Fluctuation Solution Theory Properties from Molecular Simulation

Abildskov, Jens; Wedberg, R.; O’Connell, John P.

Published in:
Fluctuation Theory of Solutions

Publication date:
2013

Citation (APA):
6. Fluctuation Solution Theory Properties from Molecular Simulation

Jens Abildskov,¹ R. Wedberg,² and John P. O’Connell³

¹CAPEC, Department of Chemical and Biochemical Engineering Building 229, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
³Department of Chemical Engineering, University of Virginia, Charlottesville, VA 22904-4741 USA
Abstract

The thermodynamic properties obtained in the Fluctuation Solution Theory are based on spatial integrals of molecular TCFs between component pairs in the mixture. Molecular simulation, via either MD or MC calculations, can yield these correlation functions for model inter- and intramolecular potential functions. However, system-size limitations and statistical noise cause uncertainties in the functions at long range, and thus uncertainties or errors in the integrals. A number of methods such as truncation, distance shifting, long-range modeling, transforms, DCF matching, finite-size scaling and adaptive resolution, have been explored to overcome these problems. This chapter reviews the issues and published work associated with using molecular simulation to obtain FST properties. The results suggest that molecular simulation should now be more fully utilized for obtaining quantitative FST thermodynamic properties of solutions.
6.1 Introduction

Although FST integrals may be obtained from simulations, most modeling has been done with empirical expressions for the integrals which appear in the final exact expressions. These are described in Chapter 9 for both pure and mixed systems. However, since the RDFs for the three molecular pairs of a binary mixture can be directly obtained from MD simulations, in principle these may be integrated numerically to yield TCFIs. With force fields capable of representing real behavior, real FST property variations can be computed. However, this task has proven to be more difficult than might have been expected. We are not aware of a fully reliable method for obtaining TCFIs from RDFs for polyatomic molecules. However, recent progress in calculating RDFs, as described in this chapter, should eventually lead to direct applications for real systems of interest, significantly expanding the knowledge and application of fluctuation solution theory. The focus here is on smaller molecules; simulations of proteins and larger substances have been described in Chapter 5.

6.2 Basics

Unlike the atomic TCFs and pair RDFs introduced in Chapter 1, which are functions of only the spatial distance \( r \) between the centers of mass of the two molecules, molecular correlation functions depend on orientations, \( \omega_1 \) and \( \omega_2 \). The TCF can be resolved into isotropic and anisotropic parts (Gray and Gubbins 1984),

\[
h_{ij}(r_{12}, \omega_1, \omega_2) = h_{ij}^{(a)}(r_{12}, \omega_1, \omega_2)
\]

where the isotropic part \( h_{ij}(r_{12}) = g_{ij}(r_{12}) - 1 \) is obtained by averaging over angles,
\[ h_{ij}(r_{ij}) \equiv \left\langle h_{ij}(r_{ij}, \omega_1, \omega_2) \right\rangle_{\omega_1, \omega_2} \]

with,

\[
\left\langle \right\rangle_{\omega_1} \equiv \frac{1}{8\pi^2} \int d\omega_1 \equiv \frac{1}{8\pi^2} \int_0^{2\pi} d\phi \int_0^\pi d(\cos \theta_1) \int_0^{2\pi} d\chi_1
\]

Thus, the anisotropic part is constructed to vanish upon averaging over orientations,

\[
\left\langle h^{(a)}_{ij}(r_{ij}, \omega_1, \omega_2) \right\rangle_{\omega_1, \omega_2} = 0
\]

For flexible molecules, the correlation functions are also functions of the molecular conformation. We consider small substances those for which conformation effects can be ignored. The molecular OZ equation defines the molecular DCF, \( c_{ij}(r_{ij}, \omega_1, \omega_2) \) (Gray and Gubbins 1984),

\[
h_{ij}(r_{ij}, \omega_1, \omega_2) = c_{ij}(r_{ij}, \omega_1, \omega_2) + \rho \sum \xi_i \left\langle h_{ij}(r_{ij}, \omega_1, \omega_2) c_{ij}(r_{ij}, \omega_1, \omega_2) \right\rangle_{\omega_1, \omega_2} d\vec{r}_3
\]

where \( \rho \) denotes the overall number density of the fluid and \( \xi_i \) is the number fraction of component \( i \). In analogy with Equations 6.1, 6.2, and 6.4 the DCF can be written as a sum of isotropic and anisotropic parts,

\[
c_{ij}(r_{ij}, \omega_1, \omega_2) = c_{ij}(r_{ij}) + c^{(a)}_{ij}(r_{ij}, \omega_1, \omega_2)
\]

Substituting Equations 6.1 and 6.6 into Equation 6.5 and angle-averaging leads to,

\[
h_{ij}(r) = c_{ij}(r) + \rho \sum \xi_i \int h_{ij}(r_{ij}) c_{ij}(r_{ij}) d\vec{r}_3 +
\]

\[
\rho \sum \xi_i \left\langle \left\langle h^{(a)}_{ij}(r_{ij}, \omega_1, \omega_2) \right\rangle_{\omega_1, \omega_2} \left\langle c^{(a)}_{ij}(r_{ij}, \omega_1, \omega_2) \right\rangle_{\omega_1, \omega_2} \right\rangle_{\omega_1, \omega_2} d\vec{r}_3
\]

