

Robin Fingerhut, Jadran Vrabec

# Kirkwood-Buff integration: A promising route to entropic properties?

Journal article | Accepted manuscript (Postprint)

This version is available at <https://doi.org/10.14279/depositonce-8391>



Fingerhut, Robin; Vrabec, Jadran (2019). Kirkwood-Buff integration: A promising route to entropic properties? *Fluid Phase Equilibria*, 485, 270–281. <https://doi.org/10.1016/j.fluid.2018.12.015>

## Terms of Use

This work is licensed under a CC BY-NC-ND 4.0 License (Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International). For more information see <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

# Kirkwood-Buff integration: a promising route to entropic properties?

Robin Fingerhut and Jadran Vrabec\*

Thermodynamics and Process Engineering, Technical University Berlin, 10587 Berlin, Germany, \*E-Mail: vrabec@tu-berlin.de.

**KEYWORDS.** Kirkwood-Buff integration, molecular dynamics, binary mixtures, Lennard-Jones, methanol, ethanol, acetone, toluene, cyclohexane

---

**ABSTRACT:** Kirkwood-Buff integration (KBI) is implemented into the massively-parallel molecular simulation tool *ms2* and assessed by molecular dynamics simulations of binary liquid mixtures. The formalism of Krüger et al. (P. Krüger et al., J. Phys. Chem. Lett. 4: 235-238, 2013) that adopts  $NVT$  ensemble data to the  $\mu VT$  ensemble is employed throughout. Taking advantage of its linear scaling with inverse system size, the extrapolation to the thermodynamic limit is analyzed. KBI are calculated with standard radial distribution functions (RDF) and two corrected RDF forms. Simulations in the  $NVT$  ensemble are carried out in the entire composition range for four Lennard-Jones mixtures, studying system size dependence by varying  $N = 4000, 8000$  and  $16000$  molecules. Moreover, four mixtures of “real” components are considered with  $N = 4000$ . Thermodynamic factor, partial molar volumes and isothermal compressibility are calculated from KBI and compared with benchmark data from  $NpT$  ensemble simulations. The assessment shows that the formalism of Krüger et al. greatly improves KBI and that extrapolation is important, particularly for smaller systems.

---

## 1. Introduction

The knowledge of accurate thermodynamic property data is essential for the understanding of natural processes and the design of technical applications [1]. Longstanding efforts in experimental thermodynamics, however, led only to a scarce data base because laboratory work is time-consuming and costly, particularly when there are safety concerns [2]. With suitable computational methods and high-performance computing, molecular modeling and simulation may substantially contribute to the thermodynamic data supply in a much more efficient manner [3].

Liquid mixture properties are of vital importance, but their prediction is still challenging. For mass transport processes, the thermodynamic factor  $\Gamma$  plays an important role because it connects the Fick diffusion coefficient  $D$  with the Maxwell-Stefan diffusion coefficient  $\mathcal{D}$ . In case of binary mixtures, their relation is simply  $D = \Gamma \cdot \mathcal{D}$ . This factor  $\Gamma = 1 + x_1(\partial \ln \gamma_1 / \partial x_1)_{T,p}$ , where  $\gamma_1$  stands for the activity coefficient of component 1 and  $x_1$  its mole fraction, cannot be measured experimentally. Instead, it has to be extracted from phase equilibrium data with the support of excess Gibbs energy models or equations of state. It also poses a challenge for molecular simulation work because it is a derivative of typically noisy entropic properties. However, local density fluctuations may provide information on it [4]. In the grand canonical ( $\mu VT$ ) ensemble, the thermodynamic factor  $\Gamma$  can be analyzed by Kirkwood-

Buff integration (KBI) [5]. Many-body molecular simulations naturally provide detailed information on the microscopic structure of fluids by means of radial distribution functions (RDF)  $g_{ij}(r)$ . On that basis, it is conceptually straightforward to obtain liquid solution properties, like the thermodynamic factor  $\Gamma$  or partial molar volumes  $v_i$ . However, even 68 years after Kirkwood-Buff theory was conceived, it is still not fully clear how to fruitfully apply KBI as a standard method because it is defined for the  $\mu VT$  ensemble only [6], which is difficult to impose in the liquid state. Numerous efforts were thus undertaken to resolve this issue [4, 6-11]. The extrapolation to the thermodynamic limit, where all ensemble types become equal, is seen as a viable approach [7, 8]. However, for large intermolecular distances  $r$  RDF may not converge to unity. Accordingly, RDF corrections are crucial [6].

To study the strengths (and weaknesses) of KBI, we implemented it into the massively-parallel molecular simulation tool *ms2* [12-14]. Users can invoke KBI in the canonical ( $NVT$ ) ensemble both for Monte Carlo (MC) or molecular dynamics (MD) simulations. It was shown by Galata et al. [9] that ensembles containing a larger number of molecules  $N$  may lead to improved results. Thus, parallelization of KBI in *ms2* was essential because the computational effort of molecular simulation basically only depends on the number of molecules in the cutoff sphere. However, to determine which molecules are located within the cutoff sphere, the intermolecular interaction matrix

has to be traversed, which offers a convenient and efficient access to RDF sampling. Users of *ms2* can specify the frequency with which the RDF are sampled (each time step was chosen in this work). Further, the *ms2* implementation does not contain any additional *if* statements because RDF are sampled over the entire simulation volume, i.e. beyond the cutoff radius that is independently specified for evaluating the intermolecular interactions explicitly. The resulting RDF data are thus available also beyond half of the edge length of the simulation volume  $L/2$ . Because *ms2* assumes a cubic volume, RDF are sampled up to  $\sqrt{3}L/2$ , which may allow for recent extended schemes [7]. Due to parallelization and adequate implementation, the computational time effort for KBI is minor, cf. Table 1. The worst case scenario is given in case of computationally cheap intermolecular interactions, whereas for realistic molecules the execution time increases only by about 4 %. Moreover, KBI is carried out by *ms2* on the fly such that cumbersome post-processing is avoided.

This work focuses on RDF corrections and the extrapolation to the thermodynamic limit, motivated by the challenges mentioned above. For this purpose, KBI was applied to three types of RDF: First, standard RDF were used. Second, RDF were corrected as suggested by Ganguly and van der Vegt [15], referred to as vdV. This correction takes into account that molecules within a distance  $r$  around a central molecule may underlie excess or depletion phenomena and it was found to be superior to other schemes [6, 10]. Third, the vdV correction was applied in combination with a subsequent and typically tiny shift of the RDF in vertical direction through a constant multiplicative factor such that the mean value of the RDF is forced to be unity in the last quarter of its explicitly sampled range. This correction is denoted as vdV+shf.

KBI was derived for the  $\mu VT$  ensemble [5] that is hard to impose in simulations of dense liquid phases. Instead, the standard procedure for KBI is to employ the  $NVT$  ensemble, which, unfortunately, leads to convergence problems. To alleviate these, an integral truncation and correction was developed by Krüger et al. [11]. Because of its linear scaling with inverse system size, it is well suited for the extrapolation to infinite system size [11]. On that basis, KBI performance with and without extrapolation to macroscopic size was analyzed.

Different KBI approaches were assessed by comparing data on the thermodynamic factor  $\Gamma$ , partial molar volumes  $v_i$  and isothermal compressibility  $\beta_T$  with independent results. These properties were determined for a set of binary mixtures in their entire composition range with MD  $NVT$  ensemble simulations. The thermodynamic factor  $\Gamma$  from KBI was assessed on the basis of simulation data for chemical potentials  $\mu_i$  supported by excess Gibbs energy models. Partial molar volumes  $v_i$  and isothermal compressibility  $\beta_T$  from KBI were compared with results from isobaric-isothermal ( $NpT$ ) ensemble simulations. Four binary Lennard-Jones (LJ) mixtures in the liquid state, cf. Table 2, with a varying number of molecules  $N = 4000, 8000$  and  $16000$  were considered to study the system size dependence. Moreover, KBI performance was assessed for four binary liquid mixtures of real components with  $N = 4000$  molecules. These mixtures were chosen from ideal over non-ideal mixing behavior up to a mixture with liquid-liquid equilibrium (LLE) phase separation. Throughout, the entire composition range was considered.

## 2. Methodology

### 2.1 Kirkwood-Buff theory

RDF are the basis of KBI and they were sampled with *ms2* for the molecules' center of mass. The RDF  $g_{ij}(r)$  are closely related to what is known as local composition. In case of a mixture containing components  $i$  and  $j$ , they are given by

$$g_{ij}(r) = \frac{1}{x_j \rho} \cdot \frac{dN_i}{4\pi r^2 dr}, \quad (1)$$

where the local density of component  $i$  in a spherical shell with a radius  $r$  centered around species  $j$  is  $dN_i/4\pi r^2 dr$  and  $dN_i$  denotes the number of molecules of component  $i$  in that shell. The mixture density is  $\rho$  and  $x_j$  is the mole fraction of component  $j$ . A binary mixture of components 1 and 2 has thus three independent RDF, i.e.  $g_{11}(r), g_{22}(r), g_{12}(r) = g_{21}(r)$ . Standard RDF are given by eq. (1).

The vdV correction [15] takes the excess or depletion of molecular species  $j$  around a given molecule  $i$  at the distance  $r$  into account such that the asymptotic behavior of the RDF should yield an improved convergence to unity. It is given by

$$g_{ij}^{vdV}(r) = g_{ij}(r) \frac{N_j(1-V(r)/L^3)}{N_j(1-V(r)/L^3) - \Delta N_{ij}(r) - \delta_{ij}}, \quad (2)$$

where  $N_j$  is the number of molecules  $j$ ,  $L^3$  the cubic simulation volume,  $\delta_{ij}$  the Kronecker delta and  $V(r) = 4\pi r^3/3$ . Excess or depletion of species  $j$  around molecule  $i$  is given by  $\Delta N_{ij}(r) = \int_0^r 4\pi r'^2 \rho_j [g_{ij}(r') - 1] dr'$ .

Next, it was attempted to further improve the vdV corrected RDF  $g_{ij}^{vdV}(r)$  by a slight shift in vertical direction by a multiplicative factor. Shifting RDF was applied e.g. by Perera and Sokolić [16] and vdV+shf corrected RDF are simply given by

$$g_{ij}^{vdV+shf}(r) = \frac{g_{ij}^{vdV}(r)}{\bar{g}_{ij}^{vdV}}. \quad (3)$$

Therein,  $\bar{g}_{ij}^{vdV}$  is the mean value of the RDF  $g_{ij}^{vdV}(r)$  in the range  $r = 3L/8$  to  $L/2$ . Hence, the mean value of the RDF is forced to be unity in that range. Figure 1 exemplarily shows an RDF in its standard form and the two corrected forms for a binary LJ mixture.

KBI are defined in the  $\mu VT$  ensemble by [5]

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

In molecular simulation work, the infinite integration limit cannot be reached because  $g_{ij}(r)$  data can only be sampled up to some finite radius  $R$ . The truncation of eq. (4) was considered through a mathematical procedure developed by Krüger et al. [11]

$$G_{ij}(R) = \int_0^{2R} [g_{ij}(r) - 1] w(r, x) dr, \quad (5)$$

with the geometrical weight function  $w(r, x)$  for a three-dimensional space

$$w(r, x) = 4\pi r^2 \left(1 - \frac{3x}{2} + \frac{x^3}{2}\right), \quad (6)$$

where  $x = r/2R$ . Note that eq. (5) is not identical with the KBI in the thermodynamic limit [11]. With eqs. (5) and (6), KBI can be applied to finite simulation volumes in the  $NVT$  ensemble, considering that they are embedded in a larger reservoir. In this way, the exchange of mass and energy of the explicitly sampled subvolume with its surrounding is accounted for. Thereby, eq. (5) considers the  $\mu VT$  ensemble for which Kirkwood-Buff theory is defined [11]. Moreover, eq. (5) scales linearly with inverse system size  $R^{-1}$  so that the extrapolation to the thermodynamic limit  $V \rightarrow \infty$  is straightforward

$$G_{ij}^{\infty}(R) = \int_0^{2R} [g_{ij}(r) - 1] (w - x \frac{\partial w}{\partial x}) dr, \quad (7)$$

where  $w - x \frac{\partial w}{\partial x} = 4\pi r^2 (1 - x^3)$  in three-dimensional space when  $R^{-1} \rightarrow 0$ . However, recent studies [7] have indicated that the following KBI approximation convergences faster than the analytically derived eq. (7) in the thermodynamic limit,

$$G_{ij}^{\infty}(R) \approx \int_0^{2R} [g_{ij}(r) - 1] u_2(r, x) dr, \quad (8)$$

where  $u_2(r, x) = 4\pi r^2 (1 - 23x^3/8 + 3x^4/4 + 9x^5/8)$ . Figure 2 exemplarily shows  $G_{ij}(R)$  over inverse radius  $R^{-1}$  for a binary LJ mixture. It can be seen that eq. (5) does exhibit an almost linear behavior and that eq. (4) is of little use for extrapolation purposes. However, values of eq. (4) are close to the extrapolated data (eqs. (7) and (8)) when RDF corrections are applied.

In the present work, KBI were evaluated throughout with eq. (5) and extrapolated to macroscopic size with eqs. (7) and (8). RDF for KBI were computed from independent blocks during the productive simulation time (after equilibration) such that statistical uncertainties of  $G_{ij}$  are accessible (blocks of  $10^4$  time steps were used). Extrapolation was carried out on the basis of RDF that were sampled over the entire simulation time, i.e. they were not re-initialized with the block frequency. Additionally, we have tested the recent extended KBI scheme for cubic volumes up to  $L/\sqrt{2}$  by Krüger and Vlucht [7]. However, these results are not discussed in detail here because significant improvements were not found for present system sizes.

## 2.2 Assessment of Kirkwood-Buff integration

This assessment of KBI is based on three thermodynamic properties that can be directly calculated from  $G_{ij}$  data.

