Camel back shaped Kirkwood–Buff integrals

Aurélien Perera, Martina Požar, Bernarda Lovrinčević

Camel back shaped Kirkwood-Buff Integrals

Aurélien Perera[†], Martina Požar[‡] and Bernarda Lovrinčević[‡]

February 10, 2022

[†]Laboratoire de Physique Théorique de la Matière Condensée (UMR CNRS 7600), Sorbonne Université, 4 Place Jussieu, F75252, Paris cedex 05, France.

[‡]Department of Physics, Faculty of Science, University of Split, Ruđera Boškovića 33, 21000, Split, Croatia.

Abstract

Some binary mixtures, such as specific alcohol-alkane mixtures, or even water-tbutanol, exhibit two humps "camel back" shaped KBI. This is in sharp contrast with usual KBI of binary mixtures having a single extremum. This extremum is interpreted as the region of maximum concentration fluctuations, and usually occurs in binary mixtures presenting appreciable micro-segregation, and corresponds to where the mixture exhibit a percolation of the two species domains. In this paper, it is shown that two extrema occur in binary mixtures when one species forms "meta-particle" aggregates, the latter which act as a meta-species, and have their own concentration fluctuations, hence their own KBI extremum. This "meta-extremum" occurs at low concentration of the aggregate-forming species (such as alcohol in alkane), and is independant of the other usual extremum observed at mid volume fraction occupancy. These systems are a good illustration of the concept of the duality between concentration fluctuations and micro-segregation.

1 Introduction

The so-called Kirkwood-Buff integrals (KBI)[1, 2] are defined as the integrals of the species-species pair correlation functions$G_{ab} = \frac{1}{\Omega}\int d\mathbf{X}_1 d\mathbf{X}_2 \left[g_{ab}(\mathbf{X}_1, \mathbf{X}_2) - 1\right]$ where $\mathbf{X}_i$ is the set of position, and if required, the orientations, of molecule i, where a and b designates the species indexes ($\Omega = V\omega^2$, where ω is the angular integral, equal to 4π or $8\pi^2$, depending on the symmetry of the molecules). It can be shown [3, 4] that this integral is identical to that involving the pair correlation between any two atoms belonging to each molecules. Following this, the KBI are more simply defined as

$$G_{ab} = 4\pi \int_0^\infty dr r^2 \left[g_{i_a j_b}(r) - 1\right] \tag{1}$$

where $i_a$and j_b designate any two atoms i and j of , respectively, species a and b, and $g_{i_a j_b}(r)$ the atom-atom pair correlation function.The Kirkwood-Buff theory relates specific thermodynamic properties, such as the compressibility, or the partial molar volumes, for example, to be related to these integrals, thus providing an appealing link between macroscopic measurable properties to the microscopic structure, albeit in an integrated form, where all microscopic details are lost. Ben-Naim [5], Matteoli and Lepori [6] and other authors [7, 8, 9, 10, 11] have managed to invert these relations, providing a way to calculate these integrals from the experimentally available data on compressibility, partial molar volume and vapour pressure or Gibbs free-energy. The corresponding expressions for a binary mixture (a, b) are well known [12]:

$$G_{aa} = G_{ab} + \frac{1}{x_b}\left[\frac{\bar{V}_b}{D} - V\right] \tag{2}$$

$$G_{ab} = \frac{\chi_T}{k_B T} - \frac{\bar{V}_a \bar{V}_b}{VD} \tag{3}$$

were x_b is the mole fraction of component b, $\bar{V}_c$ is the partial molar volume of component c ($c = a$ or b), V is the molar volume, $\chi_T$is the isothermal compressibility (with T the temperature and k_B the Boltzmann constant) and D is related to the concentration fluctuations and given by

$$D = x_i \left(\frac{\partial \beta \mu_i}{\partial x_i}\right)_{TP} \tag{4}$$

where μ_i is the chemical component of species i and $\beta = 1/k_B T$ is the Boltzmann factor.

Fig.1 of the seminal Matteoli and Lepori paper [6] displays KBI for various types of aqueous mixtures, and it can be seen that most of them have a single extremum, which is positive for like KBI G_{aa} and negative for then unlike G_{ab} with $(a \neq b)$. This extremum has its origin in the shape of the coefficient D above, which is often a U-like shaped curve with a single minimum. An example of such typical KBI and D are given in Fig.1, and other similar examples can be found in various place in the literature such as [13, 14, 15, 16, 12]. Note that, while for a given mixture, the D coefficient has a single extremum, the corresponding extrema in the KBI are not necessarily at the same x-position, as illustrated in Fig.1, as Eqs.(2,3) tend to alter this position from that of D.

A single extremum for D is expected for a curve which reaches the same value at both end points. In fact, Eq.(4) can be shown (see Appendix) to be related to the stability limit criteria for a binary mixture, that is precursor for demixing, namely

$$D \propto C_{11} C_{22} - C_{12}^2 \tag{5}$$

where $C_{ij} = \delta_{ij} - \sqrt{\rho_i \rho_j}\tilde{c}_{ij}(k = 0)$, with the $\tilde{c}_{ij}(k = 0)$ being the integrals of the direct correlation functions and i,j refer to species index.Indeed, when $D \to 0$ the mixture becomes mechanically unstable [3], such when approaching

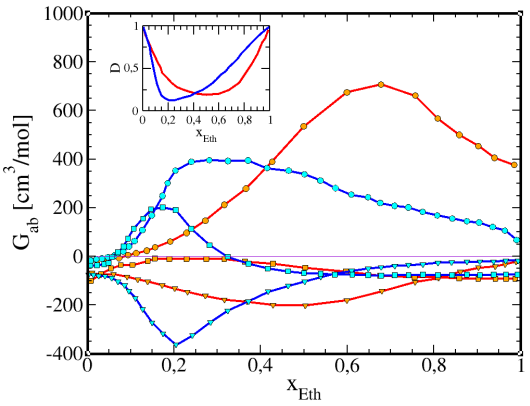

Figure 1: Examples of single extremum KBI. Red curves experimental KBI curves for aqueous-acetone from Ref.[17]; blue curves for aqueous-2propanol from Ref.[12]. Circles for Water-water KBI, squares for cross water-solute KBI, and triangles for solute-solute KBI. The inset shows the corresponding D (Eq.(4)) with same color codes, both showing a single minimum.

a spinodal or a critical point. Therefore, a single extremum in D can be physically interpreted as the point where concentration fluctuations are maximal, and with the eventuality that when these grow beyond the mechanical stability, the mixture undergoes demixing. Subsequently, the extremum in the KBI is a signature of maximal concentration fluctuations[18] . Since many binary mixtures, such as aqueous mixtures, for example, show stable local nano-segregation of constituents [19, 20, 21, 22, 23], the extremum can be interpreted as maximal amplitude in local segregation[18].

In this context, it came as a surprise that Jan Zielkiewicz published experimental KBI for various alcohols-heptane mixtures [24], which clearly showed the existence of two extremum. In Fig.2 of his paper, he shows KBI written as $\rho_a G_{ab}$ where $\rho_a = N_a/V$ is the partial density of species a (N_a is the number of particle in volume V), which amplifies the second extremum, the resulting KBI looking like "camel back" shaped, as opposed to the usual single extremum KBI which are "dromadary camel back" shaped.

The purpose of this paper is to understand the origin of this dual extremum in terms of fluctuations. The driving idea is the following. It is well known that alcohols generally tend to cluster their hydroxyl groups into linear patterns, forming chains, loops, etc... [25, 26, 27, 28, 29, 30] When put in small concentration in an inert alkane solvent, such as heptane, for example, small "chain-micelles" of alcohols form and act as independant meta-particles. Herein, we will adopt the convention of using the wording "meta" to designate local aggregated assemblies which act as micelle-like objects, and cannot be considered as mere local concentration fluctuations. Consequently, such exhibit their own concentration fluctuations, which corresponds to one of the observed extrema

in the corresponding KBI. With increasing concentration, the frontier between well separated micelles and alcohol micro-segregated domains become loose, and the mixture becomes a simple micro-segregated mixture, which has its own independent maximum in concentration fluctuations. It is not obvious that such scenario should happen, and the purpose of this paper is to use computer simulation to provide evidence that this scenario is correct, principally through the analysis of pair correlation functions and the KBI.

2 Theoretical and simulation details

The study of equilibrium locally micro-segregated mixtures has proven to be very difficult in the past decades, particularly in the case of aqueous mixtures. Perhaps a canonical example is that of aqueous *tert*-butanol (TBA) mixtures, for which the force field induced slow demixing at very small TBA concentrations around $x_{\mathrm{TBA}} \approx 0.01$ was shown [31, 32, 33, 34, 35] to rule out many classically robust force field models such as OPLS [36, 37] or TraPPE [38, 39]. One may even consider this problem to remain unsolved. Similarly, the water-tetrahydrofuran mixture remains difficult to simulate, in particular for putting into evidence the loop coexistence phase diagram [40, 41, 42]. One of the principal obstacle for successfully simulating such mixtures is the fact that both water and the alcohol tend to form hydrogen bonded clusters, and it would seem that it is precisely the competition between these cluster formation which tend to drive the simulated mixture into demixing, often after unusually very long times.

These problems might be avoided if water is replaced by a more inert solvent such as an alkane. In a previous work [43], we have studied ethanol-benzene mixtures and found that the description of the strong micro-heterogeneity (MH) in the low ethanol content region posed statistical issues which necessitated large size simulations. We expect to find similar issues in the present study.

Herein, we focus on the ethanol-heptane mixtures, simulated in the following range of ethanol mole fractions: 0.02, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. All of the mixtures contain N=16000 particles, in order to properly describe extented MH. The program package Gromacs was used to perform molecular dynamics simulations [44, 45]. The simulation protocol has been the same for all mole fractions of ethanol. The initial random configurations of 16000 molecules were created with the program Packmol [46]. These initial configurations were first energy minimized and then equilibrated in the NpT ensemble 5 ns. The length of the production runs varied for each mole fraction of ethanol and can be found in Table 1. The shortest production runs of 10 ns sampled on average 1000 configurations, whereas the longest runs of 35 ns sampled more than 3500 configurations.