Neglecting the last term in Equation 6.7 gives a simplified version of the OZ equation, in which the isotropic DCFs and TCFs are related without the anisotropic terms,
\[ h_y(r) = c_y(r) + p \sum_i x_i \int h_i(|r - r'|)c_y(r') dr' \]

6.8

The various approaches to spatial integration of molecular simulation data described in section 6.3 use this equation instead of the full OZ equation. While this may not be fully rigorous, it is supported by several different analyses. First, Equation 6.8 is exact in some integral equation theories of fluids with anisotropic interactions, such as the mean-spherical approximation and the generalized mean field theory (Gray and Gubbins 1984). Second, Wang et al. showed from MC simulations of LJ particles with significant dipole and quadrupole moments, that anisotropic forces have limited effects on \( h_y(r) \) (Wang et al. 1973). Also, Gubbins and O’Connell showed that, for dense fluids, compressibility data for water and argon could be scaled with only two parameters, meaning that the anisotropic effects were not apparent in the water data (Gubbins and O’Connell 1974). In addition, several studies show successful corresponding-states scaling for the DCFIs (Brelvi and O’Connell 1972, 1975a, 1975b; Campanella, Mathias, and O’Connell 1987; Huang and O’Connell 1987; Abildskov, Ellegaard, and O’Connell 2009, 2010a, 2010b), as described in detail by O’Connell (O’Connell 1994). Finally, the approximation of Equation 6.8 is the first term of the spherical harmonic expansions of the molecular correlation functions. It is worth noting that Equation 6.8 can be systematically improved by considering the spherical harmonic expansions of the orientation-dependent TCFs and DCFs (Gray and Gubbins 1984). The significance of this improvement is currently unknown, and probably depends on the system.

6.2.1 Equivalence of Ensembles

FST is based on the \( \mu VT \) ensemble, so the KBIs are integrals over RDFs for an open system. However, simulations are most conveniently performed in the \( NpT, NVT, \) or \( NVE \) ensembles.
MD simulations in the $\mu VT$ ensemble are possible (Cagin and Pettitt 1991), but nontrivial due to the problems associated with inserting new particles (Beutler et al. 1994). For this and other reasons, simulations are normally done on closed systems, though rigorously, the corresponding KBIs are equal to 0 for unlike pairs and -1 for like pairs, $\mu VT$ and $NpT$ RDFs differ by a term of the order of $1/N$, and the principle that $g_{ij}(r\to\infty) = 1$ is violated for closed systems (Ben-Naim 1990a). Fortunately, as illustrated in previous computational studies, RDFs in open and closed systems are extremely similar (Weerasinghe and Pettitt 1994). This means that while the original KB theory cannot be rigorously applied to a closed system, calculations converge to correct results with increasing $N$. Thus, it has become standard to use the equivalence of ensembles and to determine TCFIs using the MD simulations in the $(NpT/NVT)$ ensembles rather than the $\mu VT$ ensemble.

6.2.2 Integration

The usual approach to determine the function $g_{ab}(r)$ between the centers of mass of particle 1 of species $a$ and particle 2 of species $b$ separated by the distance $r$, is the accumulation of the number of particles $b$ lying in the interval $[r, r + dr]$ from a given particle $a$, and for all available values of $r$ within the central box. Numerical integration of the RDFs from molecular simulation is less straightforward. Theoretically, $h_{ij}(r)$ goes to zero when $r$ goes to infinity. However, because the integral is evaluated numerically, convergence requires that $h_{ij}(r)$ goes faster to zero than $r^2$ goes to infinity. For practicality, since the upper limit of the integral is infinite, for a convergent integral, there must be some upper separation distance, $R_{lim}$, beyond which the integrand no longer contributes significantly to the value of the integral. Thus, one defines
\[ H_j(R_{\text{lim}}) = \rho \int_0^{R_{\text{lim}}} r^2 h_j(r) \, dr \] 6.9

with \( R_{\text{lim}} \) is chosen sufficiently large that the integral is converged. This means that \( H_j(R_{\text{lim}}) \) should be insensitive to larger values of \( R_{\text{lim}} \). Experience shows, however, that \( H_j(R_{\text{lim}}) \) frequently does not include all of the nonzero values of \( r^2 h_j(r) \) from an MD simulation (Salacuse, Denton, and Egelstaff 1996). An example of this is shown in Figure 6.1 (Wedberg 2011).

The lack of convergence is mainly due to the RDFs retaining subtle structure over relatively long distances that contribute to the integral. While using a larger system might minimize the effects of omitting such contributions, the result would be significantly increased computational time without necessarily ensuring accuracy or reliable extrapolation. There are techniques (Theodorou and Suter 1985; Nichols, Moore, and Wheeler 2009) to allow calculation of pair correlation functions for distances up to \( \sqrt{3}/2 \) times the box dimension. However, convergence may not be obtained with the RDFs exhibiting substantial noise as the upper limit is approached (Salacuse, Denton, and Egelstaff 1996).

6.2.2 Asymptotic Properties of RDFs and Potential Truncations

One clever approach to obtaining better convergence is to include asymptotic properties of the pair correlation functions (Lebowitz and Percus 1963). In particular, exact asymptotic expressions have been obtained by Attard and co-workers (Attard 1990; Attard et al. 1991) such as for dipolar fluids. Other work has extended simulation results for a system with a truncated potential to give those for the full potential (Lado 1964). The effects on pair distribution
functions of potential truncations are important, for example, when the long-range tail of a potential must be truncated at some finite distance and the effect of the neglected part of the potential must be determined.