The thermodynamic factor  $\Gamma$  of a binary mixture of components 1 and 2 is given by [17]

$$\Gamma = 1 - \frac{x_1 x_2 \rho (G_{11} + G_{22} - 2G_{12})}{1 + x_1 x_2 \rho (G_{11} + G_{22} - 2G_{12})} \quad (9)$$

For comparison, chemical potentials  $\mu_i$  were calculated with  $NpT$  ensemble simulations and were sampled with Widom's test particle insertion [18] for the LJ mixtures and with thermodynamic integration [19, 20] for the real mixtures. Figure 3 (top) depicts the activity coefficients  $\gamma_i$  extracted from chemical potential data over the entire composition range of mixture LJ1. A fourth order polynomial was fitted to the  $\gamma_i$  data to derive the thermodynamic factor  $\Gamma$ . The accuracy of this fit for the thermodynamic factor  $\Gamma$  was estimated to be  $\pm 0.01$  in absolute terms. Well-known excess Gibbs energy models, such as Wilson, NRTL and Margules, were considered as well, however, they were insufficient to properly fit these statistically precise LJ simulation data. For quantitative analysis, the absolute relative average deviation (AAD) between  $\Gamma$  from KBI and benchmark (B) data was calculated over the entire composition range by

$$\text{AAD } \Gamma = \frac{100}{n} \cdot \sum_{x_1}^n \frac{|\Gamma_{\text{KBI}} - \Gamma_{\text{B}}|}{\Gamma_{\text{B}}} \quad (10)$$

The partial molar volumes  $v_1$  and  $v_2$  of a binary mixture are given by [17]

$$v_1 = (1 + \rho_2 (G_{22} - G_{12}))/\eta, \quad (11)$$

$$v_2 = (1 + \rho_1 (G_{11} - G_{12}))/\eta, \quad (12)$$

with  $\eta = \rho_1 + \rho_2 + \rho_1 \rho_2 (G_{11} + G_{22} - 2G_{12})$ . These properties were compared with results from  $NpT$  ensemble simulations such that a model-independent assessment of KBI was feasible.  $NpT$  simulations were carried out over the entire composition range to determine the varying mixture volume  $v$ . Secants were used to derive the partial molar volumes  $v_1$  and  $v_2$  because

three and four suffix Margules equations were insufficient to properly fit the statistically precise simulation data. Figure 3 (bottom) shows the excess volume  $v^E$  over the entire composition range of mixture LJ1. The mean deviation between simulation values and data from secants over the entire composition range is around 2 % in terms of  $v^E$ , which indicates that the partial molar volumes from secants are precise.

Finally, the isothermal compressibility  $\beta_T$  of a binary mixture is given by [17]

$$\beta_T k_B T = \frac{(1 + \rho_1 G_{11} + \rho_2 G_{22} + \rho_1 \rho_2 (G_{11} G_{22} - G_{12}^2))}{\eta}, \quad (13)$$

where  $k_B$  is the Boltzmann constant and  $T$  the temperature. This property was also compared with results from  $NpT$  ensemble simulations, where it can be sampled from volume fluctuations [21]. The mean statistical uncertainty of these benchmark data for the LJ mixtures is about 0.5 % and for the real mixtures about 3 % (cf. Table S.1 in the Supplementary Material)

## 3. Binary Lennard-Jones mixtures

Four binary LJ mixtures were selected to study the performance of KBI for varying system size. Table 2 shows their LJ parameters for size  $\sigma$  and energy  $\epsilon$  as well as the studied temperature  $T$  and pressure  $p$  pairs. These mixtures are characterized by a variation of LJ size parameter  $\sigma$ , LJ energy parameter  $\epsilon$  or both and a combination that shall stand for a real mixture, i.e. acetone + carbon dioxide [22]. All mixtures were simulated with *ms2* [12–14], throughout with the same settings. MC  $NpT$  simulations were performed to determine the mixture density. The KBI  $G_{ij}$  were calculated at that density with MD  $NVT$  simulations in the mixtures' mole fraction range  $x_1 = 0.05, 0.1, 0.2, \dots, 0.9, 0.95$  mol mol<sup>-1</sup>. Simulations for all four mixtures, denoted as LJ1, LJ2, LJ3 and LJ4, were carried out with  $N = 4000, 8000$  and 16000 molecules, always for a period of  $1.5 \cdot 10^7$  time steps. This sampling duration is extensive, but still reasonable for routine applications of MD. In the following, KBI performance is discussed for mixture LJ1 only. Analogous results for the other LJ mixtures are shown in the Supplementary Material. Therein, all MD  $NVT$  simulation results are tabulated.

Figure 4 shows  $G_{ij}$  over mole fraction  $x_1$ . From this figure, it is not possible to properly assess the KBI approaches because there are no physical benchmark data. However,  $G_{ij}$  are shown here because they are the input for all properties that are discussed below. It was found that the  $G_{ij}$  data scatter over  $x_1$  although they should form a continuous smooth curve. It becomes clear from Figure 4 that the better the RDF statistics, the smaller the difference between extrapolated and non-extrapolated values. For instance, in the composition range  $x_1 \rightarrow 1$ , the difference between extrapolated and non-extrapolated KBI  $G_{11}$  is small for all system sizes. On the other hand, extrapolated KBI differ significantly from non-extrapolated KBI if the RDF statistics is poor (i.e. for the diluted component  $G_{22}$  at  $x_1 \rightarrow 1$ ). Moreover, pure component KBI  $G_{11}$  and  $G_{22}$  at molar fractions  $x_1 \rightarrow 0$  or  $x_1 \rightarrow 1$ , respectively, are more uncertain with increasing system size. This can be explained by RDF noise for larger  $r$  with increased system size. Comparing both types of extrapolations  $G_{ij}^{\infty}$  (eqs. (7) and (8)) reveals that  $G_{ij}^{\infty}$  based on eq. (8) are closer to the non-extrapolated  $G_{ij}$ .

Figure 5 (top) shows the thermodynamic factor  $\Gamma$  derived from KBI over mole fraction  $x_1$  for varying system size  $N = 4000, 8000$  and 16000. The thermodynamic factor  $\Gamma$  based on

non-extrapolated  $G_{ij}$  with standard RDF has large deviations from the benchmark data and values based on  $G_{ij}^\infty$  deviate even more. Data based on RDF that were corrected with vdV or vdV+shf are similar and in good agreement with the benchmark if extrapolated  $G_{ij}^\infty$  are used. For extrapolation purposes, eq. (8) should be applied because results are more convincing compared to those based on eq. (7). It becomes clear that the deviation between  $\Gamma$  from KBI and the benchmark decreases from  $N = 4000$  to  $N = 8000$  molecules. For system size  $N = 8000$ , a slightly better agreement between  $\Gamma$  data based on extrapolated  $G_{ij}^\infty$  can be seen for vdV+shf corrected RDF than for vdV corrected RDF. The even larger system size  $N = 16000$  did not lead to further improvements.  $\Gamma$  based on  $G_{ij}^\infty$  (eq. (8)) with vdV+shf corrected RDF show the most adequate results for system size  $N = 4000$ . Table 3 gives an overview on AAD for all  $\Gamma$  data.

Figure 5 (center) presents the partial molar volumes  $v_1$  and  $v_2$  over mole fraction  $x_1$ . For all types of underlying RDF, the partial molar volumes  $v_i$  from extrapolated  $G_{ij}^\infty$  are in much better agreement with  $NpT$  simulations data than from  $G_{ij}$ , which holds for system sizes  $N = 4000$  and  $8000$ , whereas system size  $N = 16000$  did not lead to further improvements. Partial molar volumes  $v_i$  from non-extrapolated KBI show better results for increased system size.  $v_i$  from extrapolated KBI are more suitable, if RDF are corrected by vdV compared to standard RDF. Partial molar volumes  $v_i$  based on  $G_{ij}^\infty$ , particularly from eq. (8), with underlying vdV+shf corrected RDF show slightly better results than vdV corrected RDF for most mole fractions.

Figure 5 (bottom) depicts the isothermal compressibility  $\beta_T$  over mole fraction  $x_1$ . Throughout, isothermal compressibility  $\beta_T$  data from KBI are in poor agreement with  $NpT$  simulations. This result can be seen for all system sizes and all RDF approaches. In part,  $\beta_T$  derived from KBI are negative, which violates the thermodynamic stability condition. It can be summarized that this property should not be calculated with KBI because it can be straightforwardly sampled in the  $NpT$  ensemble, which is in accordance with the findings of Weerasinghe and Smith [23].

#### 4. Binary mixtures of real components

Four binary mixtures of real components were selected to analyze the performance of KBI for one system size  $N = 4000$ : methanol + ethanol, acetone + methanol, methanol + toluene and cyclohexane + methanol. Rigid, united-atom type models were used to describe their intermolecular interactions. The applied molecular models account for these interactions, including hydrogen bonding, by a set of LJ sites, point charges, point dipoles and point quadrupoles, which may or may not coincide with the LJ site positions [24-28]. These force fields were satisfactorily assessed in previous works and the interested reader is referred to the original publications [24-28] for detailed information about the molecular pure substance models and their parameters. The four binary real mixtures were simulated with *ms2* [12-14], throughout with the same settings, see Appendix. MD  $NpT$  simulations at 298.15 K and 0.1 MPa were carried out to calculate the mixture density and chemical potentials  $\mu_i$  were sampled simultaneously with thermodynamic integration [19, 20] and the Wilson model [29] was fitted to these data. Clearly, these chemical potential data are not as accurate as those of the

LJ mixtures. The accuracy of the Wilson fit in terms of the thermodynamic factor  $\Gamma$  was estimated to be below  $\pm 0.04$  for methanol + ethanol and for the others about  $\pm 0.1$ . The  $NpT$  simulations indicated that methanol + toluene and cyclohexane + methanol have a negative excess volume  $v^E$ , whereas the other two mixtures are volumetrically almost ideal. Secants were used to derive the partial molar volumes  $v_1$  and  $v_2$ . KBI  $G_{ij}$  were sampled with MD  $NVT$  simulations in the entire composition range  $x_1 = 0.05, 0.1, 0.2, \dots, 0.9, 0.95$  mol mol<sup>-1</sup>.

Figure 6 shows the thermodynamic factor  $\Gamma$ , partial molar volumes  $v_1$  and  $v_2$  and the isothermal compressibility  $\beta_T$  over mole fraction  $x_1$  of methanol + ethanol. Thermodynamic factor  $\Gamma$  predictions from NRTL [30], UNIQUAC [31], Wilson [29] models fitted to experimental vapor-liquid equilibrium data as well as Ref. [32] confirm the ideal behavior ( $\Gamma \approx 1$ ) of that mixture. The thermodynamic factor  $\Gamma$  from  $G_{ij}^\infty$  (eqs. (7) and (8)) based on standard RDF shows large deviations from that benchmark. Non-extrapolated  $G_{ij}$  from the other RDF approaches yield better results. Throughout, data for  $\Gamma$  based on KBI scatter around unity. It can be seen that results based on  $G_{ij}$  are in best agreement, where the vdV and vdV+shf corrected RDF lead to the most convincing results. Table 4 gives an overview on AAD for all  $\Gamma$  data. Partial molar volumes  $v_1$  and  $v_2$  from non-extrapolated KBI exhibit for all RDF approaches minor deviations from  $NpT$  simulation data, which are almost constant due to a negligible excess volume. Partial molar volumes based on  $G_{ij}^\infty$  are in good agreement, if the RDF were corrected with vdV+shf and  $G_{ij}^\infty$  approximated by eq. (8) lead to slightly better results than those based on eq. (7). Throughout, the isothermal compressibility  $\beta_T$  from extrapolated KBI is in poor agreement with the  $NpT$  simulations for all RDF approaches. Most of the  $\beta_T$  data from extrapolated KBI are even negative.

Figure 7 presents the results for acetone + methanol that has a moderately non-ideal mixture behavior. The thermodynamic factor  $\Gamma$  from  $G_{ij}^\infty$  (eq. (7)) based on the standard RDF shows the largest deviations from the Wilson model, whereas  $\Gamma$  from  $G_{ij}^\infty$  (eq. (8)) are improved. The vdV and vdV+shf corrected RDF without extrapolation leads to good results, however, the most adequate agreement is obtained by vdV corrected RDF. Both extrapolations  $G_{ij}^\infty$  (eqs. (7) and (8)) increase deviations from the benchmark for all types of RDF (cf. Table 4). Further,  $\Gamma$  data based on vdV corrected RDF and  $G_{ij}$  are most suitable in the pure component limits, where  $\Gamma$  must converge to unity. Partial molar volumes derived from non-extrapolated KBI show for all RDF approaches slightly smaller deviations from the almost constant  $NpT$  simulation data than their extrapolations. Partial molar volumes from  $G_{ij}^\infty$  (eq. (8)) are in good agreement, if RDF are corrected with vdV+shf. Throughout, the isothermal compressibility  $\beta_T$  data from extrapolated KBI (all RDF approaches) are in poor agreement with the  $NpT$  simulations and are in part below zero.

Figure 8 shows the results for methanol + toluene, which has a non-ideal mixture behavior. The thermodynamic factor  $\Gamma$  from  $G_{ij}^\infty$  based on all RDF approaches is in good agreement with the Wilson model. Best agreement was found for extrapolated data (eq. (7)) with vdV corrected RDF and  $\Gamma$  data from  $G_{ij}^\infty$  are superior than those from  $G_{ij}$ . Table 4 gives an overview on AAD of  $\Gamma$  data. Partial molar volumes  $v_1$  and  $v_2$

based on  $G_{ij}$  show for all RDF approaches slightly smaller deviations from  $NpT$  simulation results than data based on  $G_{ij}^\infty$ . This mixture has a finite excess volume so that the benchmark data for  $v_i$  are not constant. Again, the isothermal compressibility  $\beta_T$  from extrapolated KBI (all RDF approaches) is in poor agreement with the  $NpT$  simulations and in part below zero.

Figure 9 shows the results for cyclohexane + methanol. This strongly non-ideal mixture exhibits a LLE over a wide composition range. Due to the strong thermodynamic factor  $\Gamma$  gradient in the composition range limits, it is challenging to see distinctions between extrapolation and non-extrapolation as well as between different RDF. However,  $\Gamma$  data based on vdV and vdV+shf corrected RDF are in good agreement. Most adequate results were achieved with vdV corrected RDF and  $G_{ij}^\infty$  based on eq. (7), cf. Table 4. It is interesting to note that KBI never yields negative numbers for the thermodynamic factor  $\Gamma$ . Cyclohexane + methanol exhibits a large excess volume. Partial molar volumes  $v_1$  and  $v_2$  from extrapolated KBI show for all RDF approaches smaller deviations from  $NpT$  simulation data than non-extrapolated KBI. Partial molar volumes from  $G_{ij}^\infty$  (eqs. (7) and (8)) are in best agreement for vdV+shf corrected RDF. Again, the isothermal compressibility  $\beta_T$  data from extrapolated KBI (all RDF approaches) are in poor agreement with the  $NpT$  simulations and are in part negative.