The TraPPE forcefield for heptane [38] and ethanol [39] was used throughout the range of ethanol mole fractions. The mixtures were simulated at T = 300 K and p = 1 bar. Temperature was maintained constant using mostly the v-rescale [47] or Nose–Hoover [48, 49] thermostat, while pressure was maintained with the

Mole fractions x_{ETH}	Production run times	Mole fractions x_{ETH}	Production run times
0.02	10 ns	0.5	25 ns
0.05	35 ns	0.6	20 ns
0.1	35 ns	0.7	15 ns
0.2	30 ns	0.8	10 ns
0.3	35 ns	0.9	10 ns
0.4	20 ns		

Table 1: Production run times for the simulated ethanol-heptane mixtures.

Parrinello–Rahman barostat [50, 51]. The temperature algorithms had a time constant of 0.2 ps, while the pressure algorithm was set at 2 ps. The integration algorithm leap-frog [52] was used at every time-step of 2 fs. The short-range interactions were calculated within the 1.5 nm cut-off radius. The electrostatics were handled with the PME method [53], and the constraints with the LINCS algorithm [54].

The snapshots were made with VMD [55].

3 Results

3.1 Experiments

The experimental KBI data for the ethanol-heptane mixtures scanned from Zielkiewicz paper are reported as lines in the main panel of Fig.2.

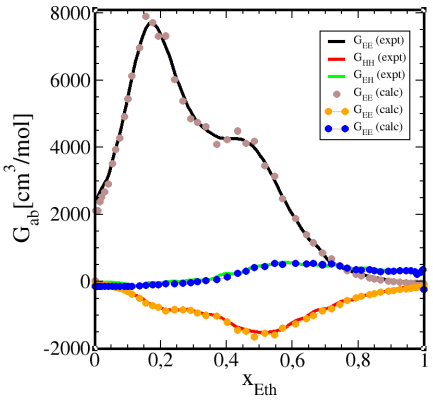

Figure 2: Experimental "camel back shaped" KBI from Ref.[24] for the ethanol-heptane mixtures, reproduced as lines, with symbols for recalculated values (see text) for consistency check.

In order to make sure that these results are fully consistent with one another, we have inverted Eqs(2,3) in order to extract D, using the following assump-

tions. The partial molar volumes have been replaced by the molar volume of each species, thus making these quantities independent of the concentrations. The isothermal compressibility term has been neglected (assuming the incompressibility of the liquid mixtures). Finally, the excess volume of the mixture has been set to be a linear function of the pure liquid volumes, hence neglecting the excess volumes. These are usually found to be less that 0.3cm^3/mol. All 3 assumptions can be justified only by the resulting KBI (as was proven by us for aqueous alcohol mixtures in Ref. [12, 22]). From Eqs(2,3) we have extracted the following equivalent expressions for D

$$D_{ab} = -\frac{V_a B_b}{V G_{ab}} \tag{6}$$

$$D_{aa} = \frac{x_b V_b - x_a V_a}{f_1 - f_2} \tag{7}$$

with $f_1 = x_a x_b \left(G_{aa} - G_{bb}\right)$ and $f_2 = V(x_a - x_b)$. The resulting values are reported in Fig.3.

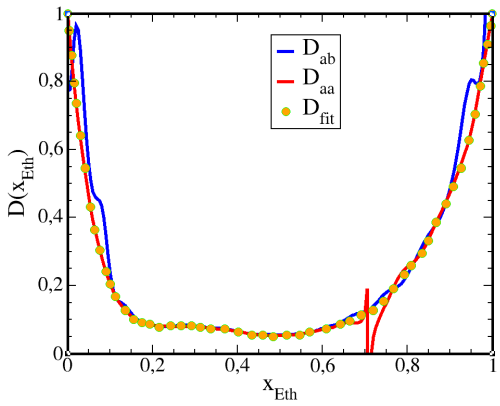

Figure 3: Dual minima D for the ethanol-heptane mixtures, as calculated from the experimental results of Ref.[24] (see Eqs.(6,7) in the text).

It is seen that they are quite similar, with the denominator in Eq.(7) producing a singularity at $x_{\rm Eth} \approx 0.7$. The dots represent the values which have been selected to best represent a good compromise between the two sets. The most prominent feature is the very apparent existence of the 2 extrema, one at $x_{\rm Eth} \approx 0.2$ and the other at $x_{\rm Eth} \approx 0.5$. Interestingly, both evaluations of D coincide almost perfectly in the entire range of the minimum part of D. Using these values of the extracted D function, we return to evaluate back the 3 KBI integrals, by still preserving the 3 approximations mentioned above. The resulting values are reported in Fig.2 as dots, and are seen to nicely superpose to the original data. This simple exercise proves that a single form of D, which must be quite close to the one reported in the inset, must have served in Ref.[24]

to calculate the KBI, and moreover, that the use of true partial molar volumes, compressibilities and excess volume did not affect much the final shapes. We also prove that the 2 extrema of the KBI originate from the 2 extrema observed in D.

3.2 Computer simulations

3.2.1 Snapshots

When mixtures present strong local heterogeneity, it is often instructive to look at snapshots in order to visually appreciate the nature and the extent of the concentration dependence of the spatial segregation. Fig.4 shows snapshots of 4 typical concentrations of ethanol, with different styles highlighting the morphological changes in the ethanol clustering in heptane solvent.

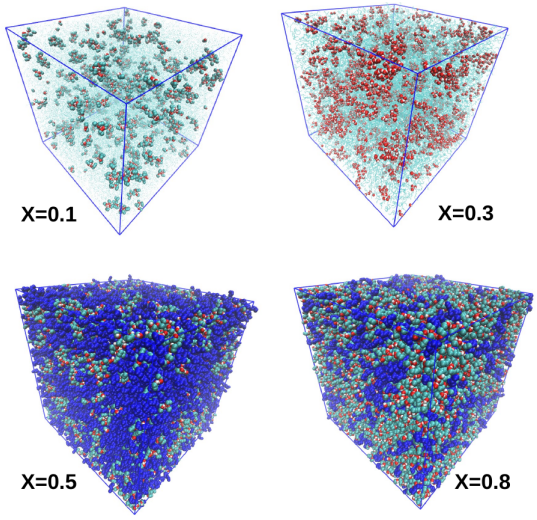

Figure 4: Snapshots of the ethanol-heptane mixtures for 4 typical ethanol concentrations. Ethanol oxygen, hydrogen and carbon groups, in red, white and cyan, respectively, and heptane carbon groups in blue. Heptane is shown as ghost pale blue for $x = 0.1$ and 0.2. Ethanol carbon groups are also shown as ghost pale blue for $x = 0.2$. (see text for details).

For ethanol concentration $x = 0.1$, full ethanol molecules are shown (oxygen atom in red, hydrogen in white and carbon groups in cyan) while heptane molecules are shown in transparent mode. It can be seen that small droplets of ethanol float in the midst of heptane solvent. For $x = 0.3$, only the hydroxyl groups are fully shown, with all carbon groups of both species shown in transparent mode. This way, hydroxyl clusters and their distribution are highlighted. For $x = 0.5$ and $x = 0.8$, ethanol molecules are fully shown like for $x = 0.1$, and

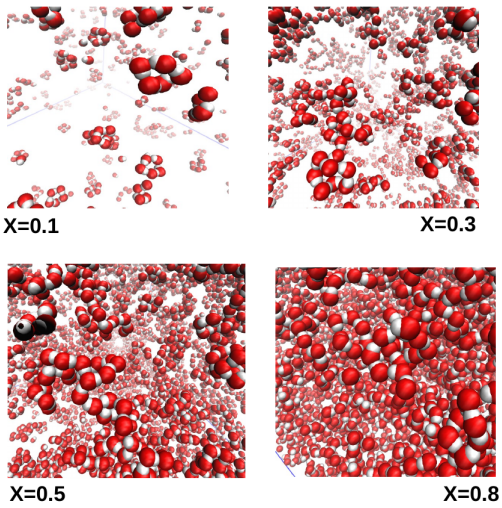

Figure 5: Zoom over the ethanol hydroxyl group clusters for the cases shown in Fig.4. (Box edges can be seen as blue lines).

heptane molecules are shown in blue. These last 2 snapshots highlight the clear micro-segregation of both species.

Fig.5 shows a zoom on the hydroxyl group clustering for the same 4 ethanol concentrations highligted in Fig.4. For $x = 0.1$ one sees that the small ethanol droplets of Fig.4 consist in fact in pentameric rings of hydroxyl groups, forming ethanol pentameric "micelles". This structure vanishes for the other higher ethanol concentrations shown in Fig.5, replaced by more complex cluster structures, such as globules and various type of chains conformations..

Detailed studies of snapshots for the entire concentration range shows that ethanol molecules start to spontaneously form these pentameric micelles at the smallest ethanol concentrations studied herein, which is $x = 0.02$. These pentameric micelle structures persist until $x = 0.3$, after which larger aggregated structures take over. The visual inspection of formed structure tends to confirm that there are 2 regime of clustering, pentameric ethanol micelles for $x < 0.3$ and larger clusters for higher concentrations. We now confirm this through cluster and pair correlation function studies.

3.2.2 Cluster structure

The cluster size probability distributions were calculated as:

$$P_n = \frac{\sum_k s\left(n,k\right)}{\sum_k \sum_n s\left(n,k\right)} \tag{8}$$

where P_n is the probability for the cluster formed of n sites, $s\left(n,k\right)$ represents the number of clusters of the size n in the configuration k. P_n is obtained

by averaging the number $s\,(n,k)$ of clusters of size n over several such configurations. The cut-off distance (r_c) for a calculation depends on the first minimum in the pair correlation function of a particular site. For the oxygens in ethanol, the $r_c = 3.7$ Å.

Fig.6 shows the size s dependence of the cluster distributions $P(s)$ in 2 different ways. Left panel (a) shows the distribution as function of the cluster size, and for different ethanol concentrations. What is apparent is the pentamer peak, which exists for all concentrations. In addition, we observe the appearance of a shoulder-like feature for large clusters, showing that the probability of larger clusters is not negligible for ethanol concentrations above $x > 0.3$. Moreover, while the probability of monomers is higher than that of pentamers for $x < 0.3$, and becomes comparable or lower for $x > 0.3$.