6.3 Methods

The task of extending the pair distribution function based on theoretical considerations has been addressed many times (Verlet 1968; Galam and Hansen 1976; Jolly, Freasier, and Bearman 1976; Ceperley and Chester 1977; Dixon and Hutchinson 1977; Foiles, Ashcroft, and Reatto 1984). Often the goal has been to study the correlation functions themselves or to calculate structure factors, not to obtain properties. Here we will emphasize applications aimed toward representing thermodynamic properties of molecular fluids that do not have conformational variations. While many publications have been confined to atomic model fluids, such as LJ particles, we focus here on applications for real molecular systems and their mixtures.

6.3.1 Direct correlation function matching

The method of Verlet (Verlet 1968) is intended to correct correlation functions from simulation for the effects of finite-sized systems, such as those summarized by Salacuse et al., and to extend computed correlation functions to long range (Salacuse, Denton, and Egelstaff 1996). The method was originally used for a pure LJ fluid in a study of the DCF and the structure factor. Later it was extended to a LJ mixture (Jolly, Freasier, and Bearman 1976). With the method, pair TCFs are extended by forcing the corresponding DCF at large separations to be consistent with the result,
Here $h_{\text{MD}}(r)$ is given by the simulation result and $a$ is a closure relation giving the DCF in terms of the TCF, $h(r)$, and the potential, $u(r)$, evaluated at the same $r$. Verlet’s method utilizes the simple structure and short range of $c$ compared to $h$. The approach can use either the Mayer $f$-function or the Percus–Yevick relation to extend pair distribution functions obtained from simulations of the pure LJ fluid. For the LJ fluid, three different relations $a(u,h)$ have been commonly used: the Mayer function,

$$a(u(r), h(r)) = f(r) = \exp(-\beta u) - 1 \quad 6.11$$

the PY relation,

$$a(u(r), h(r)) = f(r)(1 + h)(1 - \exp(-\beta u)) \quad 6.12$$

and the first-order virial expansion of $c$ (Wedberg et al. 2010),

$$a(u(r), h(r)) = f(r)\left[1 + \rho \int f(|r - r'|) f(r')dr'\right] \quad 6.13$$

Note that these theories for $c$ are all consistent with the asymptotic result $c(r) \rightarrow u(r)/kT$, when $r \rightarrow \infty$ (Lebowitz and Percus 1963). As shown in Foiles, et al., Equations 6.11-6.13 usually yield similar results and there is no rigorous basis for selecting one over another (Foiles, Ashcroft, and Reatto 1984). Calculations of isothermal compressibilities by integration of the extended pair correlation functions were reported only for three state conditions by Verlet (Verlet 1968), and the statistical uncertainties were large. It is not clear whether these uncertainties were due to the quality of the simulations and their analysis or to the assumptions made by the extension method. Furthermore, while the compressibilities were fairly reasonable for noble gases, agreement with experiment was not reported in quantitative terms, due to the focus of the paper being on other properties of the correlation functions.
For some time, the numerical solution of the liquid structure integral equations remained a challenge. Progress along these lines was made by the numerical implementation of Verlet’s method, based on the Newton-Raphson method, known as the Gillan scheme for solving the OZ equation (Gillan 1979; Abernethy and Gillan 1980; Enciso 1985). Also a factorization approach to the OZ equation was demonstrated by Jolly et al. (Jolly, Freasier, and Bearman 1976). Before describing applications of our variant of the Verlet method to systems resembling real molecular mixtures in Section 6.4, we review other approaches designed to improve convergence.

6.3.2 Truncation

One of the simplest strategies is called truncation (Weerasinghe and Smith 2003b). With this method, the $ \mu VT $ RDF is approximated by the $ NpT $ RDF truncated at a specific distance chosen for $ R_{\text{lim}} $. This distance is chosen to be, “The range over which the intermolecular forces dominate the distribution of particles.” Ideally, the truncated $ NpT $ RDF captures the major features of the $ \mu VT $ RDF, and its integral provides a good approximation to the desired KBIs. Truncated RDFs obtained from $ NpT $ simulations have been obtained for several different mixtures, and used to obtain properties, such as partial molar volumes (Lin and Wood 1996), and to express the KB equation in terms of local compositions (Mansoori and Ely 1985).

As described in Chapter 5, Smith and co-workers have employed the truncation method in order to develop accurate force fields for solutions, especially those with biochemicals and proteins. Over the past decade, a series of force-field development and validation studies have been published using KBIs (Chitra and Smith 2001a; Weerasinghe and Smith 2003d, 2004; Gee et al.}
The truncation approach can be successful if the TCFIs converge within the range of distance sampled by simulation. If not, the results will depend sensitively on the choice of $R_{\text{lim}}$. It is therefore common to average $H(R_{\text{lim}})$ with $R_{\text{lim}}$ varying over a selected interval. There seem to be no general rules for selecting the interval, other than suggesting that it should cover one oscillation of the TCF.