## 5. Conclusion

Kirkwood-Buff integration was applied to MD  $NVT$  simulations of four binary LJ mixtures and four binary mixtures of real components. The LJ mixtures were studied for varying system size  $N = 4000, 8000$  and  $16000$ . To calculate KBI for finite simulation volumes, the formalism of Krüger et al. [11] was used. The difference between extrapolated and non-extrapolated KBI was evaluated together with two extrapolation schemes (eqs. (7) and (8)). KBI was analyzed with standard RDF, vdV [15] and vdV+shf corrected RDF. Because the employed simulation code *ms2* assumes a cubic volume, RDF were sampled up to  $\sqrt{3}L/2$ , i.e. beyond half of the edge length of the simulation volume  $L/2$ . The recent extended KBI scheme for cubic volumes up to  $L/\sqrt{2}$  by Krüger and Vlucht [7] was tested. However, significant improvements were not found for present system sizes. The KBI study was based on results for thermodynamic factor  $\Gamma$ , partial molar volumes  $v_1, v_2$  and isothermal compressibility  $\beta_T$  that were compared with benchmark data from  $NpT$  simulations. In addition, excess Gibbs energy models were considered for the thermodynamic factor  $\Gamma$  of the real mixtures.

The assessment indicates that the standard KBI approach leads to large statistical uncertainties. Significant improvements were found by applying the formalism of Krüger et al. [11], which adopts  $NVT$  ensemble data to the  $\mu VT$  ensemble for which Kirkwood-Buff theory is defined. Moreover, it was found that standard RDF mostly provide little convincing results. The physically meaningful vdV correction of RDF enhances KBI, similar results were achieved by applying the vdV+shf correction.

The purpose of this work was to determine a KBI approach that is suitable to routinely sample the thermodynamic factor  $\Gamma$  which connects the Fick diffusion coefficient  $D$  with the Maxwell-Stefan diffusion coefficient  $\mathcal{D}$ . It was shown that predicting this property from KBI is particularly challenging

when a mixture is ideal. In this case (LJ3 and methanol + ethanol),  $\Gamma$  from non-extrapolated  $G_{ij}$  led to better results than from  $G_{ij}^\infty$ . In case of non-ideal LJ mixtures,  $\Gamma$  from  $G_{ij}^\infty$  showed the most convincing results with vdV+shf corrected RDF. In the case of the non-ideal mixtures of real components,  $\Gamma$  data based on vdV corrected RDF revealed the best agreement with extrapolation, except for acetone + methanol. Extrapolations with the approximation of KBI in the thermodynamic limit by eq. (8) lead to more promising results than eq. (7).

Partial molar volumes  $v_1$  and  $v_2$  were assessed with accurate benchmark data from  $NpT$  simulations. The LJ mixtures showed the best agreement with the benchmark data when extrapolated  $G_{ij}^\infty$  were applied with vdV+shf corrected RDF. No clear distinction could be made for the real mixtures. However, in most cases  $G_{ij}^\infty$  from vdV and vdV+shf corrected RDF were promising, except for cyclohexane + methanol, where standard RDF showed a better agreement.

The isothermal compressibility  $\beta_T$  was directly compared with  $NpT$  simulation data. For the LJ and real mixtures hardly any convincing results were obtained. Because this property can be easily sampled by  $NpT$  simulations, it should not be determined by KBI.

The system size dependence of KBI was assessed with LJ mixtures. Systems with  $N = 4000$  molecules were found to be sufficiently large, but in part better results were achieved with  $N = 8000$ . Further improvements were not obtained for systems with  $N = 16000$  molecules.

In summary, KBI is an important avenue to determine the thermodynamic factor  $\Gamma$  and should be applied to corrected RDF. The formalism of Krüger et al. [11] with extrapolation to the thermodynamic limit [7] is crucial for using KBI in conjunction with  $NVT$  ensemble simulations. KBI was implemented into the massively-parallel molecular simulation tool *ms2* [12–14]. Due to efficient implementation and parallelization, the execution time of mixtures of realistic molecules increases by about 4 % only, if KBI is invoked. Moreover, no post-processing is required.

## 6. Appendix - Simulation details

All simulations were carried out with the molecular simulation tool *ms2* [12–14]. For the four binary LJ liquid mixtures, MC  $NpT$  simulations were performed in the mole fraction range  $x_1 = 0.0, 0.025, \dots, 1.0$  mol mol<sup>-1</sup> with  $N = 1372$  molecules to determine the mixtures' density  $\rho$ . 25,000  $NVT$  and 40,000  $NpT$  equilibration cycles were made, while the production contained 4,000,000  $NpT$  cycles, consisting of 1372 displacement and one volume move (cf. Table 1). The chemical potentials  $\mu_i$  were sampled with Widom's test particle insertion [18]. During that MC run, also the isothermal compressibility  $\beta_T$  was calculated. Moreover, the mixture density  $\rho$  over mole fraction  $x_1$  was used to derive the partial molar volumes  $v_i$ . After the MC  $NpT$  runs, MD  $NVT$  simulations were carried out to compute the KBI  $G_{ij}$  for the four binary LJ liquid mixtures. Three system sizes containing  $N = 4000, 8000$  and  $16000$  molecules were sampled over the mole fraction range  $x_1 = 0.05, 0.1, 0.2, \dots, 0.9, 0.95$  mol mol<sup>-1</sup>. MD  $NVT$  simulations were performed with  $8 \cdot 10^5$  equilibration time steps and  $1.5 \cdot 10^7$  production time steps. RDF for KBI were sampled every time step, the KBI block frequency for error estimation was set to  $10^4$  time steps and RDF for extrapolation were sampled over the entire simulation time. The equations of motion were solved numerically by applying the

Gear predictor-corrector integrator [33] with a time step of  $\Delta t/\sigma_1\sqrt{m_1/\varepsilon_1} = 0.0003$ . Velocities were isokinetically rescaled to maintain a constant temperature. For all MC and MD simulations, the cutoff radius was set to  $5\sigma_1$  (exception LJ4:  $3.5\sigma_1$ ) and the intermolecular interactions were described with the LJ 12-6 potential. Unlike interactions were specified with the Lorentz-Berthelot combination rule [34, 35]. Beyond the cutoff radius, the LJ interactions were corrected with the standard method.

For the binary mixtures of real components, methanol + ethanol, acetone + methanol, methanol + toluene and cyclohexane + methanol, MD  $NpT$  simulations were performed in the mole fraction range  $x_1 = 0.0, 0.05, 0.1, 0.2, \dots, 0.9, 0.95, 1.0$  mol mol<sup>-1</sup> with  $N = 4000$  molecules to determine the mixture density  $\rho$ . Therefore,  $1.2 \cdot 10^5$  equilibration time steps and  $5 \cdot 10^5$  production time steps at 298.15 K and 0.1 MPa were carried out. The chemical potentials  $\mu_i$  were sampled with thermodynamic integration [19, 20] and the resulting data were used to fit the Wilson model [29]. During the MD  $NpT$  run, the isothermal compressibility  $\beta_T$  was calculated. Moreover, the mixture density  $\rho$  over the mole fraction  $x_1$  was used to derive the partial molar volumes  $v_i$ . It was found that the excess volume of methanol + ethanol and acetone + methanol is almost negligible, except for methanol + toluene and cyclohexane + methanol. Partial molar volumes of the latter two were calculated with secants. According to the MD  $NpT$  runs, MD  $NVT$  simulations were carried out to compute the KBI of the binary liquid mixtures. One system size  $N = 4000$  was considered. The MD  $NVT$  simulations were performed with  $8 \cdot 10^5$  equilibration time steps and  $1.5 \cdot 10^7$  production time steps. RDF for KBI were sampled every time step, the KBI block frequency for error estimation was set to 10,000 time steps and RDF for extrapolation was sampled over the entire simulation time. The equations of motion were solved numerically with Gear's predictor-corrector integrator [33] and a time step  $\Delta t = 0.877$  fs for all  $NpT$  and  $NVT$  simulations. The velocities were isokinetically rescaled to maintain a constant temperature. For all  $NpT$  and  $NVT$  simulations, the LJ 12-6 interactions beyond a cutoff radius of 17.5 Å were corrected by applying angle averaging [36]. The long-range electrostatic interactions were considered with the reaction field method [33]. Unlike interactions were specified with the Lorentz-Berthelot combination rule [34, 35]. The pressure of all  $NpT$  ensemble simulations was controlled with Andersen's barostat [37].

#### Acknowledgement

This research was funded by BMBF under the grant 01IH16008E "SkaSim: Skalierbare HPC-Software für molekulare Simulationen in der chemischen Industrie" and computational support was given by the High Performance Computing Center Stuttgart (HLRS) under the grant MMHBF2. Furthermore, we gratefully acknowledge the Paderborn Center for Parallel Computing (PC2) for the generous allocation of computer time on the OCuLUS cluster. We thank Jean-Marc Simon, Franjo Sokolić and Tomislav Primorac for their valuable support.

#### ASSOCIATED CONTENT

**Supplementary Material.** Additional data and figures. This material is available free of charge via the Internet at <http://.....>

#### AUTHOR INFORMATION

##### Corresponding Author

\* Jadran Vrabec, Thermodynamics, Technical University Berlin, 10587 Berlin, Germany, \*E-Mail: [vrabec@tu-berlin.de](mailto:vrabec@tu-berlin.de)

##### Funding Sources

Federal Ministry of Education and Research (BMBF)

#### REFERENCES

- [1] E. Hendriks, G.M. Kontogeorgis, R. Dohrn, J.C. de Hemptinne, I.G. Economou, L. Fele Zilnik, V. Vesovic, Industrial Requirements for Thermodynamic and Transport Properties, *Ind. Eng. Chem. Res.* 49 (2010) 11131–11141.
- [2] K.E. Gubbins, J.D. Moore, Molecular modeling of matter: Impact and prospects in Engineering, *Ind. Eng. Chem. Res.* 49 (2010) 3026–3046.
- [3] E.J. Maginn, From Discovery to Data: What Must Happen for Molecular Simulation to Become a Mainstream Chemical Engineering Tool, *AIChE J.* 55 (2009) 1304–1310.
- [4] M. Heidari, K. Kremer, R. Potestio, R. Cortes-Huerto, Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: Revisiting the spatial block analysis method, *Entropy.* 20 (2018) 222–238.
- [5] J.G. Kirkwood, F.P. Buff, The statistical mechanical theory of solutions. I, *J. Chem. Phys.* 19 (1951) 774–777.
- [6] J. Milzetti, D. Nayar, N.F.A. van der Vegt, Convergence of Kirkwood-Buff Integrals of Ideal and Non-Ideal Aqueous Solutions Using Molecular Dynamics Simulations, *J. Phys. Chem. B.* 122 (2018) 5515–5526.
- [7] P. Krüger, T.J.H. Vlugt, Size and shape dependence of finite volume Kirkwood-Buff integrals, *Phys. Rev. E.* 97 (2018) 051301.
- [8] S.K. Schnell, X. Liu, J.M. Simon, A. Bardow, D. Bedeaux, T.J.H. Vlugt, S. Kjelstrup, Calculating thermodynamic properties from fluctuations at small scales, *J. Phys. Chem. B.* 115 (2011) 10911–10918.
- [9] A.A. Galata, S.D. Anogiannakis, D.N. Theodorou, Thermodynamic analysis of Lennard-Jones binary mixtures using Kirkwood-Buff theory, *Fluid Phase Equilib.* 470 (2018) 25–37.
- [10] N. Dawass, P. Krüger, S.K. Schnell, D. Bedeaux, S. Kjelstrup, J.M. Simon, T.J.H. Vlugt, Finite-size effects of Kirkwood-Buff integrals from molecular simulations, *Mol. Simul.* 44 (2018) 599–612.