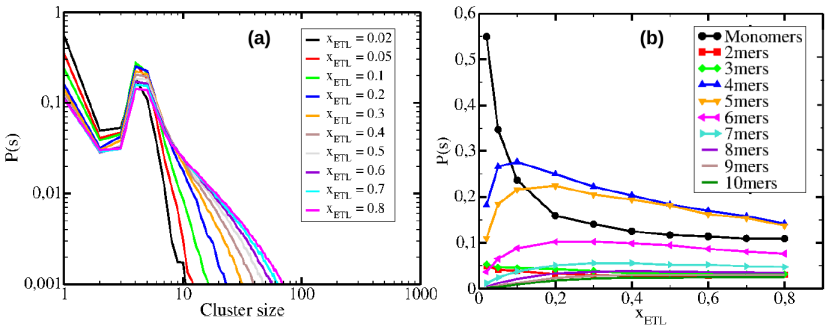

Figure 6: Cluster distribution probability function $P(s)$ as function of **(a)** cluster size s and **(b)** ethanol concentration. (See text for details).

The right panel (b) shows $P(s)$ as function of ethanol concentration, and for different types of n-mers. What is strikingly apparent here is that, for the entire concentration range, there is a sharp fall of P(s) from monomer to dimers and 3-mers, followed by a second dramatic increase for 4-mers and 5-mers, followed by 6-mers, while higher n-mers are the same level as monomers and dimers. An important feature is that the maximum of the curves seems to saturate around $x = 0.3$, indicating a turnover of cluster structures above this concentration.

The cluster study complements the visual inspection of Fig.5 by showing more details, such as for example the existence of 4-mers at small concentrations, and not only pentamers, as the visual inspection may suggest.

3.2.3 Atom-atom pair correlation functions

In order to further confirm the sharp separation of cluster structure highlighted in the previous sections, we show in Fig.7 the evolution of 3 typical pair correlation functions as a function of ethanol concentration. The right panel shows the ethanol oxygen-oxygen correlation function $g_{OO}(r)$ for all the ethanol con-

centrations calculated. Focusing on the second peak, which represent second neighbour correlations, hence the influence of clustering beyond the first neighbour peak, we clearly a difference between two successive concentrations: the gap is wider for $x < 0.3$. In order to facilitate this observation, the curve for x=0.3 is marked in thicker orange color.

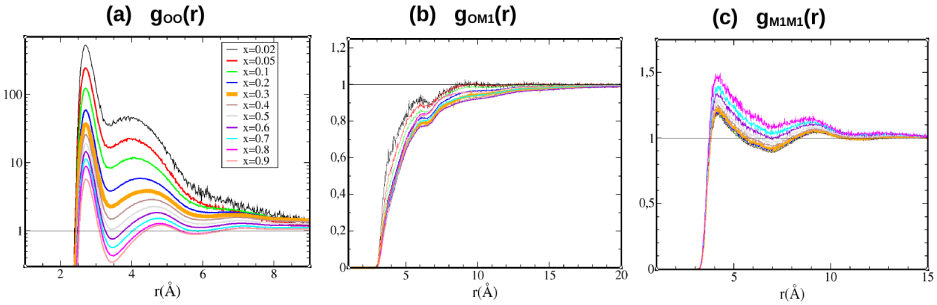

Figure 7: Ethanol concentration dependance of selected atom-atom pair correlation functions. **(a)** Ethanol oxygen-oxygen pair correlation functions $g_{OO}(r)$; **(b)** ethanol oxygen heptane 1st carbon group cross correlations $g_{OM_1}(r)$: **(c)** heptane 1st carbon group correlations $g_{M_1M_1}(r)$. The ethanol concentration $x = 0.3$ is marked at thicker orange line.

The narrower gap for $x > 0.3$ can be interpreted as lesser difference in the second neighbour correlations as a mark of insensitivity for the concentration dependence. This is consistent with hydroxyl clusters being of more varied shapes. In contrast, the almost similar gap for x<0.3 shows a linear dependence in concentration, which would be expected if clustering was the same and would only depend on the concentration of the pentamers.

In order to further confirm this trend, we examine in the middle panel of Fig.7 the cross species correlations $g_{OM_1}(r)$ between the ethanol oxygen and the heptane first(last) carbon group termed M1. These correlations are seen to be less concentration dependent than the previously examined $g_{OO}(r)$ correlations. This is expected, since heptane site are not charged, hence only Lennard-Jones like correlations exist, which are less prominent then for those between charged groups [43]. Nevertheless, we observe that for $x < 0.3$ all correlation functions are nearly superposed (below the thick orange curve), and start to show appreciable differences only when $x > 0.3$. Again, this is fully consistent with the existence of same type of clusters for x<0.3. Indeed, is clustered objects are the same, the correlations between the carbon sites of heptane and the meta objects would be nearly similar, and very weakly dependent on x. In contrast, if the clustered objects are very different in shape, and if this depends strongly on ethanol concentration, we would indeed expect a larger concentration dependence of $g_{OM_1}(r)$.

The last right panel Fig.7 shows a weak concentration correlations between

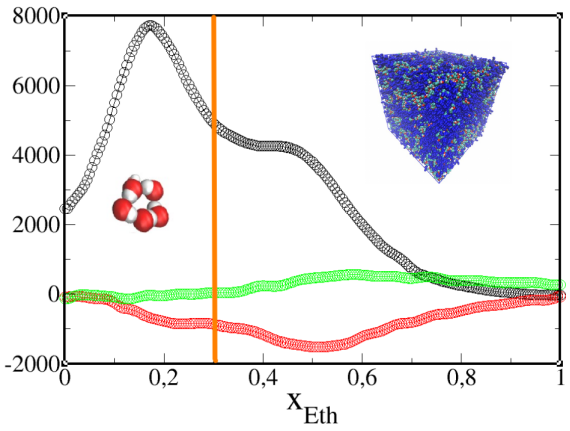

Figure 8: Illustration of the two ethanol concentration regions for each extremum of the KBI, separated by $x = 0.3$ (vertical orange line). Left part for $x < 0.3$ concerns ethanol micellar entities, while right part $x > 0.3$ concerns standard micro-segregated ethanol-heptane domains.

the heptane last/first carbon site $g_{M_1M_1}(r)$, very similar to that observed for $g_{OM_1}(r)$. First of al these are Lennard-Jones like correlations, and correlations below x<0.3 are also nearly superimposed, witnessing the same insensitivity to concentration in presence of same type of aggregates of the ethanol.

If we reconsider now the experimental KBI in the light of this sharp separation of the behaviour of the correlations for $x < 0.3$ and for $x > 0.3$, we obtain the result shown in Fig.8, where the thick orange vertical line at x=0.3 perfectly separates the

concentrations under the first KBI extremum for x< 0.3, where mostly pentameric ethanol micelles are observed, from the concentrations under the second weaker peak at right for $x < 0.3$, for which usual micro-segregation of the 2 species is observed. In order to fully confirm this picture of sharp separation between a meta-object mixture and a usual mixture, we need to calculate the KBI from the simulations and reproduce the same KBI shapes as observed in experiments.

3.3 Kirkwood-Buff integrals from simulations

The evaluation of the KBI through computer simulation requires that the asymptotes of the various site-site correlations are well defined and converging to 1 as expected. Recent investigations have shown that there are 2 major obstacles. The first obstacle is that computer simulations conducted in the Canonical or Isobaric ensemble cannot lead to the proper asymptote 1, this value being reached only in the Grand Canonical ensemble [56, 57, 58]. The second obstacle concerns the existence of micro-segregation, whose spatial extent and slow kinet-

ics alters the statistics of the correlations at large atom-atom separations. Both cases are illustrated below, where we examine both the tail of the correlation functions $g_{ab}(r)$,but also the so-called running KBI defined as [1]:

$$G_{ab}(r) = 4\pi \int_0^r ds\, s^2 \left[g_{ab}(s) - 1\right] \tag{9}$$

which, when r is large enough, is expected to converge to the KBI defined in Eq.(1).

In a first example, we examine the case of ethanol concentration $x = 0.7$, for which we have shown above that micro-segregation is dominant. The typical order parameter for micro-segregation is $g_{OM_1}(r)$, the cross species pair correlation between the ethanol oxygen atom and the first/last carbon group atom of heptanol. Fig.9 shown this function in the main panel for 3 different runs of 5ns each. The strong depletion between adverse ethanol and heptane nano-domains is clearly visible, as all 3 curves stay below 1 until $20\mathring{A}$ or so.

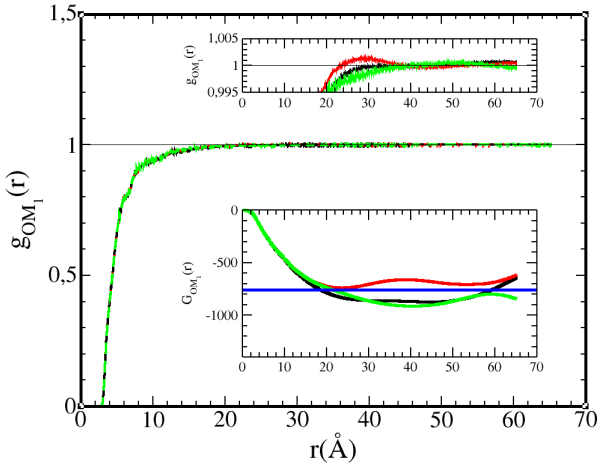

Figure 9: Long range tail contribution for ethanol-heptane cross correlation function $g_{OM_1}(r)$ for ethanol concentration $x = 0.7$ shown in the main panel for 3 different runs of 5ns each. The upper inset is a close-up of the tail part. The lower inset shows the respective RKBI $G_{OM_1}(r)$ functions, along with the experimental KBI (in blue).