### 6.3.3 Distance-Shifting

Distance-shifting is another method employed with data on systems resembling real molecules. Both Perera and Sokolić (Perera and Sokolic 2004) and Hess and van der Vegt (Hess and van der Vegt 2009) attempt to correct the RDFs obtained from simulation by rescaling according to

$$g_{ij}(r) = \alpha_{ij}(r)g_{ij}^{(0)}(r)$$

Here $\alpha_{ij}$ is chosen in order to enforce that $g_{ij}(r)$ approaches unity at long distances. Perera and Sokolić (Perera and Sokolic 2004) presented $NpT$ simulations of the water + acetone binary mixture. Although a correction of the order $1/N$ is required, the result $g_{ab}(r) \rightarrow 1$ (for $r \rightarrow \infty$) is often valid for simple fluids after a few molecular diameters. This can be realistic in simulation boxes with a few hundred particles. However, for systems characterized by microscopic aggregation, the RDFs decay in irregular fashions with the range of correlations in the RDF differing from that of pair interactions, even at conditions remote from a critical point. The apparent problem in evaluating these quantities is the upper bound of the integral relative to the range of the correlations described by the simulations. If the system is large enough, one may consider that the correct asymptotic behavior is attained at some cutoff $R_{\text{lim}}$ smaller than the half-box length $L_{\text{box}}/2$. Accordingly, the KBI can be computed by replacing the infinite upper bound by $R_{\text{lim}}$, as in truncation methods. Though this may be satisfactory for simple fluids, it is
probably incorrect for fluids with long-range correlations. If the upper bound is not large enough to capture the correct asymptotic behavior, it will lead to incorrect estimation of the KBIs. For water/solute systems, these correlations probably extend over five to six water diameters, which is too large even for a system with $N = 1024$.

Since the $L_{\text{box}}/2$ values of all partial $g_{ij}(L_{\text{box}}/2)$ are always close to unity, Perera and Sokolić (Perera and Sokolic 2004) restored the correct asymptotic value at the natural half-box cutoff by shifting the value to unity. The expression used is,

$$g_{ij}(r) = \alpha_{ij}(r)g_{ij}^{(0)}(r), \quad \alpha_{ij}(r) = \frac{1}{1 + r/L_{\text{box}} - 1}$$

where $g_{ij}^{(0)}(r)$ is the uncorrected RDF. This procedure leaves values of the RDF at contact nearly unchanged if $g_{ij}^{(0)}(L_{\text{box}}/2)$ is close to unity. Perera and Sokolić (Perera and Sokolic 2004) found that $N = 864$ is just enough to satisfy this condition, though $N = 2048$ is much better. The merit of this equation is the use of all $r < L_{\text{box}}/2$ values in the calculation of the KBI, and, in particular, avoiding artifacts in the evaluations of the canonical ensemble KBIs.

Hess and van der Vegt (Hess and van der Vegt 2009) studied cation binding affinity with carboxylate ions. They computed the excess coordination numbers, $N_{ij}$, defined in Section 1.1.5, for water ($w$) or cations ($c$) about cations,

$$N_{jc} = \rho c 4\pi \int_0^\infty (g_{jc}(r) - 1)r^2 dr$$

where $j = w, c$ and computed the chemical potential derivatives of Equation 1.49 from closed system $NpT$ simulations. The finite-size correction is accounted for in a novel way. The idea is to consider a small part of a large system, such that this small part can be considered as open, and
then evaluate the integral up to a finite distance where it has converged within the larger system. Since the RDFs can not converge exactly to unity, the Hess/van der Vegt approach employs a scaling factor to account for the fact that the fluid composition far from a given molecule is different from the overall composition. The scaling factor is chosen, “Such that the RDF becomes exactly 1.” Assuming that $N_{jc}(R)$ is constant beyond a distance $R_{lim}$, the RDF can be normalized to 1 by dividing it by the ‘observed’ number of particles and multiplying by the ‘expected’ number,

$$g_{jc} = \frac{N_j (1 - V(R_{lim})/V_{box})}{N_j (1 - V(R_{lim})/V_{box}) - \Delta N_{jc}(R_{lim}) - \delta_{jc}} g_{jc}$$  \hspace{1cm} (6.17)

where $N_j$ is the number of particles of species $j$, $V(R)$ is the volume of a sphere with radius $R$ and $V_{box}$ is the volume of the simulation box, $\delta_{cc} = 1$ and $\delta_{wc} = 0$. For a system with 100 ion pairs, the scaling factor is around 1.005 (or one particle in 200). The correction for $g_{wc}$ is two orders of magnitude smaller. Although the methods are straightforward to implement, they still require selecting an appropriate truncation distance, and there seems to be no systematic way to choose the value of $R_{lim}$. This aspect can be a limitation, since in our experience the results are very sensitive to the selection of this value (Wedberg 2011).