- [11] P. Krüger, S.K. Schnell, D. Bedeaux, S. Kjelstrup, T.J.H. Vlucht, J.M. Simon, Kirkwood-Buff integrals for finite volumes, *J. Phys. Chem. Lett.* 4 (2013) 235–238.
- [12] S. Deublein, B. Eckl, J. Stoll, S. V. Lishchuk, G. Guevara-Carrion, C.W. Glass, T. Merker, M. Bernreuther, H. Hasse, J. Vrabec, ms2: A molecular simulation tool for thermodynamic properties, *Comput. Phys. Commun.* 182 (2011) 2350–2367.
- [13] C.W. Glass, S. Reiser, G. Rutkai, S. Deublein, A. Köster, G. Guevara-Carrion, A. Wafai, M. Horsch, M. Bernreuther, T. Windmann, H. Hasse, J. Vrabec, ms2: A molecular simulation tool for thermodynamic properties, new version release, *Comput. Phys. Commun.* 185 (2014) 3302–3306.
- [14] G. Rutkai, A. Köster, G. Guevara-Carrion, T. Janzen, M. Schappals, C.W. Glass, M. Bernreuther, A. Wafai, S. Stephan, M. Kohns, S. Reiser, S. Deublein, M. Horsch, H. Hasse, J. Vrabec, ms2: A molecular simulation tool for thermodynamic properties, release 3.0, *Comput. Phys. Commun.* 221 (2017) 343–351.
- [15] P. Ganguly, N.F.A. Van Der Vegt, Convergence of sampling Kirkwood-Buff integrals of aqueous solutions with molecular dynamics simulations, *J. Chem. Theory Comput.* 9 (2013) 1347–1355.
- [16] A. Perera, F. Sokolić, Modeling nonionic aqueous solutions: The acetone-water mixture, *J. Chem. Phys.* 121 (2004) 11272–11282.
- [17] A. Ben-Naim, *Molecular theory of solutions*, Oxford University Press, Oxford, 2006.
- [18] B. Widom, Some Topics in the Theory of Fluids, *J. Chem. Phys.* 39 (1963) 2808–2812.
- [19] T.P. Straatsma, Free Energy Simulations, *Encycl. Comput. Chem.* V. 2 (1998) 1083–1089.
- [20] P. Jedlovszky, A. Idrissi, G. Jancsó, Can existing models qualitatively describe the mixing behavior of acetone with water?, *J. Chem. Phys.* 130 (2009) 124516.
- [21] T.L. Hill, *Statistical Mechanics*, Mc-Graw-Hill Book Company, New York, 1956.
- [22] A.Z. Panagiotopoulos, U.W. Suter, R.C. Reid, Phase Diagrams of Nonideal Fluid Mixtures from Monte Carlo Simulation, *Ind. Eng. Chem. Fundam.* 25 (1986) 525–535.
- [23] S. Weerasinghe, P.E. Smith, Kirkwood-Buff derived force field for mixtures of acetone and water, *J. Chem. Phys.* 118 (2003) 10663–10670.
- [24] T. Schnabel, A. Srivastava, J. Vrabec, H. Hasse, Hydrogen Bonding of Methanol in Supercritical CO<sub>2</sub>: Comparison between 1 H NMR Spectroscopic Data and Molecular Simulation Results, *J. Phys. Chem. B.* 111 (2007) 9871–9878.
- [25] T. Schnabel, J. Vrabec, H. Hasse, Henry’s law constants of methane, nitrogen, oxygen and carbon dioxide in ethanol from 273 to 498 K: Prediction from molecular simulation, *Fluid Phase Equilib.* 233 (2005) 134–143.
- [26] T. Windmann, M. Linnemann, J. Vrabec, Fluid phase behavior of nitrogen + acetone and oxygen + acetone by molecular simulation, experiment and the peng-robinson equation of state, *J. Chem. Eng. Data.* 59 (2014) 28–38.
- [27] G. Guevara-Carrion, T. Janzen, Y.M. Muñoz-Muñoz, J. Vrabec, Mutual diffusion of binary liquid mixtures containing methanol, ethanol, acetone, benzene, cyclohexane, toluene, and carbon tetrachloride, *J. Chem. Phys.* 144 (2016) 124501.
- [28] Y.M. Muñoz-Muñoz, G. Guevara-Carrion, M. Llano-Restrepo, J. Vrabec, Lennard-Jones force field parameters for cyclic alkanes from cyclopropane to cyclohexane, *Fluid Phase Equilib.* 404 (2015) 150–160.
- [29] G.M. Wilson, Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing, *J. Am. Chem. Soc.* 86 (1964) 127–130.
- [30] H. Renon, J.M. Prausnitz, Local compositions in thermodynamics excess functions for liquids mixtures, *AIChE J.* 14 (1968) 135–144.
- [31] D.S. Abrams, J.M. Prausnitz, Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems, *AIChE J.* 21 (1975) 116–128.
- [32] G. Guevara-Carrion, C. Nieto-Draghi, J. Vrabec, H. Hasse, Prediction of Transport Properties by Molecular Simulation: Methanol and Ethanol and Their Mixture, *J. Phys. Chem. B.* 112 (2008) 16664–16674.
- [33] D.J. Allen, M.P. Tildesley, *Computer simulation of liquids*, Oxford University Press, Oxford, 1989.
- [34] H.A. Lorentz, Über die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase, *Ann. Phys.* 248 (1881) 127–136.
- [35] D. Berthelot, Sur le mélange des gaz, *Comptes Rendus Hebd. Des Séances l’Académie Des Sci.* 126 (1898) 1703–1855.



- [36] R. Lustig, Angle-average for the powers of the distance between two separated vectors, Mol. Phys. 65 (1988) 175–179. Phys. 72 (1980) 2384–2393.
- [37] H.C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, J. Chem.

**Table 1. Scalar and parallel simulation using *ms2* with and without KBI for a LJ mixture and a real mixture with  $N = 4000$  molecules; execution time per 1000 MD time steps was measured on Hazel hen at the High Performance Computing Center Stuttgart; RDF were sampled every time step in case of MD *NVT* + KBI.**

mixture	cores	MD <i>NVT</i>	MD <i>NVT</i> + KBI
LJ1;	1	80.9 s	115.7 s
$x_1 = 0.5 \text{ mol mol}^{-1}$	48	2.7 s	3.8 s
methanol + toluene;	1	615.2 s	641.1 s
$x_1 = 0.2 \text{ mol mol}^{-1}$	48	22.8 s	23.5 s

**Table 2. LJ size and energy parameters of four binary mixtures, studied at the given temperature and pressure pairs.**

	$\sigma_2/\sigma_1$	$\varepsilon_2/\varepsilon_1$	$k_B T/\varepsilon_1$	$p\sigma_1^3/\varepsilon_1$
LJ1	1.5	0.75	0.85	0.03
LJ2	1.5	1.00	1.00	0.03
LJ3	1.0	0.75	0.75	0.03
LJ4	0.7677	0.5968	3.50	0.6286

**Table 3. AAD of thermodynamic factor data for the LJ mixtures; table elements in bold indicate the smallest AAD for a system size and RDF method field.**

RDF	KBI	LJ1			LJ2			LJ3			LJ4		
		4000	8000	16000	4000	8000	16000	4000	8000	16000	4000	8000	16000
standard	$\Gamma(G_{ij})$	<b>12.48</b>	<b>10.37</b>	<b>12.06</b>	<b>4.41</b>	<b>4.28</b>	<b>6.50</b>	<b>9.96</b>	<b>9.20</b>	<b>11.28</b>	<b>10.73</b>	<b>9.89</b>	<b>9.33</b>
	$\Gamma(G_{ij}^\infty \text{ eq. (7)})$	36.72	33.21	44.67	36.40	42.32	47.01	37.85	33.77	45.37	40.35	37.39	36.05
	$\Gamma(G_{ij}^\infty \text{ eq. (8)})$	19.66	17.52	23.81	18.91	22.52	25.49	19.65	18.34	24.90	23.58	21.25	20.19
vdV	$\Gamma(G_{ij})$	5.27	<b>3.21</b>	<b>4.65</b>	4.76	<b>3.38</b>	<b>2.62</b>	<b>3.72</b>	<b>2.19</b>	<b>4.40</b>	<b>3.73</b>	3.44	<b>3.11</b>
	$\Gamma(G_{ij}^\infty \text{ eq. (7)})$	7.10	5.81	10.73	7.09	7.42	11.25	9.09	4.79	12.57	4.06	3.30	4.35
	$\Gamma(G_{ij}^\infty \text{ eq. (8)})$	<b>4.77</b>	4.47	7.54	<b>4.59</b>	5.09	7.93	6.07	3.40	7.83	4.01	<b>3.12</b>	3.20
vdV+shf	$\Gamma(G_{ij})$	4.55	<b>3.41</b>	<b>4.40</b>	4.88	<b>3.80</b>	<b>2.34</b>	<b>2.36</b>	<b>2.69</b>	<b>3.96</b>	<b>3.75</b>	3.55	3.00
	$\Gamma(G_{ij}^\infty \text{ eq. (7)})$	3.50	5.26	8.75	3.40	6.70	6.09	4.75	4.92	7.43	4.06	<b>3.39</b>	3.13
	$\Gamma(G_{ij}^\infty \text{ eq. (8)})$	<b>2.77</b>	4.61	6.91	<b>2.91</b>	5.41	5.19	3.74	3.90	5.78	4.07	3.40	<b>2.81</b>

**Table 4. AAD of thermodynamic factor data for the mixtures of real components; table elements in bold indicate the smallest AAD for a RDF method field.**

RDF	KBI	methanol+ethanol	acetone+methanol	methanol+toluene	cyclohexane+methanol
		4000	4000	4000	4000
standard	$\Gamma(G_{ij})$	<b>7.92</b>	<b>7.13</b>	516.29	355.20
	$\Gamma(G_{ij}^\infty \text{ eq. (7)})$	39.20	24.96	443.95	292.25
	$\Gamma(G_{ij}^\infty \text{ eq. (8)})$	21.29	9.98	<b>399.46</b>	<b>278.85</b>
vdV	$\Gamma(G_{ij})$	<b>2.46</b>	<b>5.62</b>	465.61	322.03
	$\Gamma(G_{ij}^\infty \text{ eq. (7)})$	8.54	9.95	<b>317.36</b>	<b>222.64</b>
	$\Gamma(G_{ij}^\infty \text{ eq. (8)})$	5.81	9.37	329.79	237.40
vdV+shf	$\Gamma(G_{ij})$	<b>2.77</b>	<b>5.89</b>	467.15	343.52
	$\Gamma(G_{ij}^\infty \text{ eq. (7)})$	5.93	9.28	<b>315.90</b>	<b>263.92</b>
	$\Gamma(G_{ij}^\infty \text{ eq. (8)})$	5.11	9.20	330.17	266.41

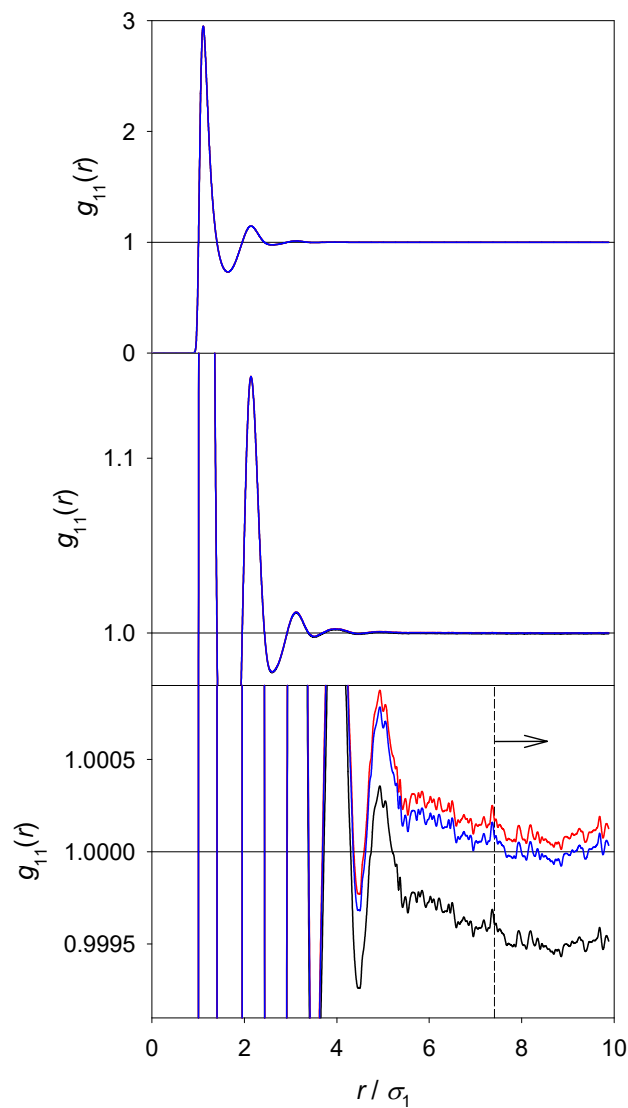


Figure 1. Correction schemes for RDF shown for a liquid state point of mixture LJ1 with mole fraction  $x_1 = 0.8 \text{ mol mol}^{-1}$ ; RDF were sampled every time step over a production period of  $1.5 \cdot 10^7$  time steps in the MD  $NVT$  ensemble containing  $N = 4000$  molecules; black line: standard RDF; red line: vdV corrected RDF; blue line: vdV+shf corrected RDF; center and bottom: magnified views, while the vertical dashed line delimits the shift range (bottom).

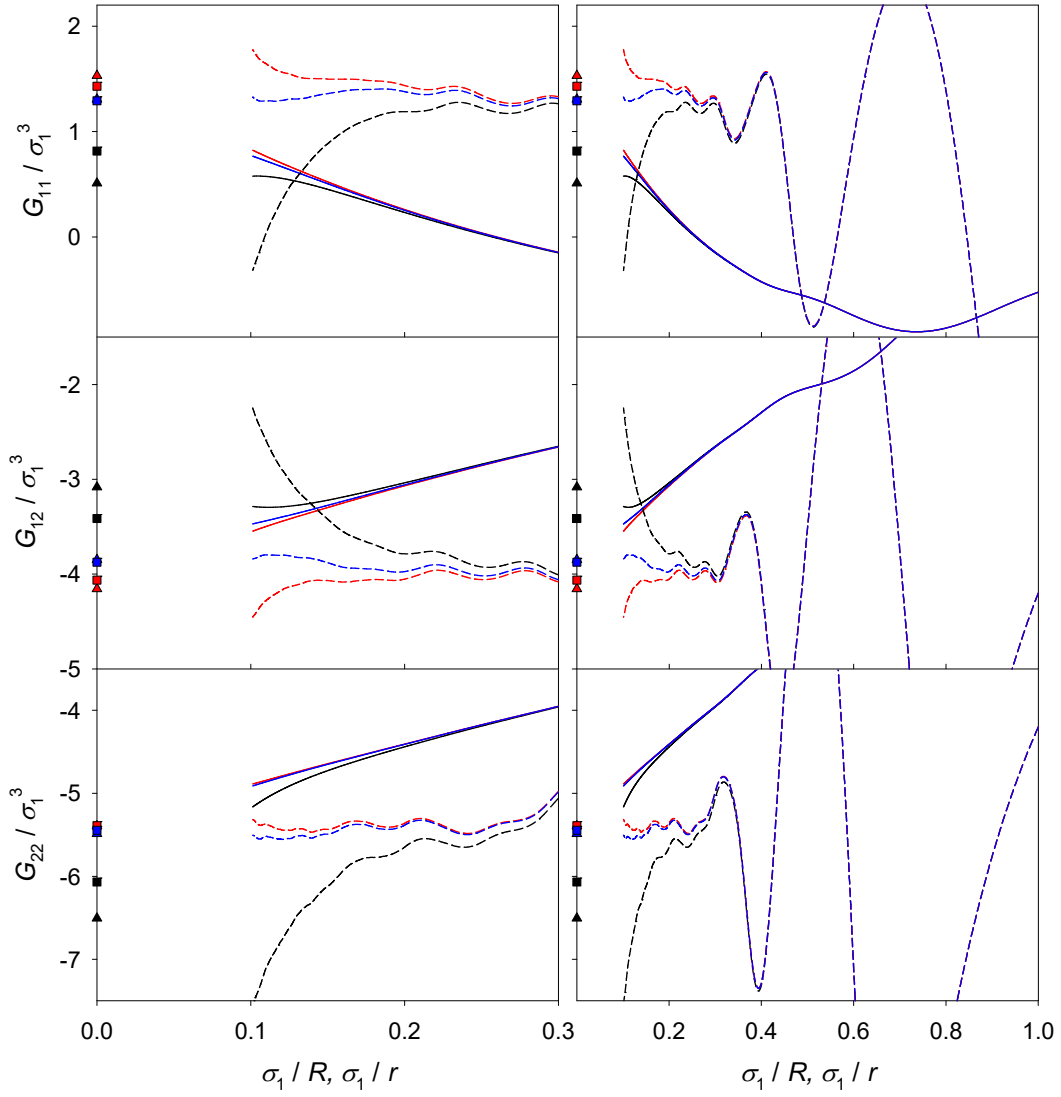


Figure 2.  $G_{ij}$  and  $G_{ij}^{\infty}$  over inverse radii  $r^{-1}$  and  $R^{-1}$  for a liquid state point of mixture LJ1 with mole fraction  $x_1 = 0.8 \text{ mol mol}^{-1}$ ; RDF were sampled every time step over a production period of  $1.5 \cdot 10^7$  time steps in the MD  $NVT$  ensemble containing  $N = 4000$  molecules; dashed lines:  $G_{ij}$  (eq. (4)); solid lines:  $G_{ij}$  (eq. (5)); triangles:  $G_{ij}^{\infty}$  (eq. (7)); squares:  $G_{ij}^{\infty}$  (eq. (8)); black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; left column: magnified view.