On this scale, all 3 curves seem nearly similar, which is what is expected from statistics. However, the zoom on tail region in the upper inset shows that there are visible differences between the 3 runs, as the segregation of domain is not the same in each of the 3 runs. What is observed is the typical domains oscillatory correlations, with half-period about $20\mathring{A}$, which is the depletion extent. All 3 curves oscillated about the asymptote 1, hence the first obstacle mentioned is not really apparent. The lower inset shows the corresponding running KBI $G_{OM_1}(r)$,

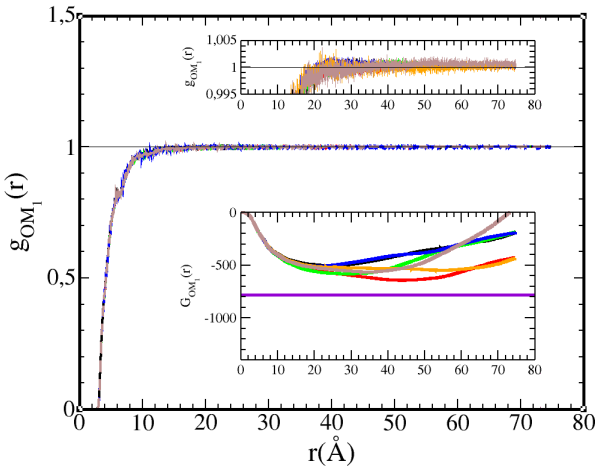

Figure 10: Long range tail contribution for ethanol-heptane cross correlation function $g_{OM_1}(r)$ for ethanol concentration $x = 0.2$ shown in the main panel for 5 different runs of 5ns each. The upper inset is a close-up of the tail part. The lower inset shows the respective RKBI $G_{OM_1}(r)$ functions, along with the experimental KBI (in purple).

as well as the expected experimental value. It is seen that the various runs have an oscillatory asymptotic feature, witnessing the domains alternation, but oscillate around the experimental value. This result shows the rather excellent agreement between the experimental KBI and the calculated one.

In the second example, we examine the case of ethanol concentration $x = 0.2$, for which we expect an homogeneous distribution of ethanol micelles in the midst of heptane solvent. In this case, micro-segregation is very different in that it generates these micelles. Fig.10 shows the same order parameter $g_{OM_1}(r)$ in the main panel, and for 5 different runs of 5ns each.

This function again shows appreciable depletion, but of the size of the ethanol micelles, which is more about $12\mathring{A}$. But now, the zoom of the tail in the upper inset shows that there are no domain oscillations, which is indeed expected if the ethanol droplets are homogeneous distributed. However, we observe that all the asymptote are clearly shifted upwards from 1. This is a direct manifestation of the first obstacle mentioned above, which occurs here since our calculations are in the isobaric ensemble instead of the Grand canonical ensemble. It can be shown that the asymptote of the correlation functions has the following form

$$\lim_{r\to\infty} g_{ab}(r) = 1 - \frac{\epsilon_{ab}}{N} \tag{10}$$

where N is the number of particles in the simulation box, and the value of ϵ_{ab} depend both on species a,b and the statistical ensemble [56, 57, 58]. For like correlations the shift is downwards from 1, and for unlike it is upwards [43]. The RKBI are shown in the lower inset, and are seen to lie above the experimental

value, and not to have the expected flat asymptote for a proper defintition of the KBI value for G_{OM_1}. In order to correct for this "spurious" behaviour, we simply multiply the pair correlation function with the appropriate coefficient $\gamma_{ab} = 1/\left(1 - \frac{\epsilon_{ab}}{N}\right)$ before applying Eq.(10) to $\bar{g}_{OM_1}(r) = \gamma_{OM_1} g_{OM_1}(r)$, which now has the correct asymptote 1. The asymptote-corrected RKBI are computed using the expression

$$\bar{G}_{ab}(r) = 4\pi \int_0^r ds\, s^2 \left[\gamma_{ab} g_{ab}(s) - 1\right] \tag{11}$$

and the corresponding KBI are given by

$$\bar{G}_{ab} = \lim_{r\to\infty} \bar{G}_{ab}(r) \approx \bar{G}_{ab}(L/2) \tag{12}$$

where L is the box size. The above approximation is appropriate if a flat asymptote is reached for $\bar{G}_{ab}(r)$ before $r < L/2$. For each correlation function an initial value of ϵ_{ab}/N is guessed, Eq.(11) applied and we visually test if the $\bar{G}_{ab}(r)$ function is becoming more horizontal. If not, a new guess of ϵ_{ab}/N is used, and the procedure repeated. At some point, the change of horizontality shifts in the opposite direction, and this determines the optimum ϵ_{ab}/N value. Clearly, this empirical procedure is not obvious when the functions $\bar{G}_{ab}(r)$ have large domain oscillations, such as in Fig.11. We have successfully used this empirical shift methodology in several of previous our works in obtaining very good agreement with the experimental data [4, 17, 22, 43] .

The corresponding results are shown in Fig.11.

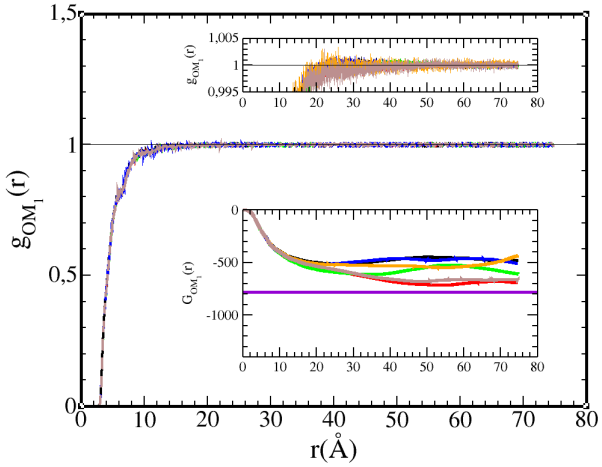

Figure 11: Tail corrected version of pair correlation functions shown Fig.10 (see text)

Due to the extreme smallness of $\frac{\epsilon_{ab}}{N}$, γ_{ab} is nearly 1 and the main panel shows no difference with respect for Fig.10. However, while the upper inset

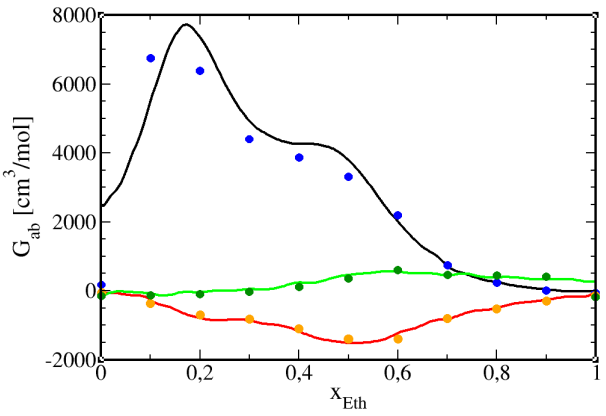

Figure 12: Comparison between the calculated KBI from simulations versus the experimental ones [24].

confirms that the correct asymptote 1 has been set by the operation above, the lower inset shows that all the RKBI have now flat asymptotes, even though it is not quite the value set by the experiments. This discrepancy is expected, since the ethanol aggregation in the real mixtures may not be well represented by the model simulations.

For the calculations of the KBI, from the cross KBI G_{OM_1} we extract D from Eq.(6) and calculate the like KBI G_{OO} and $G_{M_1M_1}$ using Eq.(7). Indeed, the simulations results for these KBI do not often converge to the consistency for D in Eqs.(7,6)), and the above trick is one way to get around this problem. The origin of the inconsistency is due to system size problems, since the like domain statistics require larger simulations, with N=32000 particles or more [33, 34, 35]. When these operations are conducted for all ethanol concentrations, we obtain the result shown in Fig.12, which shows a remarkable agreement with the experimental results.

The very low ethanol concentrations for $x = 0.02$ and $x = 0.05$ pose particular problems for the proper evaluation of the asymptotes. Indeed, the corresponding RKBI have very distorted shapes, which cannot be arranged to look flat by simple shifting of the asymptotes of the pair correlation functions. These distortions, which vary considerably from one run the other, witness the strong kinetics of the ethanol micelle formation. Since these micelles are about $15\mathring{A}$ in diameter, while the half box length extends to $70\mathring{A}$, if the micelles were replaced by soft spheres of same diameter, the statistics on the tail of the soft sphere correlations would be much better. Therefore, it is really the micelle-monomer exchange which affect those statistics at very small ethanol concentrations.

Even though the calculated KBI shown in Fig.12 are in good agreement with the experimental ones, and more importantly, allowing to explain the experimental evidence of the 2 extrema, one could ask if the techniques used in Eqs.(11,12)

are the appropriate ones. Conversely, one may question if other methodologies developped for the calculations of the KBI are suitable for the demanding cases of very large heterogeneity, such as that studied in the present work. We examine this issue in the Supplementary Information (SI). The fact that other methodologies often give results similar to that proposed here ,demonstrates that the current method is equally appropriate, with the added advantage of being intuitive and very simple to implement.

4 Discussion and Conclusion

It is not a priori obvious that the segregation of ethanol in small concentration in oil-like solvent should evolve in a discontinuous manner into the micro-segregation as this concentration is increased, leading to the separation illustrated in Fig.8. For instance, in a previous study of ethanol-benzene mixtures [43], we have found a single extremum shaped KBIs. Yet, this study has shown that small micelle-like aggregates at low ethanol mole fraction mixtures. This difference suggests that closed carbon group molecules, such as benzene, affect differently the cluster structural changes from micelle to domain, than chain-like shaped alkanes. Indeed, these latter molecules can more easily merge with the micellar alkane corona than the disc-shaped benzene molecules. In order to explain this apparent discrepancy, we hypothetize here that it is solvent shape induced depletion entropic effects [59, 60, 61] which affect the structural transition between the ethanol micelles regime and the micro-segregated domain regime. Following this hypothesis, chain-like solvent would allow a smoother transition between micelles and domains, hence clearly separating the two manifestations in terms of fluctuations, thus leading to separate fluctuation regimes.

In the abstract, we mention that water tert-butanol mixtures equally show two extrema in KBI. This was indeed reported our previous work[12] , where the KBI calculated by us (see Fig.6b in Ref.[12]) were obtained by Y. Koga from the vapour pressure measurements of the chemical potentials (see the plot of D in Fig.6a of Ref.[12]) . The double extrema is absent from the KBI data from other authors (see Fig.6b in Ref.[12]), but present in the SANS data obtained in the same work. Interestingly, in boths cases, the double extrema occurs clearly only for the water-water KBI. In addition, the equivalent of the micelle extrema - in the present case it would be tbutanol micellar aggregates, occurs for very small tbutanol concentrations in the interval $0.1 < x < 0.2$, which is similar to that observed in the present work for an entirely different system. Interestingly, the aggregation in aqueous tbutanol is not clearly understood to date, despite several experimental [62, 63] and simulation [32, 33, 35, 64] investigations.

From these two points, it appears that the topic of the existence of camel back shaped KBI deserves further investigations as to the conditions where it might occur.