6.3.4 Long Range Modeling

To avoid searching for an $R_{lim}$ that gives the value of a converged, KBI, an alternative is to find a model, and its parameters, for the long-range tail of the RDF to effectively extend the simulation results to infinite separation. The intention is to minimize sensitivity to the location where the simulation results are considered unreliable.
Matteoli and Mansoori (Matteoli and Mansoori 1995) gave a parametric expression for the RDFs of LJ fluids and their mixtures. That work arrived at a final form of the RDF based on the asymptotic conditions for zero density and infinite distance, as required by statistical thermodynamics, rather than rigorous geometrical and spatial considerations. Seven adjustable parameters were fitted to literature data on RDFs for each LJ fluid at different temperatures and densities. These were in turn expressed as functions of reduced temperature and density, so the complete parameterization used a total of 21 parameters. The capability of the expression to fit to RDFs of mixtures was checked against literature simulations of binary LJ mixtures with different diameters, molar fractions and $e_{AA}/e_{BB}$ ratios. The agreement between calculated and simulation curves was satisfactory. The values of the reduced pressure and internal energy calculated by numerical integration of the completely parameterized equation compared reasonably well with literature MD simulations. This approach allows calculation by integration of related quantities such as compressibility, internal energy, pressure and, using FST, the chemical potentials and partial molar volumes of the LJ mixture components for which RDF results are available. The works of Christensen et al. (Christensen et al. 2007a; Christensen et al. 2007b; Christensen et al. 2007c) addressed molecular simulations of systems resembling real molecules.

As shown in Figure 6.2 (a) for the benzene(1)/ethanol(2) binary mixture, the solid line is $g_{22}(r)$ which deviates much less than 1% from unity near $r_{\text{max}}$. However, divergence is clearly seen in Figure 6.2b, starting at $r \approx 15$ Å, where a local maximum in the integrand, $dG_{22}$, has a negative value. This is magnified in the TCFI because $r^2$ is a factor in the integrand, $dG_{22}$. To overcome this issue, a data reduction procedure based on tail modeling, in the sense of Matteoli/Mansoori
(Matteoli and Mansoori 1995) was tried. Unfortunately, the assumptions underlying the Matteoli/Mansoori RDF are not valid in the direct correlation range (first peak) for the kinds of substances frequently encountered in chemical engineering applications, such as hydrogen bonding species. For example, the Matteoli correlation assumes that the variations of the RDF from unity decrease from peak to peak so a combination of the exponential and cosine functions can be used. However, this is not always the case with real molecules. For example, the ethanol-ethanol RDF at high concentrations of benzene shows an irregular multi-peak behavior not seen in LJ mixtures. This reflects dilute solution association effects of the ethanol molecules. As a result, Christensen et al. abandoned the LJ expression and integrated the simulation results numerically for the range of direct interactions. Then, since the indirect part of the RDF does not change dramatically when different interaction potentials are used, especially when averaging over angles, the long-range shape of $g_{ij}(r)$ is simple and similar to that of hard spheres. Further, when a model expression is used, it is possible to analytically integrate the indirect $g_{ij}$ out to $r_{max}^*$ and ultimately to $r_{nc}$, beyond which there is no contribution to the TCFI. The model expression selected was,

$$g_{ij}^{\text{indirect}}(r) = 1 + a \cdot \exp[-h(r-e)] \sin \left( d(r-e) \right)$$

6.18

The five-parameters are determined by regression with the objective to reproduce $g_{ij}(r)$ from the third unity, $r_{u3}$, to $r_{max}$, with initial guesses found from,

$$a_{init} = g_{ij}(r_{max,2}) - 1,$$
$$b_{init} = \frac{-1}{r_{min,2} - r_{max,2}} \ln \left( \frac{1 - g_{ij}(r_{min,2})}{g_{ij}(r_{max,2}) - 1} \right),$$
$$c_{init} = r_{max,2},$$
$$d_{init} = \frac{\pi}{r_{u4} - r_{u3}},$$
$$e_{init} = r_{u3}$$

6.19

Here $r_{u3}$ and $r_{u4}$ are the radii of the third and fourth zeros of $h_{ij}(r)$, respectively, while $r_{max,2}$ is the
radius of the second maximum of $g_{ij}(r)$, or the first maximum of the indirect interaction, located between the third and fourth zeros of $h_{ij}(r)$. The quantities $g_{ij}(r_{\text{max,2}})$ and $g_{ij}(r_{\text{min,2}})$ are the values of the second maximum and minimum, respectively. The dashed curve in Figure 6.2 is the result of such a regression. The contribution to $H_{ij}$ from integration from $r_{\text{max}}$ to $r_{nc}$ is viewed as a long distance correction term.

[Insert Figure 6.2]

Thus, the final form for TCFIs has three contributions as shown in Figure 6.3: 1) the direct interaction part of the RDF, $H_{ij}^{\text{direct}}$, integrated numerically; 2) the integral of a trial function from the end of the direct interaction to the maximum distance determined by the box size, $H_{ij}^{\text{box}}$; and 3) the long distance contribution, $H_{ij}^{\text{ld}}$,

$$H_{ij} = \rho \int_{r_{\text{min}}}^{r_{\text{max}}} r^2 (g_{ij}(r) - 1) dr + \rho \int_{r_{\text{min}}}^{r_{\text{max}}} r^2 (g_{ij}(r) - 1) dr = H_{ij}^{\text{direct}} + H_{ij}^{\text{indirect}} + H_{ij}^{\text{ld}}$$

The statistical uncertainty of the direct interaction part of the RDFs is normally negligible. Uncertainties are small for the first portion of the indirect part, though they could be significant at greater distances. Finally, the contribution of $H_{ij}^{\text{ld}}$ to $H_{ij}$ is the least, so its uncertainty can be ignored.