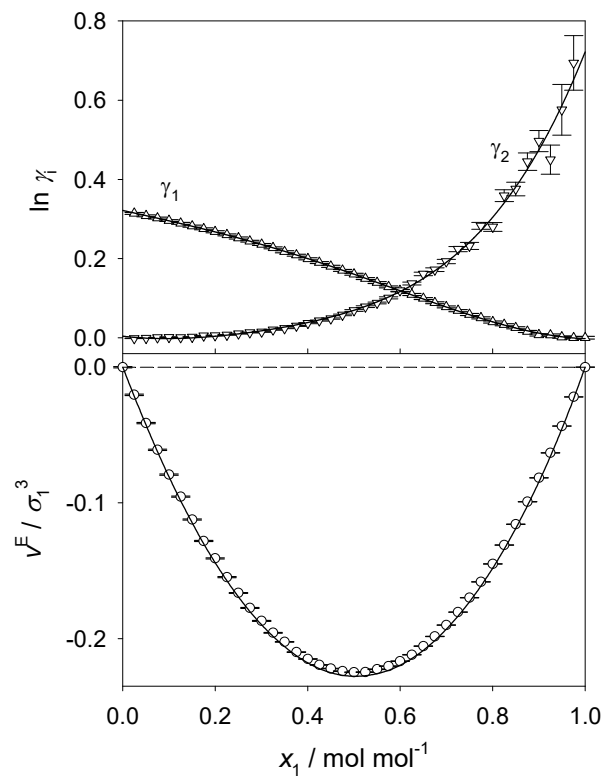


Figure 3. Activity coefficients  $\gamma_i$  (top) and excess volume  $v^E$  (bottom) over mole fraction  $x_1$  of mixture LJ1; triangles:  $\gamma_i$  derived from chemical potentials  $\mu_i$  sampled by Widom's test particle insertion [18] with  $NpT$  ensemble simulations; solid lines, top: fourth order polynomial fit; circles: simulation data  $v^E$ ; solid line, bottom:  $v^E$  calculated from the partial molar volumes  $v_1$  and  $v_2$  determined from secants.

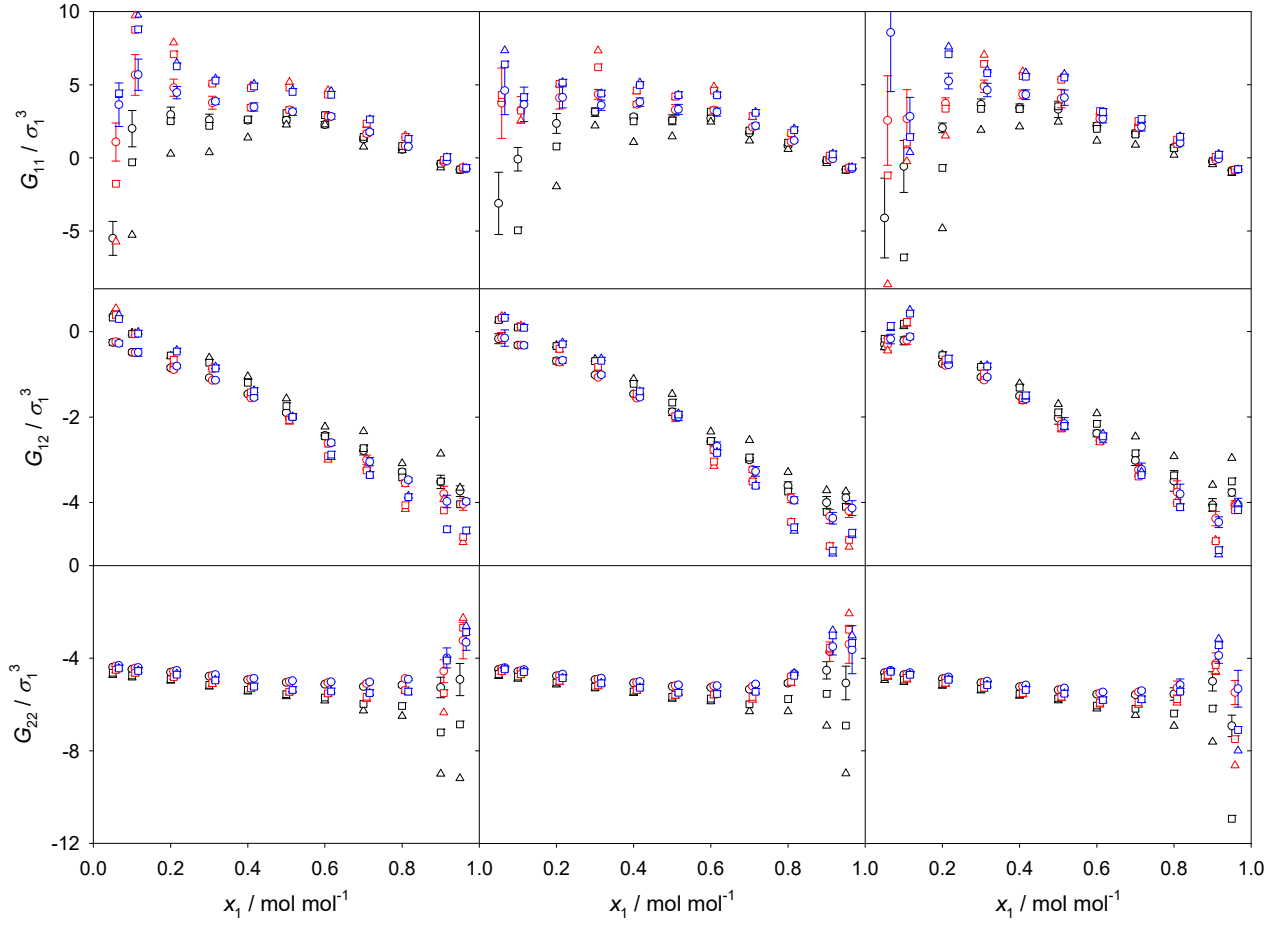


Figure 4.  $G_{ij}$  and  $G_{ij}^{\infty}$  over mole fraction  $x_1$  of mixture LJ1; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles:  $G_{ij}$  (eq. (5)) sampled via RDF in independent blocks; triangles (eq. (7)) and squares (eq. (8)):  $G_{ij}^{\infty}$  sampled via RDF calculated over the entire simulation time having the same statistical uncertainty as the circles; left:  $N = 4000$ ; center:  $N = 8000$ ; right:  $N = 16000$ ; data points at a given composition are slightly shifted in horizontal direction for visibility reasons.

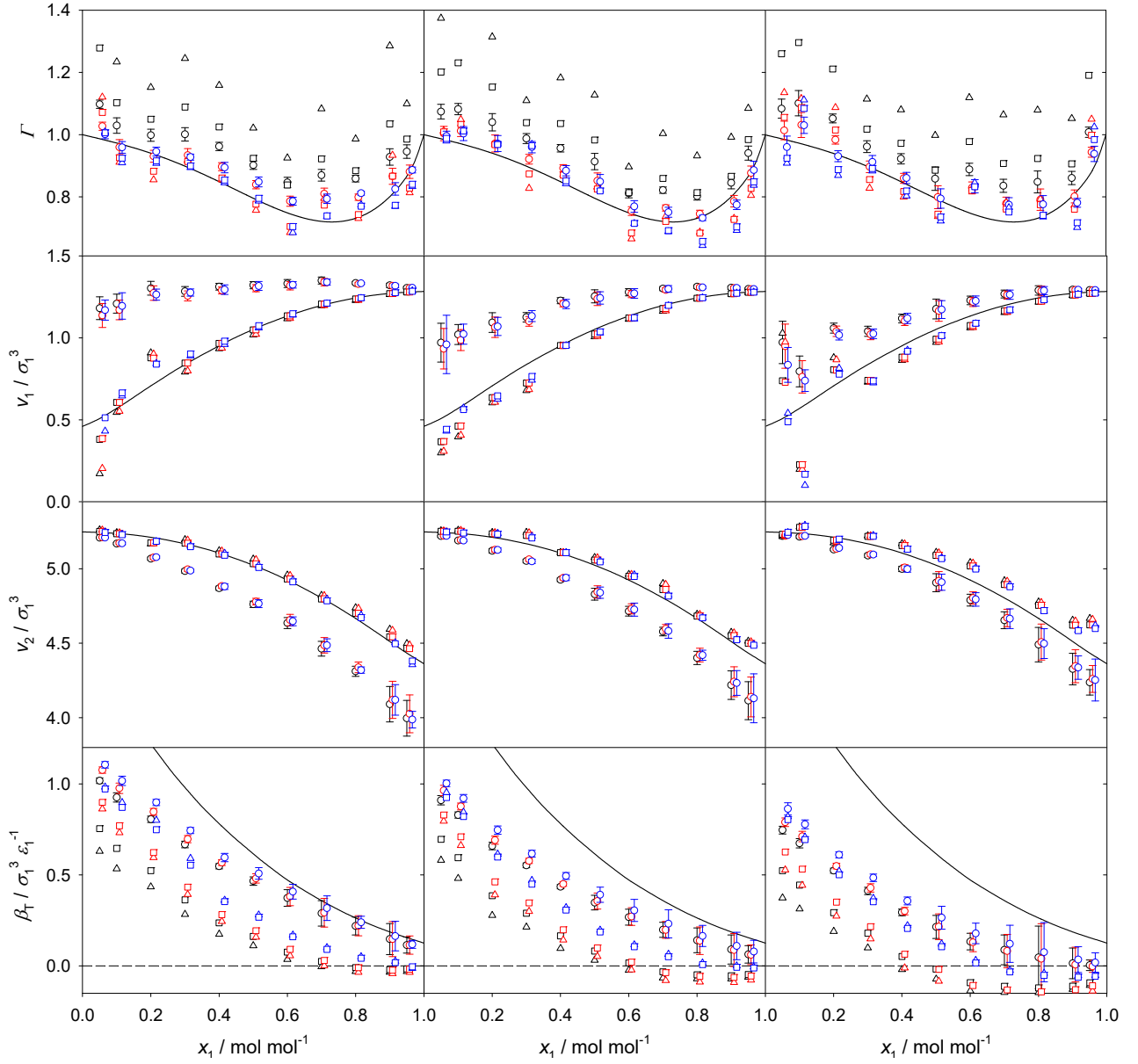


Figure 5. Thermodynamic factor  $\Gamma$  (top), partial molar volumes  $v_1$ ,  $v_2$  (center) and isothermal compressibility  $\beta_T$  (bottom) over mole fraction  $x_1$  of mixture LJ1; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles: data based on  $G_{ij}$ ; triangles (eq. (7)) and squares (eq. (8)): data based on  $G_{ij}^{\infty}$  having the same statistical uncertainty as the circles; left:  $N = 4000$ ; center:  $N = 8000$ ; right:  $N = 16000$ ; data points at given compositions are slightly shifted in horizontal direction for visibility reasons; solid lines: fourth order polynomial fitted to chemical potential  $\mu_i$  data sampled with Widom's test particle insertion [18] during  $NpT$  simulations (top); results from  $NpT$  simulations (below).

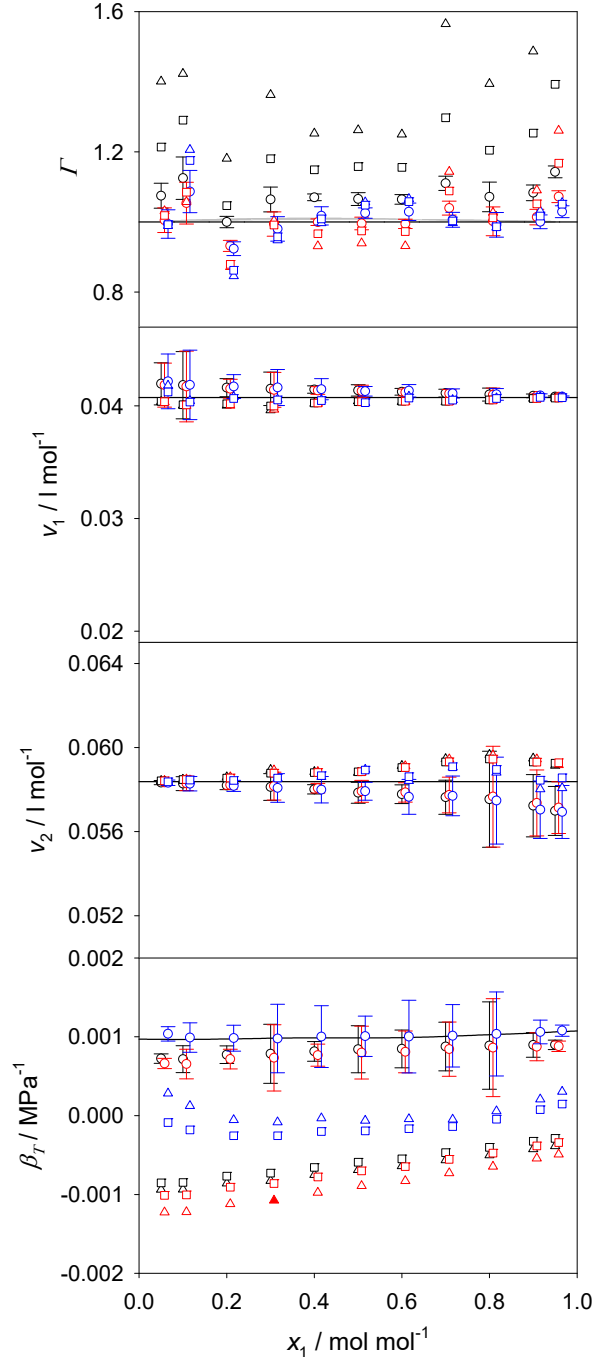


Figure 6. Thermodynamic factor  $\Gamma$  (top), partial molar volumes  $v_1$ ,  $v_2$  (center) and isothermal compressibility  $\beta_T$  (bottom) over mole fraction  $x_1$  of methanol + ethanol with  $N = 4000$ ; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles: data based on  $G_{ij}$ ; triangles (eq. (7)) and squares (eq. (8)): data based on  $G_{ij}^\infty$  having the same statistical uncertainty as the circles; data points at given compositions are slightly shifted in horizontal direction for visibility reasons; solid lines: Wilson model [29] fitted to chemical potential  $\mu_i$  data sampled with thermodynamic integration [19, 20] during  $NpT$  simulations (top); results from  $NpT$  simulations (below); hardly visible grey area: predictions from NRTL [30], UNIQUAC [31], Wilson [29] models fitted to experimental data [27] (top).