To conclude, the present study illustrates the difference between concentration fluctuations and micro-segregation both in the microscopic and macroscopic thermodynamic level. It is generally believed that micro-segregation would cor-

respond to a more microscopic $k \neq 0$ part of the concentration fluctuations, related to the finite extent of segregated domains, while concentration fluctuations would be their thermodynamic limit, which correspond to $k = 0$. The present study shown that the KBI, which are related to the macroscopic $k = 0$ limit of the structure factors, themselves contain the difference in both manifestations. This was illustrated in Fig.8, through the sharp separation between concentration fluctuations of meta-objects (the ethanol micelles) for $x < 0.3$ and the micro-segregation regime for $x > 0.3$. To be more specific, while micro-segregation concerns the initial microscopic objects, namely ethanol and heptane molecules, the concentration fluctuations described here concern the meta-objects formed by the ethanol micelles, themselves floating in molecular heptane solvent. This study has been possible, precisely because heptane is an inert solvent, due to its uncharged carbon group sites, which cannot form associated entities. In this context, heptane concentration fluctuations are very small, and the neutrality of this oil-solvent enhances the charge association of ethanol molecules. This is not possible to observe with water-solvent and small alcohol molecules, because both species tend to associate. In order to observe a similar phenomena in water, one requires much larger solutes molecules, such as surfactants, which can form micelles and other self-assembled objects.

The present study also illustrates the importance of the pair correlation functions and their asymptotes, and through these quantities, the issues related to domain segregation and concentration fluctuations in finite size simulations, in particular through the Lebowitz-Percus shift of the asymptotes, illustrate here for the fluctuations between meta-objects, as opposed to the usual illustration for molecular objects.

Appendix

Eq.(2) in the Kirkwood-Buff paper [1] expresses the derivative of the chemical potential $\mu_i$of species i with respect to its mole fraction x_i in terms of the KBI, which we rewrite below using a trivial rearrangement of the original equation as:

$$\left(\frac{\partial \beta \mu_i}{\partial x_i}\right)_{T,P} = \frac{1}{x_i \left[1 + \rho_j x_i \left(G_{11} + G_{22} - 2G_{12}\right)\right]}$$

The Ornstein-Zernike equation for a binary mixture can be written in a matrix equation as

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix} = I$$

where I is the identity matrix, and where the structure matrix elements are $S_{ij}(k) = \delta_{ij} + \sqrt{\rho_i \rho_j}\tilde{h}_{ij}(k)$ and the C-matrix elements are defined in terms of the direct correlation functions as $C_{ij}(k) = \delta_{ij} - \sqrt{\rho_i \rho_j}\tilde{c}_{ij}(k)$, with the Fourier transform notation and definition $\tilde{f}(k) = \int d\vec{r} f(r) \exp(i\vec{r}.\vec{k})$. Since the KBI are defined as in Eq.(1), one has $G_{ij} = \tilde{h}_{ij}(k = 0)$, and $S_{ij} = \delta_{ij} + \sqrt{\rho_i \rho_j} G_{ij}$.

Therefore, one can rewrite the equation above as (using $\rho_i = x_i\rho$)

$$\left(\frac{\partial\beta\mu_i}{\partial x_i}\right)_{T,P} = \frac{1}{x_i\sqrt{x_1x_2}\left[S_{11}+S_{22}-2S_{12}\right]}$$

which can be rewritten by using the OZ equation as

$$\left(\frac{\partial\beta\mu_i}{\partial x_i}\right)_{T,P} = \frac{\det C}{x_i\sqrt{x_1x_2}\left[C_{11}+C_{22}+2C_{12}\right]}$$

Using Eq.(4) which defines the KBI coefficient D, one has the relation relating D to the determinant of C, which controls the stability limit of a mixture [3], as written in Eq.(5)

$$D = \alpha\left[C_{11}C_{22}-C_{12}^2\right]$$

with $\alpha = 1/(\sqrt{x_1x_2}\left[C_{11}+C_{22}+2C_{12}\right])$.

References

[1] J.G. Kirkwood and F.P. Buff. The statistical mechanical theory of solutions. i. *The Journal of Chemical Physics*, 19(6):774, 1951.

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

[3] J.-P. Hansen and I.R. McDonald. *Theory of Simple Liquids*. Academic Press, Elsevier, Amsterdam, 3rd edition, 2006.

[4] B. Lovrinčević, A. Bella, I. Le Tenoux-Rachidi, M. Požar, F. Sokolić, and A. Perera. Methanol-ethanol "ideal" mixtures as a test ground for the computation of kirkwood-buff integrals. *Journal of Molecular Liquids*, 293:111447, 2019.

[5] A. Ben-Naim. Inversion of the kirkwood-buff theory of solutions: Application to the water-ethanol system. *The Journal of Chemical Physics*, 67(11):4884, 1977.

[6] E. Matteoli and L. Lepori. Solute-solute interactions in water. ii. an analysis through the kirkwood-buff integrals for 14 organic solutes. *The Journal of Chemical Physics*, 80(6):2856, 1984.

[7] Y. Marcus. Preferential solvation in mixed solvents. part 5.-binary mixtures of water and organic solvents. *J. Chem. Soc., Faraday Trans.*, 86:2215–2224, 1990.

[8] Y. Marcus. Preferential solvation in mixed solvents. part 6.-binary mixtures containing methanol, ethanol, acetone or triethylamine and another organic solvent. *J. Chem. Soc., Faraday Trans.*, 87:1843–1849, 1991.

[9] J.D. Pandey and Richa Verma. Inversion of the kirkwood-buff theory of solutions: application to binary systems. *Chemical Physics*, 270(3):429–438, 2001.

[10] P. E. Smith. On the kirkwood-buff inversion procedure. *The Journal of Chemical Physics*, 129(12):124509, 2008.

[11] I. Shulgin and E. Ruckenstein. Kirkwood-buff integrals in aqueous alcohol systems: comparison between thermodynamic calculations and x-ray scattering experiments. *The Journal of Physical Chemistry B*, 103:2496, 1999.

[12] A. Perera, F. Sokolić, L. Almásy, and Y. Koga. Kirkwood-buff integrals of aqueous alcohol binary mixtures. *The Journal of Chemical Physics*, 124(12):124515, 2006.

[13] M. C. A. Donkersloot. The structure of binary liquids. the kirkwood-buff theory of liquid mixtures, illustrated on the basis of the systems water/methanol, water/ethanol, and cyclohexane/2,3-dimethylbutane, as a link between thermodynamic data and x-ray and neutron scattering results. *Journal of Solution Chemistry*, 8:293, 1979.

[14] K. Nishikawa and T. Iijima. Small-angle x-ray scattering study of fluctuations in ethanol and water mixtures. *The Journal of Physical Chemistry*, 97(41):10824–10828, 1993.

[15] Y. Marcus. Preferential solvation in mixed solvents x. completely miscible aqueous co-solvent binary mixtures at 298.15 k. *Monatshefte fuer Chemie*, 132:1387–1411, 2001.

[16] Y. Marcus. Preferential solvation in mixed solvents. part 11. eight additional completely miscible aqueous co-solvent binary mixtures and the relationship between the volume-corrected preferential solvation parameters and the structures of the co-solvents. *Phys. Chem. Chem. Phys.*, 4:4462–4471, 2002.

[17] A. Perera, F. Sokolić, L. Almásy, P. Westh, and Y. Koga. On the evaluation of the kirkwood-buff integrals of aqueous acetone mixtures. *The Journal of Chemical Physics*, 123(2):024503, 2005.

[18] Keiko Nishikawa. The solution chemistry of mixing states probed via fluctuations: a direct description of inhomogeneity in mixing. *Bulletin of the Chemical Society of Japan*, 94(9):2170–2186, 2021.

[19] S. Dixit, J. Crain, W.C.K. Poon, J.L. Finney, and A.K. Soper. Molecular segregation observed in a concentrated alcohol-water solution. *Nature*, 416:829–832, 2002.

[20] J.-H. Guo, Y. Luo, A. Augustsson, S. Kashtanov, J.-E. Rubensson, D. K. Shuh, H. Ågren, and J. Nordgren. Molecular structure of alcohol-water mixtures. *Physical Review Letters*, 91:157401, 2003.

[21] S.K. Allison, J.P. Fox, R. Hargreaves, and S.P. Bates. Clustering and microimmiscibility in alcohol-water mixtures: Evidence from molecular-dynamics simulations. *Physical Review B*, 71:024201, 2005.

[22] M. Mijaković, B. Kežić, L. Zoranić, F. Sokolić, A. Asenbaum, C. Pruner, Emmerich Wilhelm, and A. Perera. Ethanol-water mixtures: ultrasonics, brillouin scattering and molecular dynamics. *Journal of Molecular Liquids*, 164(1):66–73, 2011. Complex liquids: Modern trends in exploration, understanding and application Selected papers on molecular liquids presented at the EMLG/JMLG annual meeting 5-9 september 2010.

[23] M. Požar, B. Lovrinčević, L. Zoranić, T. Primorac, F. Sokolić, and A. Perera. Micro-heterogeneity versus clustering in binary mixtures of ethanol with water or alkanes. *Physical Chemistry Chemical Physics*, 18:23971–23979, 2016.

[24] J. Zielkiewicz. Kirkwood-buff integrals in the binary and ternary mixtures containing heptane and aliphatic alcohol. *The Journal of Physical Chemistry*, 99(10):3357–3364, 1995.

[25] A.H. Narten and A. Habenschuss. Hydrogen bonding in liquid methanol and ethanol determined by x-ray diffraction. *The Journal of Chemical Physics*, 80:3387–3391, 1984.

[26] K. S. Vahvaselkä, R. Serimaa, and M. Torkkeli. Determination of liquid structures of the primary alcohols methanol, ethanol, 1-propanol, 1-butanol and 1-octanol by x-ray scattering. *Journal of Applied Crystallography*, 28:189–195, 1995.

[27] A.K. Karmakar, P.S.R. Krishna, and R.N. Joarder. On the structure function of liquid alcohols at small wave numbers and signature of hydrogen-bonded clusters in the liquid state. *Physics Letters A*, 253:207–210, 1999.

[28] M. Tomšič, A. Jamnik, G. Fritz-Popovski, O. Glatter, and L. Vlček. Structural properties of pure simple alcohols from ethanol, propanol, butanol, pentanol, to hexanol: Comparing monte carlo simulations with experimental saxs data. *The Journal of Physical Chemistry B*, 111:1738–1751, 2007.