[Insert Figure 6.3]

The 2006 International Fluid Properties Simulation Challenge (IFPSC) (http://fluidproperties.org/) competitions were initiated in 2001 to stimulate and assess prediction methods for properties of industrially important fluids, by comparing methods, assessing the
state of the art in simulation, and enhancing alignment of academic efforts with industrial needs. The 3rd IFPSC was held from March to September 2006. The focus of this contest was on the transferability of force fields and simulation methods for bubble pressures of mixtures of 1,1,1,2,3,3,3-heptafluoropropane (HFC-227ea refrigerant) and ethanol, based on limited data. In addition to their interesting pure component chemical properties, HFCs are often mixed with other fluids to be replacements for environmentally damaging chlorofluorocarbon refrigerants, and for cleaning-solutions, fire-retardants, and propellants. In the system of interest, hydrogen bonding interactions could occur between the HFC-227ea proton and the ethanol hydroxyl, leading to attractive unlike interactions. It was expected that this system would not be well-modeled by simple EoS mixing rules.

Entrants were provided with the experimental bubble points for 15 mixture compositions of 1,1,1,2,3,3,3-heptafluoropropane (HFC-227ea) and ethanol at 283.17 K, and properties of the pure materials at 343.13 K. The challenge was to compute bubble points for seven mixture compositions at 343.13 K, using any experimental data for the pure components but the only mixture points were those at 283.17 K. Entries using any theory/modeling/simulation method were accepted. Entries were judged based on the criterion,

$$\text{SCORE} = \frac{100}{7} \sum_{i=1}^{7} \left| \frac{P_{i,exp} - P_{i,calc}}{P_{i,calc}} \right|$$

The experimental data for the mixtures at 343.13 K, measured at DuPont, were not released until all entries had been received. The four contestants used a range of different techniques including statistical-mechanical and molecular-simulation approaches that gave significantly more accurate predictions than from the semi-empirical NRTL model (Prausnitz, Lichtenthaler, and Gomes de Azevedo 1999). The entry provided by Christensen, et al. (Christensen et al. 2007a) was based
on an NPT-MD simulation of the liquid phase at each mixture composition with FST linking the predicted microscopic structure (from the calculated RDF), to parameters that optimized a $G^E$-model for the liquid phase. It was assumed that the vapor phase was an ideal gas. For this contest the CHARMM force field was modified by revising the LJ parameters for the –CHF- part of the HFC-227ea molecule to fit experimental densities. The long-range modeling method of Christensen et al. (Christensen et al. 2007a) was sufficiently accurate to win (Case et al. 2007) in the ‘State Conditions Transferability’ category. Their predictions of VLE behavior shown in Figure 6.4 gave a SCORE of 1.52. Activity coefficients found for this system are unusual in behavior, as shown in Figure 6.5. The total pressures show positive deviations from Raoult’s law, but the activity coefficients have a maximum for HFC-227ae and a minimum for ethanol near $x_1 = 0.45$.

[Insert Figures 6.4 and 6.5]

The approach of Christensen was later modified by Wedberg et al. (Wedberg, Peters, and Abildskov 2008). First, the 5-parameter form of Equation 6.18 was changed to a 4-parameter form,

$$g^{\text{indirect}}(r) = 1 + a \cdot \exp\left[\frac{1}{b(r-c)}\right] \sin\left(d(r-c)\right)$$  \hspace{1cm} 6.22

Next, a tail model corresponding to the anti-derivative of the trial expression for $g(r)$ was fitted to the truncated numerical integral of $g(r)$ as a function of the upper integration limit. This tail model was then used to extrapolate $H(R_{\text{lim}})$ to $R_{\text{lim}} = \infty$ which yielded the value of the TCFI. Defining the running integrals of the RDF, $G(r)$,

$$G(r) = 4\pi \int_0^r r'^2 \left[ g(r') - 1 \right] dr', \quad H = \lim_{r \to \infty} r \rho G$$  \hspace{1cm} 6.23
the anti-derivative of Equation 6.23 used is,

\[
G_{\text{smooth}}(r) = -4\pi a \left[ p(r) \cos(d(r-c)) + q(r) \sin(d(r-c)) \right] e^{-b(r-c)} + f
\]

\[
p(r) = \frac{d}{b^2 + d^2} r^2 + \frac{4bd}{(b^2 + d^2)^2} r - 2 \frac{d^3 - 3b^2 d}{(b^2 + d^2)^3}
\]

\[
q(r) = \frac{b}{b^2 + d^2} r^2 + \frac{2(b^2 - d^2)}{(b^2 + d^2)^2} r - 2 \frac{3d^2 b - b^3}{(b^2 + d^2)^3}
\]

6.24

The parameters for \(G_{\text{smooth}}(r)\) were fitted to the sampled \(G(r)\) for \(r\) ranging from \(r_{\text{min}}\) to \(r_{\text{max}}\), as defined above. The method was used to obtain isothermal compressibilities of five pure alkanes at three different state points and these were compared to the values derived from simulated overall density fluctuations. Results showed that the two approaches were fully consistent in values and uncertainties. Further, the computations converged in approximately the same simulation times. This suggests that computation of TCFIs is a route to isothermal compressibility, as accurate and fast as well-established benchmark techniques, with the advantage that it can be used in any ensemble (Puliti, Paolucci, and Sen 2011). Note that this approach has only been successfully tested on pure fluids. The main limitation of both methods (Christensen et al. 2007a; Christensen et al. 2007b; Christensen et al. 2007c; Wedberg, Peters, and Abildskov 2008) is that they apply only to systems where the TCF tails can be approximated by the model equations, which may not be true in general.