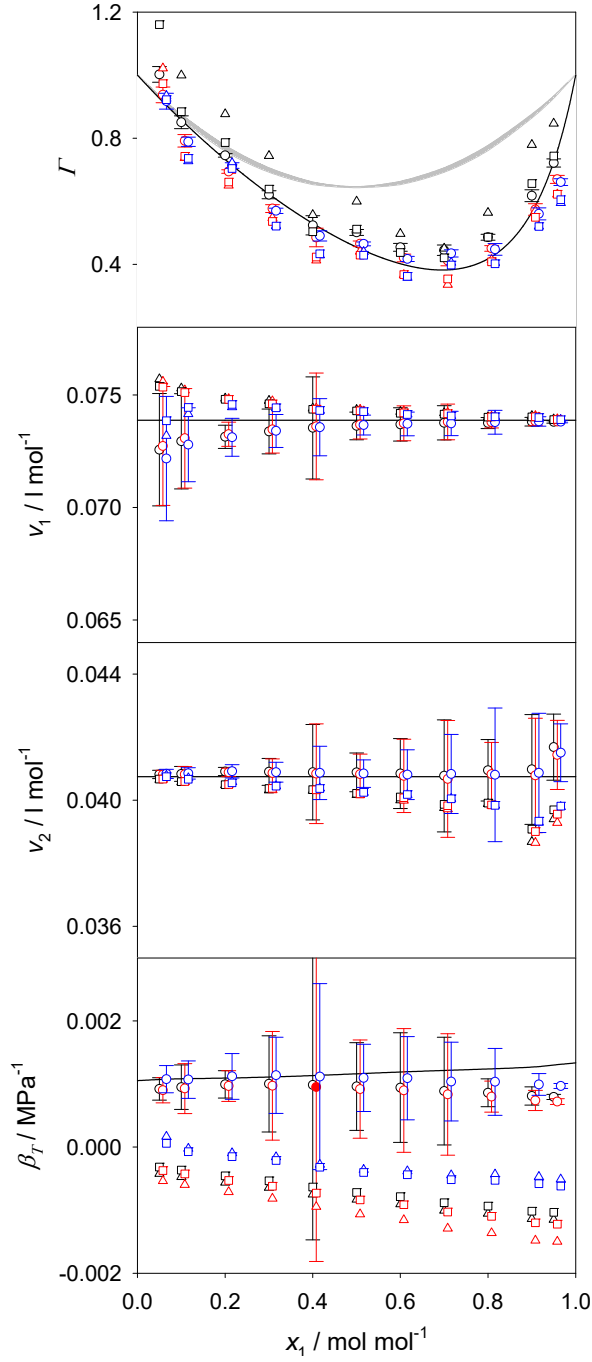


Figure 7. Thermodynamic factor  $\Gamma$  (top), partial molar volumes  $v_1$ ,  $v_2$  (center) and isothermal compressibility  $\beta_T$  (bottom) over mole fraction  $x_1$  of acetone + methanol with  $N = 4000$ ; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles: data based on  $G_{ij}$ ; triangles (eq. (7)) and squares (eq. (8)): data based on  $G_{ij}^\infty$  having the same statistical uncertainty as the circles; data points at given compositions are slightly shifted in horizontal direction for visibility reasons; solid lines: Wilson model [29] fitted to chemical potential  $\mu_i$  data sampled with thermodynamic integration [19, 20] during  $NpT$  simulations (top); results from  $NpT$  simulations (below); grey area: predictions from NRTL [30], UNIQUAC [31], Wilson [29] models fitted to experimental data [27] (top).

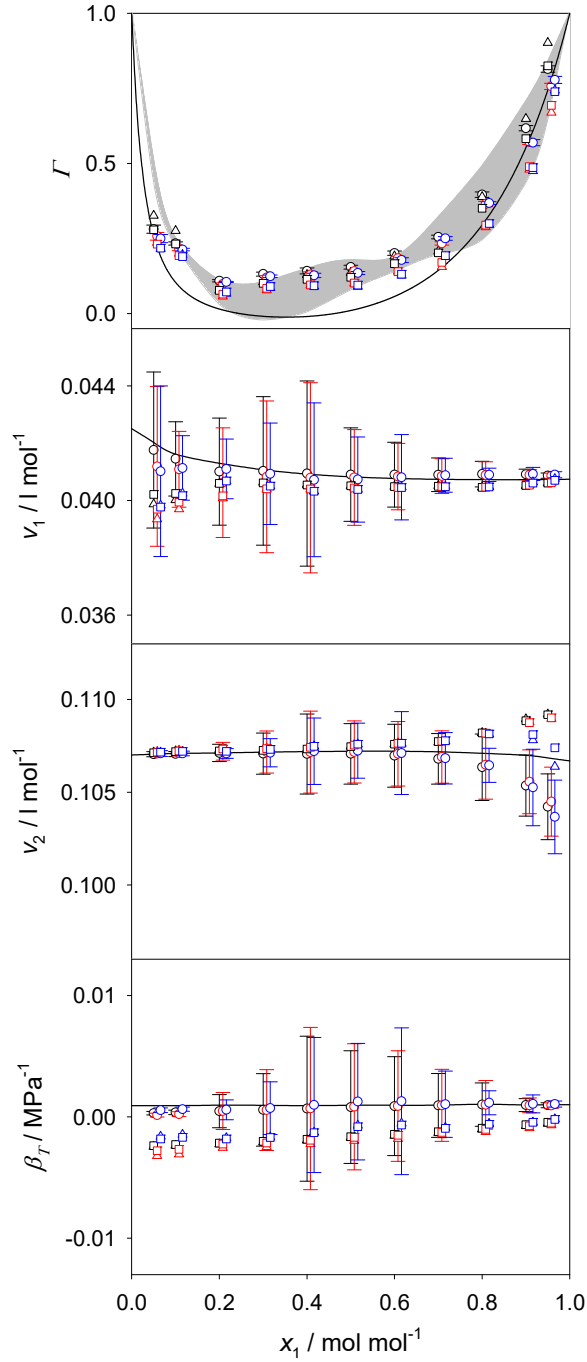


Figure 8. Thermodynamic factor  $\Gamma$  (top), partial molar volumes  $v_1$ ,  $v_2$  (center) and isothermal compressibility  $\beta_T$  (bottom) over mole fraction  $x_1$  of methanol + toluene with  $N = 4000$ ; standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles: data based on  $G_{ij}$ ; triangles (eq. (7)) and squares (eq. (8)): data based on  $G_{ij}^\infty$  having the same statistical uncertainty as the circles; data points at given compositions are slightly shifted in horizontal direction for visibility reasons; solid lines: Wilson model [29] fitted to chemical potential  $\mu_i$  data sampled with thermodynamic integration [19, 20] during  $NpT$  simulations (top); results from  $NpT$  simulations (below); grey area: predictions from NRTL [30], UNIQUAC [31], Wilson [29] models fitted to experimental data [27] (top).

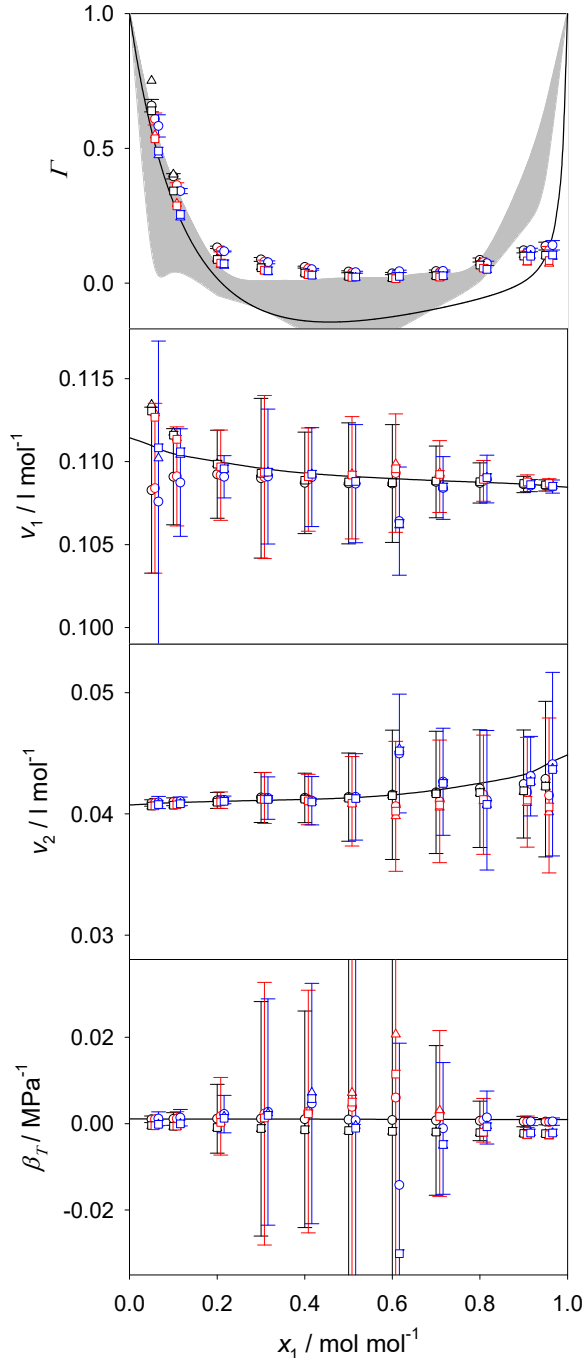


Figure 9. Thermodynamic factor  $\Gamma$  (top), partial molar volumes  $v_1$ ,  $v_2$  (center) and isothermal compressibility  $\beta_T$  (bottom) over mole fraction  $x_1$  of cyclohexane + methanol with  $N = 4000$ ; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles: data based on  $G_{ij}$ ; triangles (eq. (7)) and squares (eq. (8)): data based on  $G_{ij}^\infty$  having the same statistical uncertainty as the circles; data points at given compositions are slightly shifted in horizontal direction for visibility reasons; solid lines: Wilson model [29] fitted to chemical potential  $\mu_i$ ; data sampled with thermodynamic integration [19, 20] during  $NpT$  simulations (top); results from  $NpT$  simulations (below); grey area: predictions from NRTL [30], UNIQUAC [31], Wilson [29] models fitted to experimental data [27] (top).

# **Supplementary Material**

to

## **Kirkwood-Buff integration: a promising route to entropic properties?**

Robin Fingerhut and Jadran Vrabec<sup>\*</sup>

Thermodynamics and Process engineering, Technical University Berlin, 10587 Berlin, Germany, \*E-Mail: vrabec@tu-berlin.de.

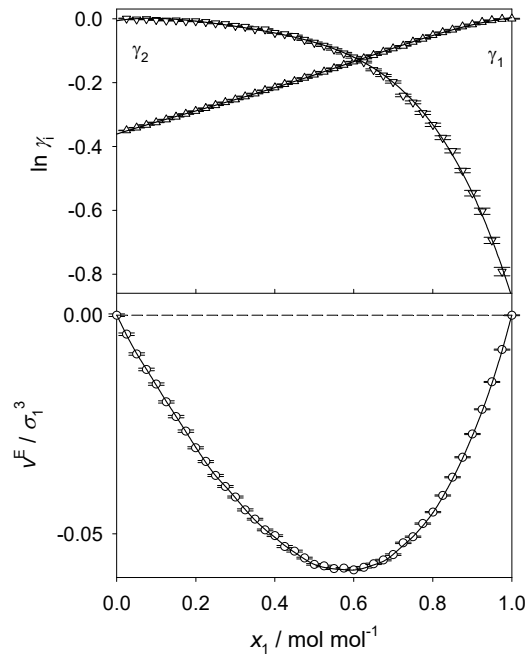


Figure S.1. Activity coefficients  $\gamma_i$  (top) and excess volume  $v^E$  (bottom) over mole fraction  $x_1$  of mixture LJ2; triangles:  $\gamma_i$  derived from chemical potentials  $\mu_i$  sampled by Widom's test particle insertion [18] with  $NpT$  ensemble simulations; solid lines, top: fourth order polynomial fit; circles: simulation data  $v^E$ ; solid line, bottom:  $v^E$  calculated from the partial molar volumes  $v_1$  and  $v_2$  determined from secants.

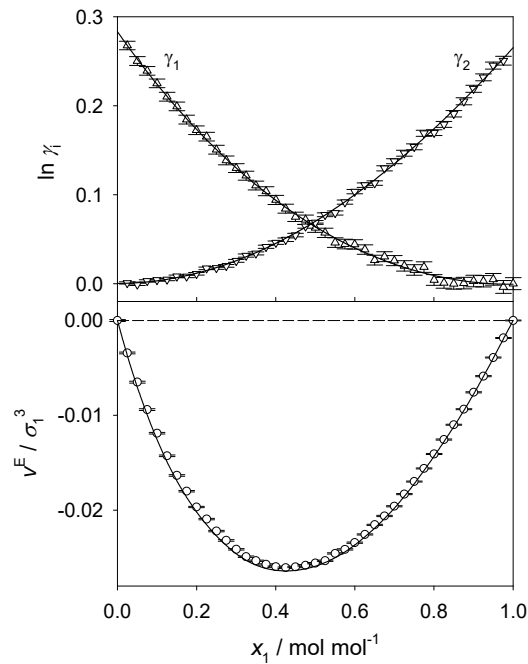


Figure S.2. Activity coefficients  $\gamma_i$  (top) and excess volume  $v^E$  (bottom) over mole fraction  $x_1$  of mixture LJ3; triangles:  $\gamma_i$  derived from chemical potentials  $\mu_i$  sampled by Widom's test particle insertion [18] with  $NpT$  ensemble simulations; solid lines, top: fourth order polynomial fit; circles: simulation data  $v^E$ ; solid line, bottom:  $v^E$  calculated from the partial molar volumes  $v_1$  and  $v_2$  determined from secants.