[29] M. Požar, J. Bolle, C. Sternemann, and A. Perera. On the x-ray scattering pre-peak of linear mono-ols and the related microstructure from computer simulations. *The Journal of Physical Chemistry B*, 124(38):8358–8371, 2020.

[30] J. Bolle, S. P. Bierwirth, M. Požar, A. Perera, M. Paulus, P. Münzner, C. Albers, S. Dogan, M. Elbers, R. Sakrowski, G. Surmeier, R. Böhmer,

M. Tolan, and C. Sternemann. Isomeric effects in structure formation and dielectric dynamics of different octanols. *Phys. Chem. Chem. Phys.*, 23:24211–24221, 2021.

[31] M.E. Lee and N.F.A. van der Vegt. A new force field for atomistic simulations of aqueous tertiary butanol solutions. *The Journal of Chemical Physics*, 122(11):114509, 2005.

[32] S. Banerjee, J. Furtado, and B. Bagchi. Fluctuating micro-heterogeneity in water-tert-butyl alcohol mixtures and lambda-type divergence of the mean cluster size with phase transition-like multiple anomalies. *The Journal of Chemical Physics*, 140(19):194502, 2014.

[33] R. Gupta and G. N. Patey. Aggregation in dilute aqueous tert-butyl alcohol solutions: Insights from large-scale simulations. *The Journal of Chemical Physics*, 137(3):034509, 2012.

[34] S. D. Overduin and G. N. Patey. Comparison of simulation and experimental results for a model aqueous tert-butanol solution. *The Journal of Chemical Physics*, 147(2):024503, 2017.

[35] S. D. Overduin, A. Perera, and G. N. Patey. Structural behavior of aqueous t-butanol solutions from large-scale molecular dynamics simulations. *The Journal of Chemical Physics*, 150(18):184504, 2019.

[36] W.L. Jorgensen, J.D. Madura, and C.J. Swenson. Optimized intermolecular potential functions for liquid hydrocarbons. *Journal of the American Chemical Society*, 106(22):6638–6646, 1984.

[37] W.L. Jorgensen. Optimized intermolecular potential functions for liquid alcohols. *The Journal of Physical Chemistry*, 90(7):1276–1284, 1986.

[38] M. G. Martin and J. I. Siepmann. Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes. *The Journal of Physical Chemistry B*, 102(14):2569–2577, 1998.

[39] B. Chen, J.J. Potoff, and J.I. Siepmann. Monte carlo calculations for alcohols and their mixtures with alkanes. transferable potentials for phase equilibria. 5. united-atom description of primary, secondary, and tertiary alcohols. *The Journal of Physical Chemistry B*, 105:3093, 2001.

[40] J. Matouš, J. Hrnčiřík, J. P. Novák, and J. Šobr. Liquid-liquid equilibrium in the system water-tetrahydrofuran. *Collect. Czech. Chem. Commun.*, 35:1904–1905, 1970.

[41] T. Narayanan and A. Kumar. Reentrant phase transitions in multicomponent liquid mixtures. *Physics Reports*, 249(3):135–218, 1994.

[42] I. Brovchenko and B. Guillot. Simulation of the liquid-liquid coexistence of the tetrahydrofuran+water mixture in the gibbs ensemble. *Fluid Phase Equilibria*, 183-184:311–319, 2001. Proceedings of the fourteenth symposium on thermophysical properties.

[43] M. Požar, J.-B. Seguier, J. Guerche, R. Mazighi, L. Zoranić, M. Mijaković, B. Kežić-Lovrinčević, F. Sokolić, and A. Perera. Simple and complex disorder in binary mixtures with benzene as a common solvent. *Physical Chemistry Chemical Physics*, 17:9885–9898, 2015.

[44] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845, 2013.

[45] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19, 2015.

[46] J.M. Martínez and L. Martínez. Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. *Journal of Computational Chemistry*, 24(7):819–825, 2003. ibid L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez. Journal of Computational Chemistry, 30(13):2157-2164, 2009.

[47] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

[48] S. Nose. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255, 1984.

[49] W.G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[50] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Physical Review Letters*, 45(14):1196, 1980.

[51] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182, 1981.

[52] R.W. Hockney. *Methods in computational physics, vol. 9*, volume 9, chapter The potential calculation and some applications, pages 135–221. Orlando Academic Press, 1970.

[53] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n*log(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[54] B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463, 1997.

[55] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33, 1996.

[56] J.L. Lebowitz and J.K. Percus. Long-range correlations in a closed system with applications to nonuniform fluids. *Physical Review*, 122(6):1675, 1961.

[57] J. J. Salacuse, A. R. Denton, and P. A. Egelstaff. Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. i. theoretical method. *Physical Review E*, 53:2382, 1996.

[58] J. J. Salacuse, A. R. Denton, P. A. Egelstaff, M. Tau, and L. Reatto. Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. ii. application to a model krypton fluid. *Physical Review E*, 53:2390, 1996.

[59] S. Asakura and F. Oosawa. On interaction between two bodies immersed in a solution of macromolecules. *The Journal of Chemical Physics*, 22(7):1255–1256, 1954.

[60] L. Sapir and D. Harries. Origin of enthalpic depletion forces. *The Journal of Physical Chemistry Letters*, 5(7):1061–1065, 2014. PMID: 26274449.

[61] L. Sapir and D. Harries. Is the depletion force entropic? molecular crowding beyond steric interactions. *Current Opinion in Colloid & Interface Science*, 20(1):3–10, 2015.

[62] R. Winkler, E. Ré, G. Arrachart, and S. Pellet-Rostaing. Impact of solvent structuring in water/tert-butanol mixtures on the assembly of silica nanoparticles to aerogels. *Langmuir*, 35(24):7905–7915, 2019. PMID: 31088054.

[63] M. Sedlák and D. Rak. On the origin of mesoscale structures in aqueous solutions of tertiary butyl alcohol: The mystery resolved. *The Journal of Physical Chemistry B*, 118(10):2726–2737, 2014. PMID: 24559045.

[64] J. Cerar, A. Jamnik, I. Pethes, L. Temleitner, L. Pusztai, and M. Tomšič. Structural, rheological and dynamic aspects of hydrogen-bonding molecular liquids: Aqueous solutions of hydrotropic tert-butyl alcohol. *Journal of Colloid and Interface Science*, 560:730–742, 2020.

Supplementary Information for : Camel back shaped Kirkwood-Buff integrals

Aurélien Perera[(1)], Martina Požar[(2)] and Bernarda Lovrinčević[(2)]

(1) Laboratoire de Physique Théorique de la Matière Condensée (UMR CNRS 7600), Sorbonne Université, 4 Place Jussieu, F75252, Paris cedex 05, France.
(2) Department of Physics, Faculty of Science, University of Split, Ruđera Boškovića 33, 21000, Split, Croatia.

In this SI document, we consider the size dependence of the KBI, as well two technical issues concerning the evaluation of macroscopic fluctuations within finite size simulation boxes, and which could potentially affect the KBI computed in the work. The analysis shows that these methods are equally affected by the strong local heterogeneity, and do not improve over the asymptote shifting method used herein.

A – System size dependence of the KBI

Although rather large system size was used in the simulations (N=16000 molecules), leading to calculated KBI in good agreement with the experimental data, and allowing to support the idea of duality in aggregation and fluctuations, we have nevertheless considered useful to simulate larger systems of N=32000 molecules, doubling the previous size. We have focused on ethanol concentration $x = 0.2$ which corresponds to the middle of the ethanol "micelle" region, where competition between aggregation and fluctuation may be the most challenging to reproduce. Two independent runs of 5ns each were performed, after the initial 10ns equilibration. The two calculated RKBI $G_{OM_1}(r)$, both for the asymptote shifted $g_{OM_1}(r)$ data (main panel) and non-shifted data (inset) are shown below in Fig.1A. The data from Fig.11 is equally shown as green curves, for comparison purpose.

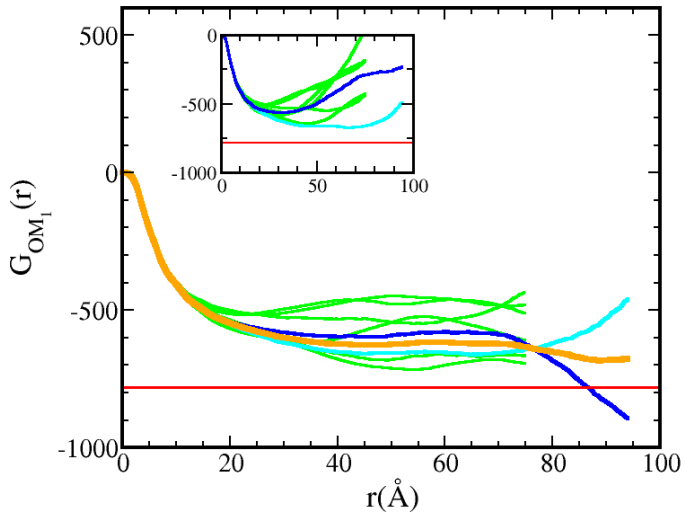

Fig.1A. Comparison of the RKBI for a N=32000 system (blue and cyan curves) with that of the N=16000 systems shown in Fig.11 (reproduced in green here). The inset shows the original asymptote uncorrected curves. The average of the 2 runs for the larger system is shown as orange curve.

It is seen that both new curves has quite an appreciable horizontal middle range, pointing toward the same mean value suggested by the smaller systems. When averaging the 2 curves, the resulting orange curve shows a rather excellent agreement with the mean value obtained by the smaller system calculations. The large deviations at the end of the half-box size are expected for 2 reasons.

First, the statistics on the aggregated entities show appreciable domain oscillations, which are genuine features, and may require excessive long simulation to stabilize. Second, small uncertainties in the large distance behaviour of the $g(r)$ are amplified by the integration procedure. These results show that very large systems are not necessarily a good strategy when dealing with strongly heterogeneous systems with inner large association-dissociation kinetics.