6.3.5 Transforms

Nichols et al. developed a method using finite-Fourier-series expansions of molecular concentration fluctuations in order to reduce systematic errors from the simulation boundary conditions (Nichols, Moore, and Wheeler 2009). The procedure was validated and compared to a truncation method for a non-ideal binary liquid mixture of LJ particles tuned to imitate the
system CF₄ and CH₄. A fluctuation expression is applied to a portion of the total volume within a closed simulation such as \textit{NVT}. Rather than the sampling volume being spherical and centered on a single moving molecule, the sampling volume is one or more rectangular-slab regions that are stationary with respect to the simulation cell. This leads to two alternative expressions,

\[
S_{ij}(q) = x_i \delta_{ij} + x_i x_j \rho \int_0^\infty \frac{\sin(qr)}{qr} [g_{ij}(r) - 1] 4\pi r^2 dr 
\]

and,

\[
S_{ij}(q) = \frac{1}{N} \langle \delta N_i(q,t) \cdot \delta N_j(-q,t) \rangle = N \langle \psi_i(q,t) \cdot \psi_j(-q,t) \rangle
\]

where \(\psi_i(q,t)\) is the Fourier mass coefficient at the time \(t\) of the component \(i\), and the wave vector \(q\) has components that are integer multiples of \(2\pi/L_{\text{box}}\). The TCFIs are found from the structure factors via \(\rho G_{ij} = S_{ij}(0) - 1\).

For an infinite system, the definitions of \(S_{ij}\) in Equations 6.25 and 6.26 are equivalent. However, for finite systems, truncation of the integrals can lead to errors in the integral of Equation 6.25. The advantage of Equation 6.26 relative to Equation 6.25 is that the structure factor can be based on a discrete Cartesian-based Fourier transform, rather than a continuous, spherically symmetric Fourier transform, Equation 6.26, is evaluated at various values of \(q\) which are most accurate at larger values. The \(q\)-dependent \(S_{ij}\) are then extrapolated to \(q = 0\) by fitting them to polynomials, the range of \(q\) and the polynomial order being selected empirically. Thus, it is not \(g_{ij}(r)\) that is corrected; it is done via the structure factors related to the RDFs using the radial Fourier transform. The sampling volumes of the method do not truncate intermolecular correlations at a particular radial distance and no assumption is made that \(g_{ij} \rightarrow 1\) at large intermolecular
separations. In effect, this approach is consistent with periodic boundary conditions, but immune
to long-range truncation effects. The inaccuracies in simulation RDFs at large \( r \) become
unreliable \( S_q(0) \) because of the system size but extrapolation to \( q = 0 \) provides the correction.
Good results are obtained for LJ mixtures, but the method has apparently not yet been tested for
molecular fluids.

6.3.6 Finite-size scaling

The approach of Schnell et al. is to sample small non-periodic systems in a (periodic) simulation
box (reservoir) and then scale the results (Schnell et al. 2011). The simulation system has sides
of \( L_i \) in each dimension. Small systems are randomly selected sub-volumes, denoted by \( L_{n-1} \), \( L_n \),
and \( L_{n+1} \), and can exchange energy and particles with the rest of the system. While the reservoir
may not be a grand-canonical ensemble, the small systems will be when \( L_n \ll L_i \). Finite-size
scaling of KBIs determined from different-sized small systems is done by fitting and
extrapolating them to the thermodynamic limit using essentially straight lines. The values
compare well with those from integrating the RDF using a truncation method. In all cases, visual
inspection is needed to identify where the subvolume results follow straight lines, but this is
straightforward. While this approach has not been extensively tested, there is promise for such
computations for practical applications with relative insensitivity to system size.

6.3.7 Adaptive Resolution

Quite recently, Mukherji et al. (Mukherji et al. 2012) have proposed an adaptive resolution
simulation method (‘AdResS’). In a way resembling the distance-shifting method of Hess and
van der Vegt (Hess and van der Vegt 2009), the ‘AdResS’ method divides the simulation domain into a small region of high resolution (atomistic) molecules and a large region of (coarse-grained) lower resolution. Mukherji et al. analyze their method with a mixture resembling methanol and water by comparing the results with all-atom simulations and existing experimental results. Excellent agreement is found with the much larger and computationally more expensive all-atom simulations as well as with data. As with other methods that need to select an integration limit such that the integral converges to a plateau value or oscillates in a well-controlled way around a mean value, the ‘AdResS’ method also selects a high resolution region width. The results indicate that a radius of 2 nm is adequate for the specific system of methanol and water, but a general approach to selecting the radius is uncertain.