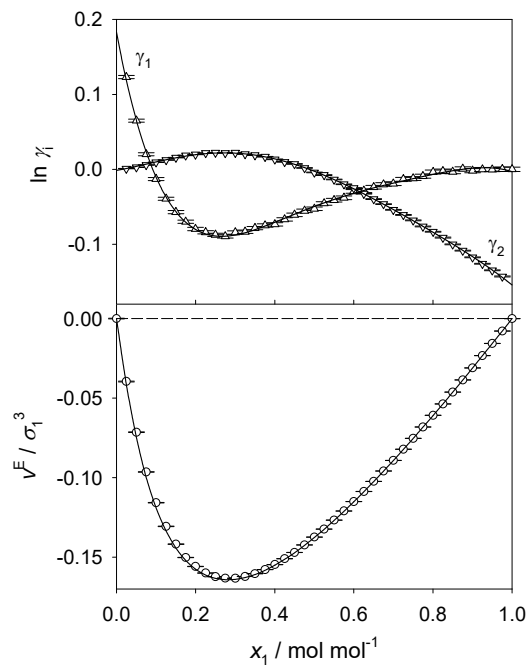


Figure S.3. Activity coefficients  $\gamma_i$  (top) and excess volume  $v^E$  (bottom) over mole fraction  $x_1$  of mixture LJ4; triangles:  $\gamma_i$  derived from chemical potentials  $\mu_i$  sampled by Widom's test particle insertion [18] with  $NpT$  ensemble simulations; solid lines, top: fourth order polynomial fit; circles: simulation data  $v^E$ ; solid line, bottom:  $v^E$  calculated from the partial molar volumes  $v_1$  and  $v_2$  determined from secants.

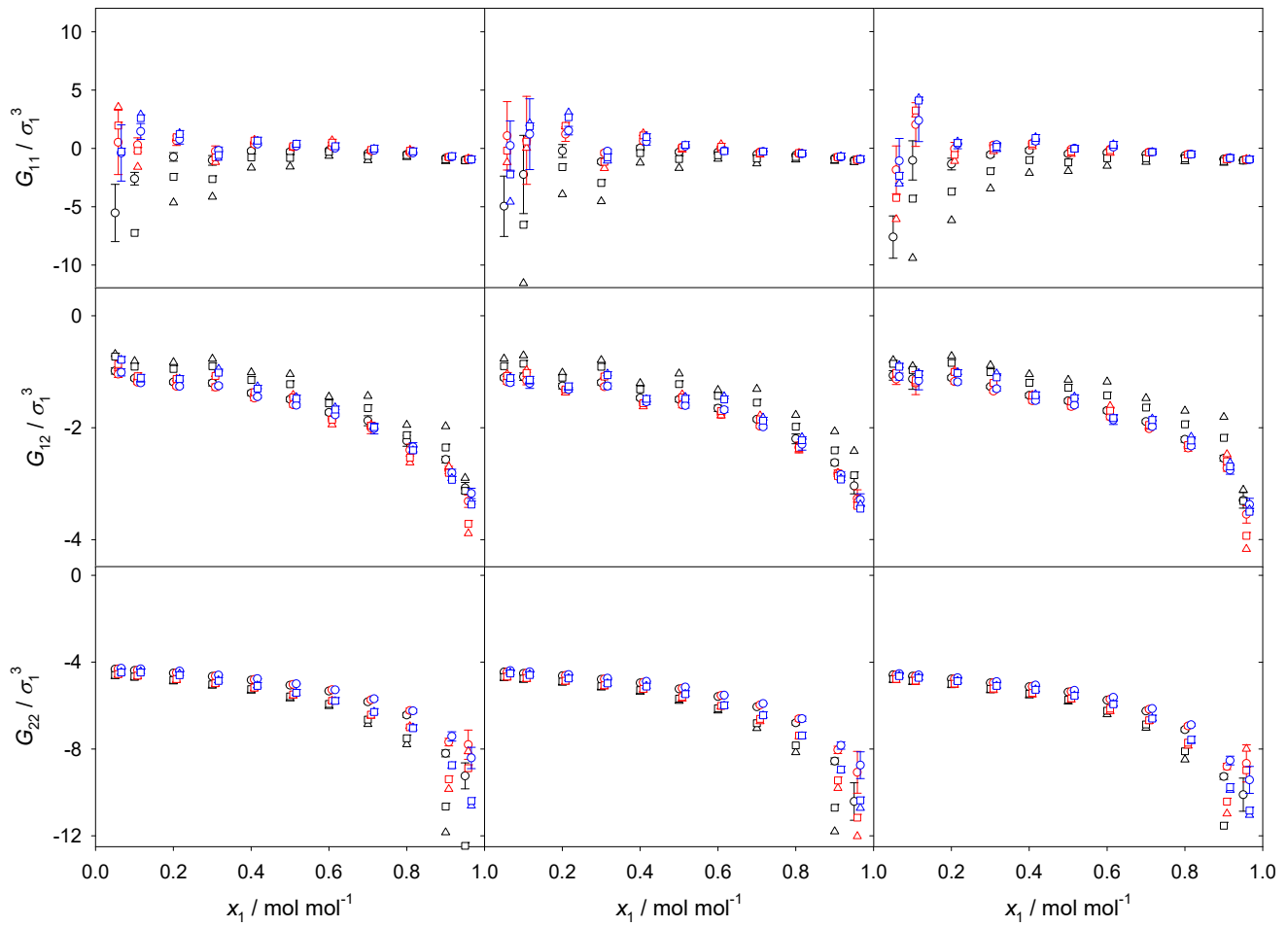


Figure S.4.  $G_{ij}$  and  $G_{ij}^{\infty}$  over mole fraction  $x_1$  of mixture LJ2; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles:  $G_{ij}$  sampled via RDF in independent blocks; triangles (eq. (7)) and squares (eq. (8)):  $G_{ij}^{\infty}$  sampled via RDF calculated over the entire simulation time having the same statistical uncertainty as the circles; left:  $N = 4000$ ; center:  $N = 8000$ ; right:  $N = 16000$ ; data points at a given composition are slightly shifted in horizontal direction for visibility reasons; for outliers cf. Excel table.

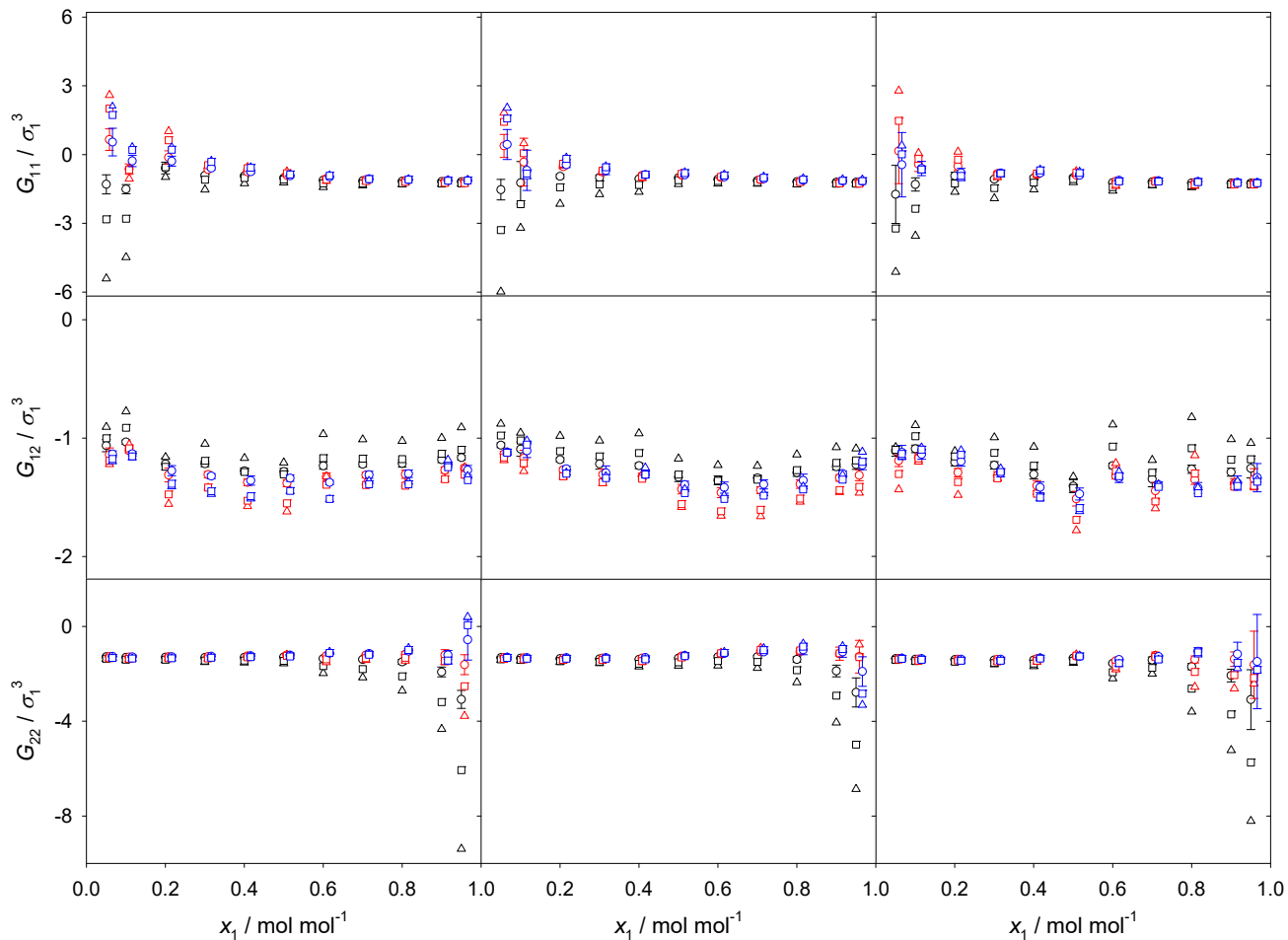


Figure S.5.  $G_{ij}$  and  $G_{ij}^{\infty}$  over mole fraction  $x_1$  of mixture LJ3; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles:  $G_{ij}$  sampled via RDF in independent blocks; triangles (eq. (7)) and squares (eq. (8)):  $G_{ij}^{\infty}$  sampled via RDF calculated over the entire simulation time having the same statistical uncertainty as the circles; left:  $N = 4000$ ; center:  $N = 8000$ ; right:  $N = 16000$ ; data points at a given composition are slightly shifted in horizontal direction for visibility reasons; for outliers cf. Excel table.



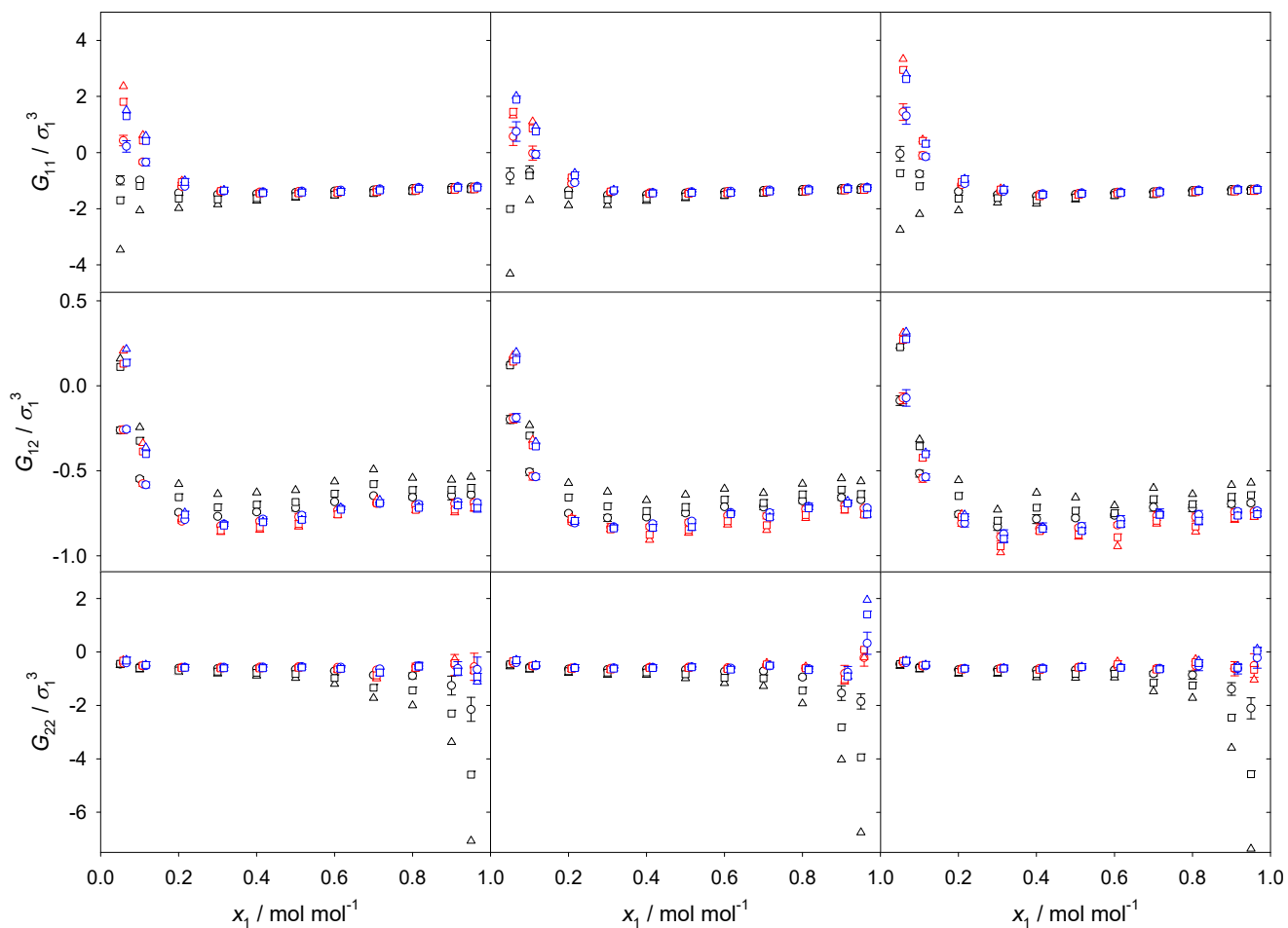


Figure S.6.  $G_{ij}$  and  $G_{ij}^{\infty}$  over mole fraction  $x_1$  of mixture LJ4; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles:  $G_{ij}$  sampled via RDF in independent blocks; triangles (eq. (7)) and squares (eq. (8)):  $G_{ij}^{\infty}$  sampled via RDF calculated over the entire simulation time having the same statistical uncertainty as the circles; left:  $N = 4000$ ; center:  $N = 8000$ ; right:  $N = 16000$ ; data points at a given composition are slightly shifted in horizontal direction for visibility reasons; for outliers cf. Excel table.