B – Finite size scaling analysis

Finite size analysis has been originally [1] used in the context of second order phase transitions for the following reason. In this type of phase transition, the correlation length ξ diverges, hence may become larger than the simulation box size L . It necessary to find the way ξ scales with L in order to predict the distribution of the order parameter with L . In the present context which is far from any demixing phase transition, the correlation is about few atom size ($\xi \approx 5$ Å), much smaller the box size $L \approx 135$ Å), and such considerations may appear as remote. Yet, the KBI are a measure of the concentration fluctuations, since , alternatively to the definition in Eq.(1), they are also defined as [2]

$$G_{ab} = V \frac{\langle N_a N_b \rangle - \langle N_a \rangle \langle N_b \rangle}{\langle N_a \rangle \langle N_b \rangle} - V \frac{\delta_{ab}}{\langle N_b \rangle} \tag{1B}$$

where N_x is the number of particles of species x . However, in the context of simulations when the total number of particles is fixed, the total fluctuation is zero (-ie- $\langle N_a N_b \rangle - \langle N_a \rangle \langle N_b \rangle = 0$. This is the case in canonical and isobaric ensembles, the latter which we have used in this work. Hence, the formula above can only give the ideal gas result for the KBI. However, one can compute local fluctuations within smaller boxes of size $L_i < L$ embedded within the simulation box, and accumulate statistics while varying the sizes L_i , hence estimate the size dependence of local G_{ab} by using Eq.(1a). While these values will be strongly distorted for $L/2 < L_i < L$, one can expect that there will be a scaling with smaller sizes $L_i < L/2$, which would allow to extract the correct KBI by extrapolating to larger L_i . This is exactly what was found in Ref.[3], with the following scaling relation for intermediate values of the L_i

$$G_{ab}^{(i)} = G_{ab} + \frac{\alpha_{ab}}{L_i} \tag{2B}$$

Below, we have applied this method to the 2 cases studied in detail in Section 3.3, namely for ethanol concentrations $x = 0.2$ and $x = 0.7$. We have followed the procedure described in Ref. [4]. We started with small boxes of size $L_0 = 15$ Å, and computed cross fluctuations between the oxygen sites of ethanol (O site) and first carbon group of heptane (M_1 site), using Eq.(1B). This was done for 100 random choices of the box center within the main box. This operation was repeated for several choices for $L_i = L_0 + \delta L$, with $\delta L = 5$ Å, which represents about 15-17 points before reaching $L/2$ where distortions appear (as one moves towards the zero global fluctuation regimes). The relation (2B) is indeed well verified, as can be seen in Fig-1B and Fig-2B below.
The finite size scaling applied to the case of $x = 0.7$ is shown in Fig.1B. The main panel shows, in addition to the scaling analysis, the RKBI of Fig.9. The inset shows the $1/r$ plot, which allows to verify the extent of the applicability of Eq.(2B). It is seen that the scaling analysis tends to predict KBI in close agreement to that guessed from the KBI. Fig.1B shows a similar analysis for $x = 0.2$, but for the asymptote shifted KBI. The close agreement demonstrates that the asymptote shift method produces correct KBI values with a simpler empirical approach. One of the data for the N=32000 particles is also shown as brown curve and symbols. Again, a rather good agreement with previous smaller size data is observed

The extrapolation $1/L_i \to 0$ shown in the insets predicts values for the KBI which are very similar to those shown in Fig.9 and Fig.11, showing the same dispersion. This is not a surprise, since each calculations have been performed using the same trajectories as for the calculations of the $g_{OM_1}(r)$ reported in section 3.3. It is particularly interesting for the case of $x=0.2$, since the newly extrapolated KBI are consistent with the KBI obtained the corrected asymptotes as in Fig.11, and not those of Fig.10.

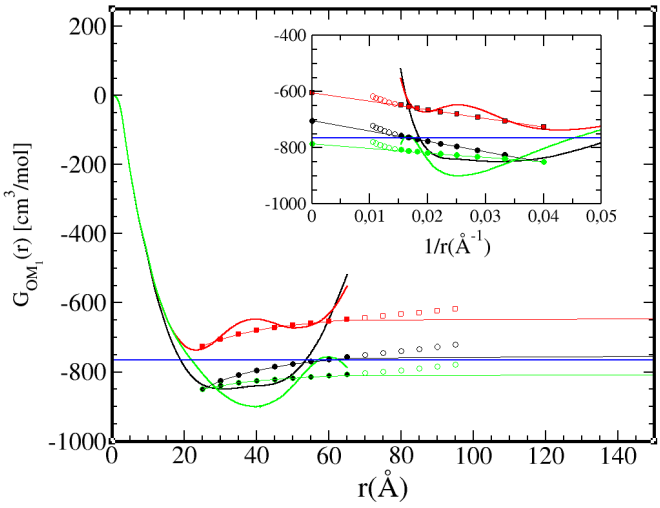

Fig.1B. Finite size scaling analysis of KBI for x=0.7, for the 3 independent runs shown in Fig.9 with the same 3 color conventions. The inset shows the 1/r dependence which obeys Eq.(2B). Corresponding RKBI are shown both in the main panel and inset. Close dots correspond to points r<L/2, and open dots for r>L/2. The horizontal blue line is the experimental value.

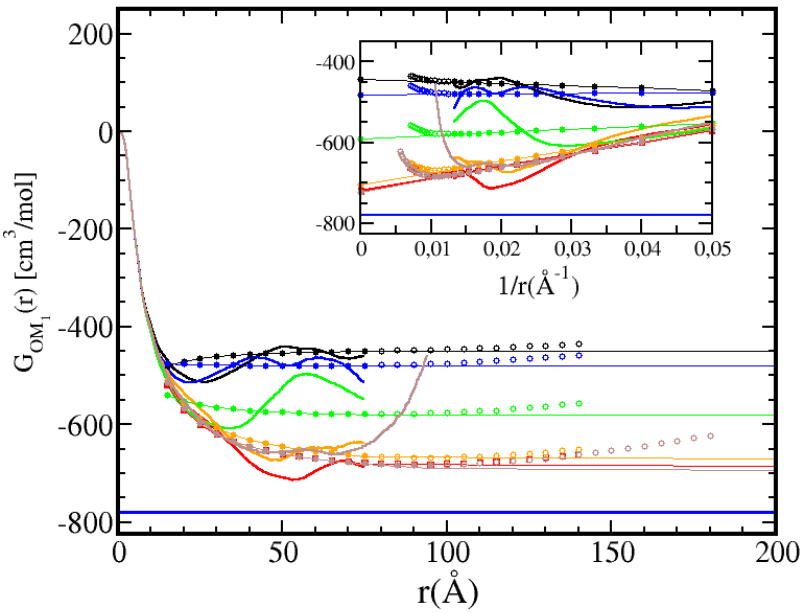

Fig.2B. Finite size scaling analysis of KBI for x=0.2, for the independent runs shown in Fig.11 with the same color conventions. In addition, one result for the N=32000 particle of Fig.1A is equally shown as brown lines and symbols. Line and symbol conventions are as in Fig.1B.

Again, this is not a surprise, since the work of Lebowitz and Percus in the 70's have demonstrated [5] that correlation functions computed in other ensembles than the Grand Canonical ensemble (GCE) (which allows for total particle number fluctuations, due to the coupling to a particles reservoir), should not tend asymptotically to 1, but rather obey Eq.(10). In particular, it should not

obey Eq.(1), and this is exactly what the current calculations have confirmed. However, the results shown in the SI equally demonstrate that calculations based on shifting the asymptote, as in Eqs. (11,12) are a proper way to use Eq.(1B) in the isobaric ensemble, since the predict KBI very similar to those of the finite scale analysis.

C – Alternate computational methodology for evaluating KBI from simulations

The comparison conducted in Section B of this SI document confirms that the incorrect behaviour of the asymptote of the $g_{ab}(r)$, is a fundamental problem for the computation the KBI from simulations conducted in ensembles other than the GCE, since Eq.(1) cannot be used. A similar conclusion was equally reached in Ref.[4] where the authors proposed a formula to compute the asymptote shift, instead of the empirical numerical approach used here.
A different method was proposed in Ref[6] which consists in weighting the integral in the KBI by a function which takes into account cross correlations between the domains surrounding the two particles 1 and 2 for which the separation is $r=|\vec{r}_1-\vec{r}_2|$. The important point we would like demonstrate here, is that this method is not suited to predict the KBI, unless the asymptote is corrected prior to using it. According to this method the KBI are given by the expression

$$G_{ab}=\lim_{R\to\infty} G_{ab}^{(w)}(R)=\lim_{R\to\infty} 4\pi\int_0^R dr\, r^2[g_{ab}(r)-1]w_R(r) \tag{1C}$$

The distance dependent weight $w_R(r)$ accounts for the overlap volume between the 2 spheres centered around particles 1 and 2, which leads to the expression

$$w_R(r)=1-\frac{3}{2}\frac{r}{R}+\frac{1}{2}\left(\frac{r}{R}\right)^3 \tag{2C}$$

An alternative expression, valid for large R , is

$$w_R(r)=1-\left(\frac{r}{R}\right)^3 \tag{3C}$$

A first remark about Eq.(1C) is that it has the same problem as Eq.(1) with respect to the asymptote, since it uses $g_{ab}(r)-1$, which is not appropriate in constant N ensembles. In other words, it does not cure the fundamental problem of the computation of the KBI noted above. A second remark concerns how the weight function $w_R(r)$ affect the integral in Eq.(1C) as compared to that in Eq. (1). In both expressions, $w_R(r)$ varies smoothly from 1 at $r=0$, to 0 for $r=R$. Hence, it will affect the value of the running integral $G_{ab}(r)$ close to $r=R$,to differ from the unweighted one, except when R is large enough such that $g_{ab}(r)-1\approx 0$. This damping effect makes the running integral to converge faster to KBI value, and in particular at intermediate R to differ from those obtained using the unweighted expression. From the expressions above, we see that the damping will be faster for Eq.(3C) than for Eq.(2C), which is indeed verified below.

Before using Eqs.(1C-3C) in the present context, and to illustrate the two remarks above, we have calculated the KBI for a simple Lennard-Jones liquid, for the reduced density $\rho^*=(N/V)\sigma^3=0.8$ and reduced temperature $T^*=T/(\epsilon/k_B)=1.5$, in a simulation with $N=16000$ particles. The RKBI function $G(r)$, as calculated from the 3 methods, namely the asymptote shifting method of the present work and the 2 weighting methods Eq.(2C) and Eq.(3C), are shown in Fig.1C, and for both the cases when the asymptote of $g(r)$ is un-shifted or shifted. This figure is very similar to Fig.(1a) in Ref.[6].
The $G(r)$ for the original data, calculated using Eq.(9) is shown in dashed orange lines and that using Eq.(11) is shown in red lines. The asymptote shift is empirically found to be extremely small (since $\epsilon N=0.0352$). Although the difference between the 2 curves in invisible in the main panel, it is very apparent in the zoom shown in the inset, the orange curve for $G(r)$ of the un-shifted $g(r)$ showing this typical tendency of curving downward.

