude this thesis, where we pushed forward the understanding and computation of finite-size thermodynamics from both analytical and computational points of view, a field of rapid development as the volume of data and resources available demand innovation in the procedures and practices to analyze simulations. We achieved this by providing clean and robust methodologies to bring computer simulations to the thermodynamic limit. Opening the doors to clear immediate applications as (a) the extension to the computation of thermodynamics of more complex systems, such as realistic systems and mixtures where the finite-size Kirkwood-Buff analysis is of crucial interest. (b) The connection of the excess of entropy with different transport properties in more complex systems, like binary mixtures (Maxwell-Stefan or Fick Diffusion) or systems with internal degrees of freedom (Thermal Conductivity). (c) Having the grand canonical formulation, i.e., a well-defined equilibrium state, allows us to consistently tackle (`AdResS+ PI`) simulations under non-equilibrium conditions.