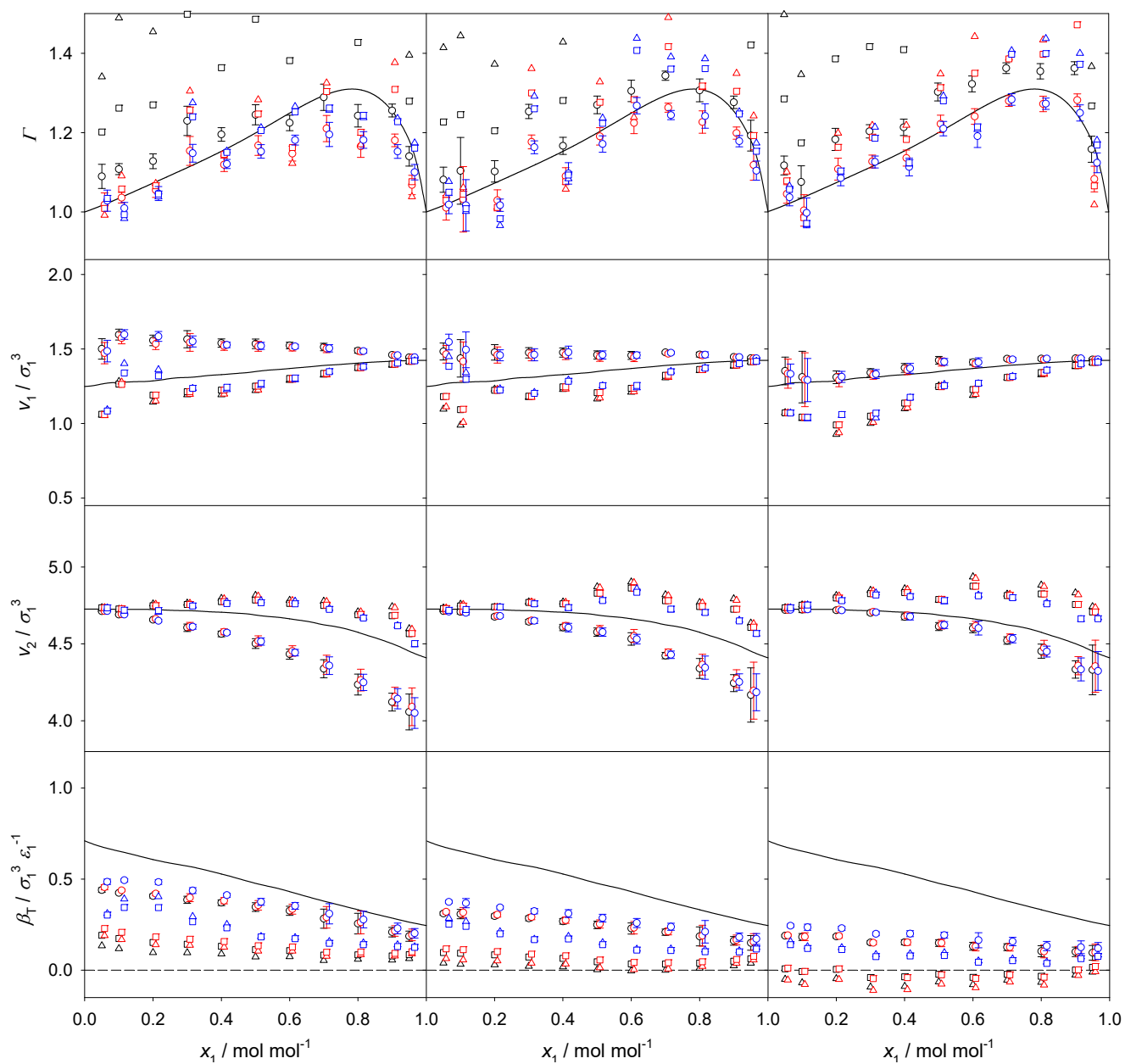
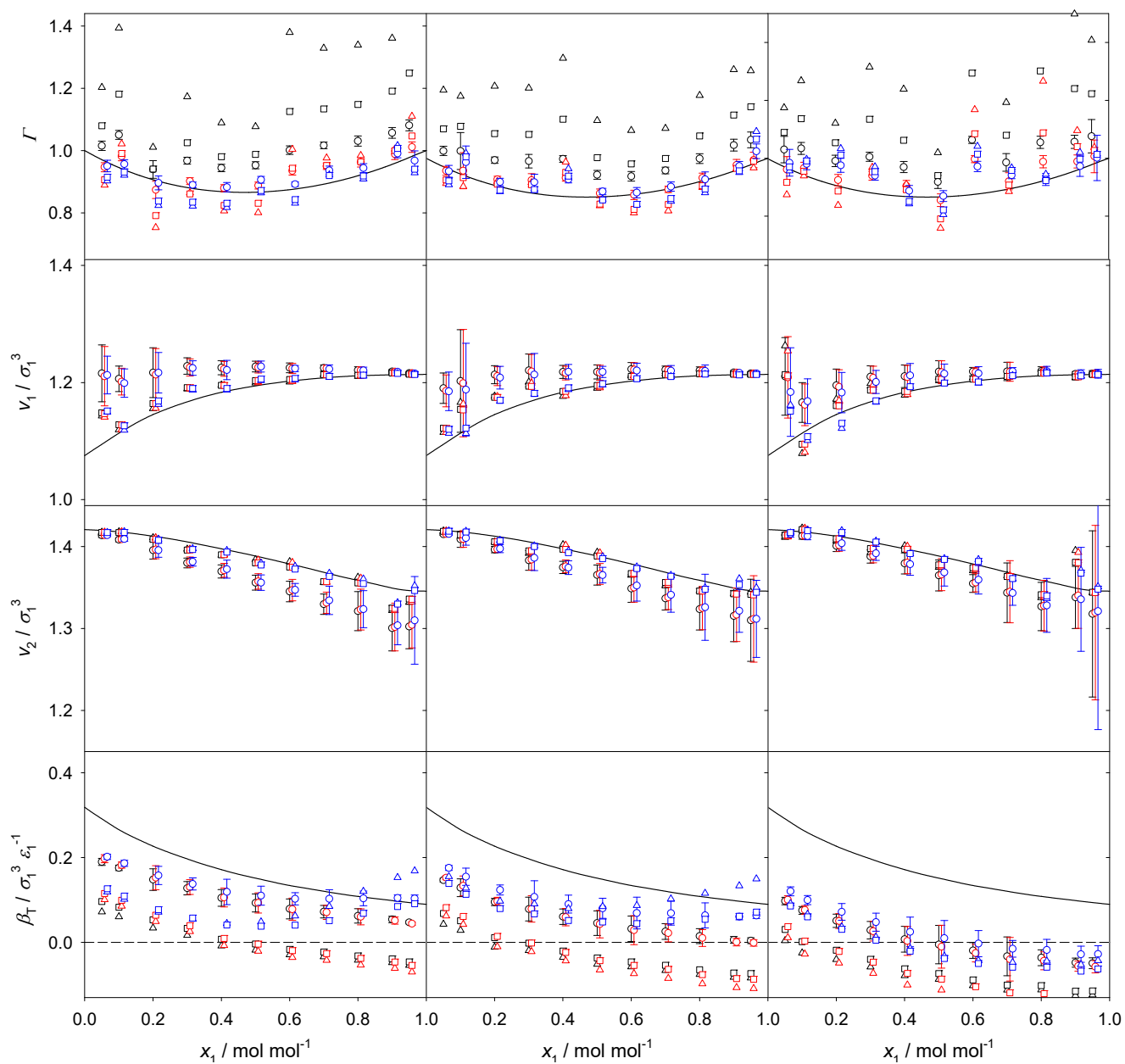


Figure S.7. Thermodynamic factor  $\Gamma$  (top), partial molar volumes  $v_1$ ,  $v_2$  (center) and isothermal compressibility  $\beta_T$  (bottom) over mole fraction  $x_1$  of mixture LJ2; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles: data based on  $G_{ij}$ ; triangles (eq. (7)) and squares (eq. (8)): data based on  $G_{ij}^\infty$  having the same statistical uncertainty as the circles; left:  $N = 4000$  molecules; center:  $N = 8000$  molecules; right:  $N = 16000$ ; data points at given compositions are slightly shifted in horizontal direction for visibility reasons; solid lines: fourth order polynomial fitted to chemical potential  $\mu_i$  data sampled with Widom's test particle insertion [18] during  $NpT$  simulations (top); results from  $NpT$  simulations (below); for outliers cf. Excel table..



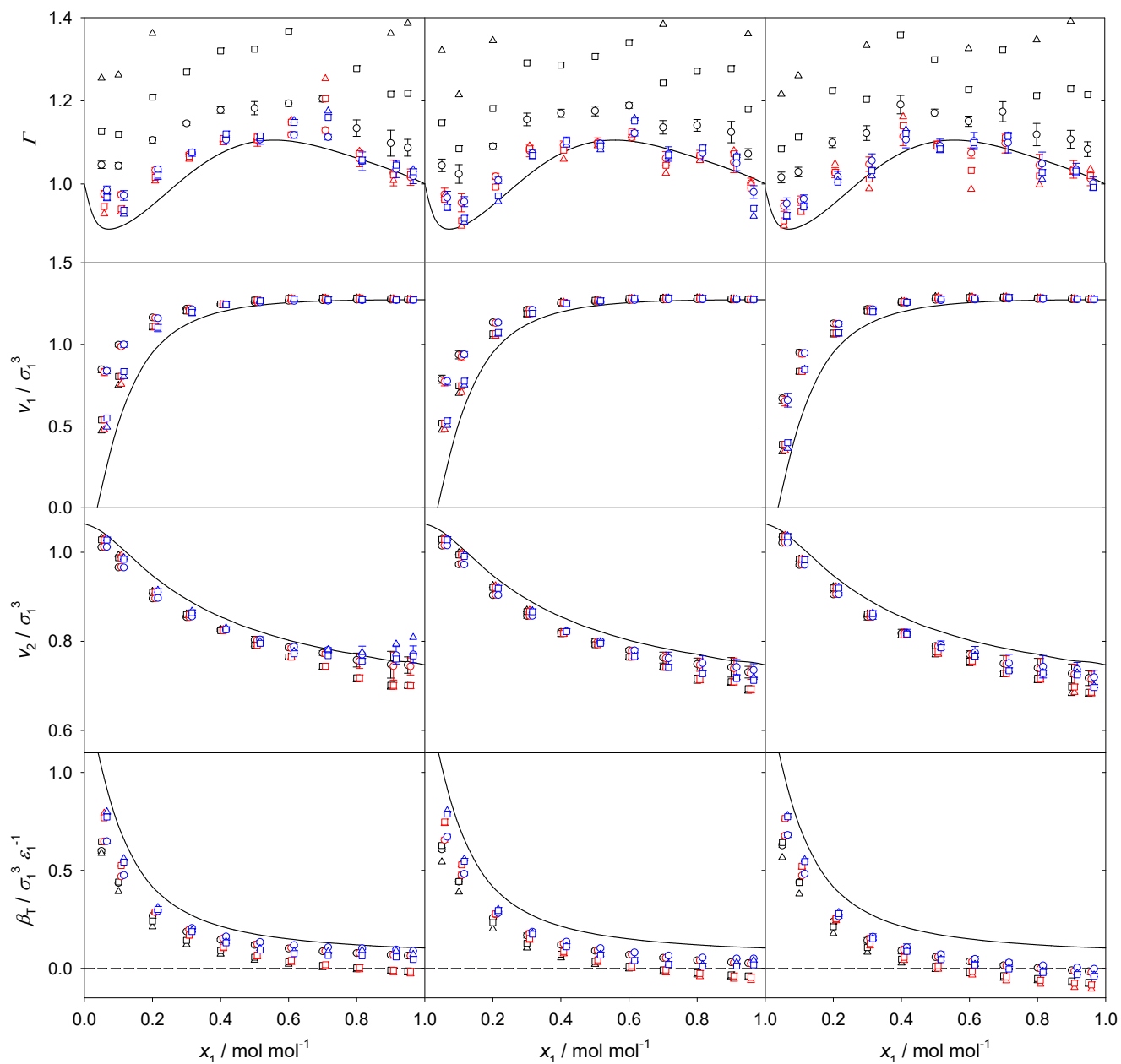


Figure S.9. Thermodynamic factor  $\Gamma$  (top), partial molar volumes  $v_1$ ,  $v_2$  (center) and isothermal compressibility  $\beta_T$  (bottom) over mole fraction  $x_1$  of mixture LJ4; black: standard RDF; red: vdV corrected RDF; blue: vdV+shf corrected RDF; circles: data based on  $G_{ij}$ ; triangles (eq. (7)) and squares (eq. (8)): data based on  $G_{ij}^\infty$  having the same statistical uncertainty as the circles; left:  $N = 4000$  molecules; center:  $N = 8000$  molecules; right:  $N = 16000$ ; data points at given compositions are slightly shifted in horizontal direction for visibility reasons; solid lines: fourth order polynomial fitted to chemical potential  $\mu_i$  data sampled with Widom's test particle insertion [18] during  $NpT$  simulations (top); results from  $NpT$  simulations (below); for outliers cf. Excel table.

**Table S.1. Mean statistical uncertainties of isothermal compressibility  $\beta_T$  data from NpT simulations over the entire composition range.**

mixture	$\Delta\overline{\beta_T} / \%$
LJ1	0.54
LJ2	0.52
LJ3	0.51
LJ4	0.49
methanol + ethanol	3.00
acetone + methanol	2.35
methanol + toluene	2.32
cyclohexane + methanol	3.02