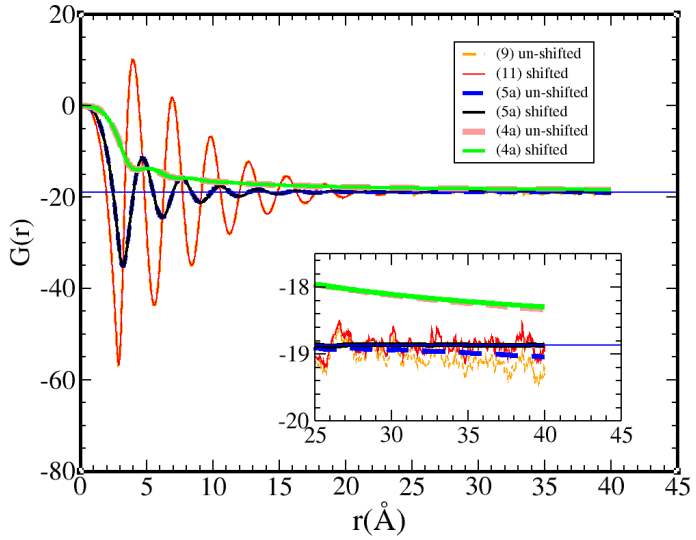

Fig.1C. Comparison of the calculation methods for the RKBI for a Lennard-Jones liquid (see text). The lower inset is a zoom on the tail behaviour of the various G(r) of the main panel. In both plots, the thin horizontal blue lines serves as a guide for the correct KBI value.

When using Eq.(2C), the $G(r)$ for the un-shifted (thick dashed pink curve) and the shifted one (thick green curve) show a very good damping of the oscillations at all distances, but very slow convergence, as illustrated in the inset. By contrast, $G(r)$ obtained by Eq.(3C) for the shifted $g(r)$ (thick black curve) converge faster, and the agreement with the method of Eq.(11) is very good, with the advantage of being less noisy, precisely because of the damping. However, if Eq. (3C) is applied to the un-shifted $g(r)$ (thick dashed blue curve), it curves downward, just as the orange curve. This plot demonstrates clearly that it is the asymptote shift which ensures the obtention of the correct KBI. The weighting methodology could be useful to spot the incorrect asymptote, since it damps the noise in the $G(r)$.

Next, in Fig.2C, we compare the RKBI for ethanol concentration $x=0.7$ in a plot similar to the lower inset of Fig.9, for the 3 different runs depicted with different colors.

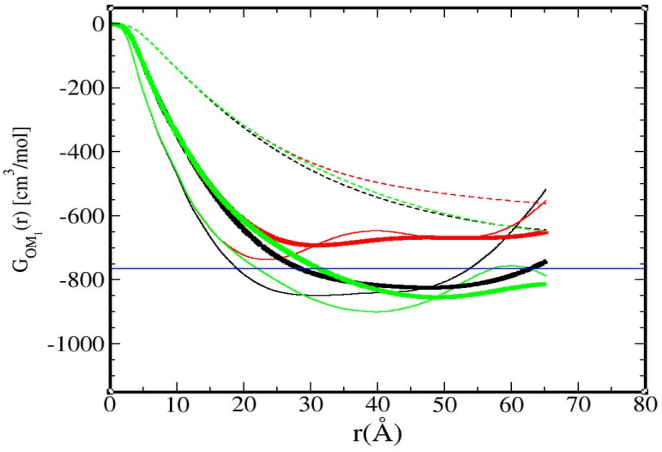

Fig.2C. Comparison of the calculation methods for the RKBI for x=0.7 for 3 different runs depicted with 3 colors (black, red, green). The data from Eq. (2C) are shown as thin dashed lines, that for Eq.(3C) in thick lines, and that from Fig.9 in thin lines. The horizontal blue line is the data the experimental KBI value.

The RKBI obtained with Eq.(2C) (dashed lines) are clearly the worst of the 3, and this is consistent with the results for the LJ fluid shown above. Those computed with Eq.(3C) (thick lines) are better than those reported in Fig.9, in the sense that they flatten out better, thus allowing for a better

estimate. However, these estimates are not so different than those predicted in this work and the total estimate is very close to the previous one, both predictions being indistinguishable on the scale of the final plot Fig.12.

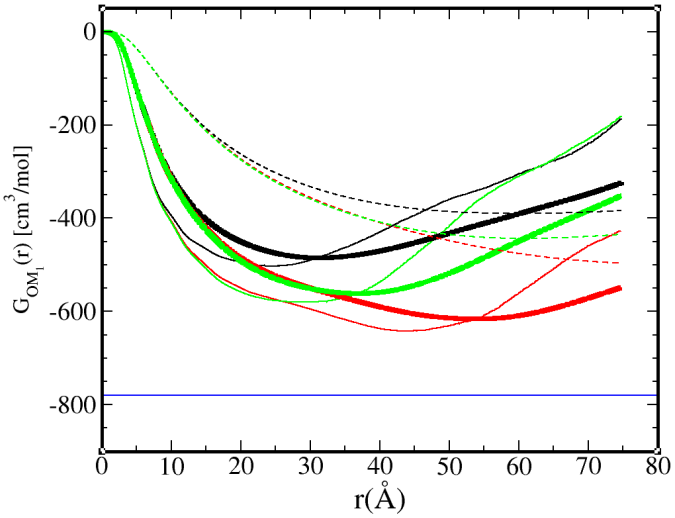

Fig.3C. Comparison of the calculation methods for the RKBI for x=0.2 for 3 different runs, and computed from the un-shifted correlation functions. The horizontal blue line is the data the experimental KBI value. Line conventions as in Fig.3C.

The same analysis for the case of $x=0.2$ is done below in Fig.4C, first for the case of the un-shifted asymptotes (as in Fig.10), and second for the case of the shifted ones (Fig.11). To make the plots less cluttered, only 3 of the 5 runs have been analyzed.

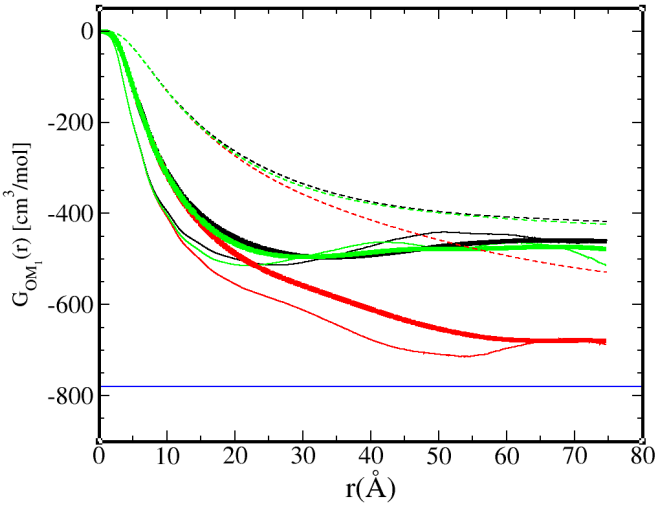

Fig.4C. Comparison of the calculation methods for the RKBI for x=0.2 for 3 different runs, and computed from the shifted correlation functions. The horizontal blue line is the data the experimental KBI value.

The conclusion of these 3 calculations can be expressed as follows.
Eq.(3C) produces better KBI results than the original expression Eq.(2C), the latter which was based on the physical picture of allowing for cross correlations between the 2 particles of interest, itself based on the claim of better mathematical treatment of the initial expression for the KBI integral [6]. In view the results shown here, but also in the original papers [3,6,7], one may raise some doubts on the entire procedure. Indeed, both in Ref.[3] and [7], the validity of Eqs.(1C-3C) have been successfully tested for a model $g(r)$ which has the correct asymptote of 1, valid for

the GCE but for the N-constant ensembles for which the entire procedure is supposed to apply. The dashed blue curve in Fig.1C clearly shows that a g(r) taken from N-constant ensemble will not produce the proper KBI.
When using properly shifted asymptotes, Eq.(3C), which is an approximation of Eq.(2C) valid for large separations, produces better results by damping the oscillating RKBI asymptotes. However, this forced damping may produce artifacts for mixtures where the oscillations are realistic physical features witnessing the alternance of segregated domains, such in the case of aqueous-1propanol mixtures [8], and possibly in more complex soft matter liquids.

These results may not be so surprising, since this methodology was generally tested in less demanding cases, where the asymptote of the correlations are well defined, because the corresponding liquids are very homogeneous at microscopic scales. The cases discussed here present heavy domain segregation, hence asymptote distortions are important physical manifestations which cannot be dismissed [9]. The method proposed in the present work (and also in our previous ones), although empirical, has the advantage of not altering the influence of domain formation statistics on the pair correlation functions.

[1] Binder, K. Monte Carlo Methods in Statistical Physics Berlin, Heidelberg, New York: Springer 1979
[2] J. G. Kirkwood and F. P. Buff, The Statistical Mechanical Theory of Solutions. I, J. Chem. Phys. 19, 774 (1951)
[3] Krűger, Peter and Vlugt, Thijs J. H., Size and shape dependence of finite-volume Kirkwood-Buff integrals, Phys. Rev. E.97, 051301 (2018)
[4] P. Ganguly and N. 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)
[5] J. L. Lebowitz and J. K. Percus, Long-range correlations in a closed system with applications to nonuniform fluids, Physical Review, 122(6):1675, (1961)
[6] P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, T. J. H. Vlugt, and J.-M. Simon, Kirkwood–Buff Integrals for Finite Volumes, Phys. Chem. Lett. 2013, 4, 2, 235–238 (2012)
[7] P. Krüger and T. J. H. Vlugt, Size and shape dependence of finite-volume Kirkwood-Buff integrals, Phys. Rev. E 97, 051301(R) (2018)
[8] A. Perera, Molecular emulsions: from charge order to domain order, Physical Chemistry Chemical Physics 19 (41), 28275-28285 (2017)
[9] W. Dednam and A. E. Botha, Evaluation of Kirkwood-Buff integrals via finite size scaling: a large scale molecular dynamics study, Journal of Physics: Conference Series 574 (2015) 012092.