6.4 DCF Matching – Mixtures of Polyatomic Molecules

Recently we (Wedberg et al. 2010; Wedberg 2011; Wedberg et al. 2011a, 2011b) have investigated more thoroughly the extension method of Verlet (Verlet 1968) for mixtures with polyatomic molecules. This section gives a more complete description of our implementation, including some results not previously published. We have analyzed MD simulations of both pure (Wedberg et al. 2010) and mixed (Wedberg et al. 2011a, 2011b) LJ and Stockmayer fluids for wide ranges of state conditions and compared the results with the truncation method of Weerasinghe and Smith (Weerasinghe and Smith 2003b) and the distance-shifting method of Hess and van der Vegt (Hess and van der Vegt 2009).
6.4.1 Method

For potentials that decay faster than $r^{-3}$, the asymptotic behavior of the DCF is (Lebowitz and Percus 1963),

$$c(r) = -\beta u(r) + O\left(\left[\beta u(r)\right]^{2}\right)$$  \hspace{1cm} (6.27)

Here $u$ is the pair intermolecular potential. In order to extend the TCF obtained from simulation, one chooses a value $R$ within the range for which $h(r)$ is sampled, and determines $h$ and $c$ according to Equation 6.10. These requirements, together with the OZ equation, define a closed-form integral equation, which can be solved in order to obtain $h(r)$ for $r > R$. Simultaneously, $c(r)$ is obtained for all $r$. In our implementation, $h$ and $c$ are discretized as linear splines and both functions are assumed to be zero beyond a finite distance $R_c$. For example, in the cases of LJ and Stockmayer particles $R_c = 15\sigma$, where $\sigma$ is the LJ diameter. For “real” fluids, other values are used. Selecting this truncation radius is straightforward since the integrals generally converge, as can be checked by plotting the running integral versus $r$. If not converged, one repeats the calculation with a larger $R_c$. The Fourier-transformed OZ equation (Lebowitz and Percus 1963) is employed to express $c$ explicitly in terms of $h$. This transforms the integral equation into a system of nonlinear equations for which a numerical solution is found using Newton’s method (Wedberg et al. 2011a). The Jacobian is evaluated analytically and 5–15 iterations are normally required for convergence. Commonly, the Wiener-Hopf factorization technique is applied when the DCF is computed numerically from the TCF or vice versa (Jolly, Freasier, and Bearman 1976; Gray and Gubbins 1984; Press et al. 1992; Ramirez, Mareschal, and Borgis 2005). For the present application, the three-dimensional Fourier-transformed OZ equation can be employed. Applying the Fourier transform to Equation 6.6 gives a product,
\[ \hat{h}_y(k) = \hat{c}_y(k) + \rho \sum_{i=1}^3 \hat{h}_y(k)x_i \hat{c}_i(k) \]  

where \( \hat{h}_y(k) \) denotes the Fourier transformation of \( h_y(r) \). Due to radial symmetry, this is the zeroth-order Hankel transform, defined by,

\[ \hat{h}_y(k) = 4\pi \int_0^\infty drr^2 \sin(kr) \frac{h_y(r)}{kr} \]  

This same formulation holds for \( \hat{c}_y(k) \). The function \( h_y(r) \) is recovered from the inverse Hankel transform and is given by,

\[ h_y(r) = \frac{4\pi}{(2\pi)^2} \int_0^\infty dk k^2 \frac{\sin(kr)}{kr} \hat{h}_y(k) \]

Utilizing that \( \hat{h}_{12}(k) = \hat{h}_{21}(k) \) and \( \hat{c}_{12}(k) = \hat{c}_{21}(k) \), Equation 6.28 can be written as a linear system,

\[ \tilde{h}(k) = [I + \rho H(k)] \tilde{c}(k) \]

with,

\[
\begin{bmatrix}
\hat{h}_{11}(k) \\
\hat{h}_{12}(k) \\
\hat{h}_{22}(k)
\end{bmatrix} = \begin{bmatrix}
\hat{c}_{11}(k) \\
\hat{c}_{12}(k) \\
\hat{c}_{22}(k)
\end{bmatrix}, \quad \begin{bmatrix}
x_1\hat{h}_{11}(k) \\
x_2\hat{h}_{12}(k) \\
x_1\hat{h}_{21}(k) \\
x_2\hat{h}_{22}(k)
\end{bmatrix} = \begin{bmatrix}
x_1\hat{c}_{11}(k) \\
x_2\hat{c}_{12}(k) \\
x_1\hat{c}_{21}(k) \\
x_2\hat{c}_{22}(k)
\end{bmatrix}
\]

where \( I \) denotes the identity matrix. Equations 6.32 assume that the fluid mixture has at most two components. Equations 6.29-6.31 provide a route for computing \( c_y(r) \) given \( h_y(r) \). The function \( h_y(r) \) is Hankel-transformed to yield \( \hat{h}_y(k) \). The linear system in Equation 6.31 is then solved for \( \hat{c}_y(k) \) at each \( k \), followed by applying the inverse Hankel transform to obtain \( c_y(r) \). Solution of the problem of Equation 6.10 requires that the long-range part of \( h_y(r) \) is adjusted until the long-range part of \( c_y(r) \) matches a trial function \( t_y(r) \). This is accomplished by a Newton iteration scheme for which grids in \( r \) and \( k \) space are introduced,
\[ r_\alpha = \alpha \cdot \Delta r, \quad \alpha = 0, \ldots, N \]
\[ k_\beta = \beta \cdot \Delta k, \quad \beta = 0, \ldots, N \]

6.33

The upper cutoffs are \( R_c = N \cdot \Delta r \) and \( K_c = N \cdot \Delta k \) for the integrals in Equations 6.29 and 6.30, respectively. Note that \( R_c \) is not the sampling limit set by the simulation box dimensions, but is typically much larger. The TCFs, DCFs, and their Hankel transforms at an iteration step \( t \) are represented by discrete vectors,

\[ \tilde{h}_{ij}^{(t)} \equiv h_{ij}^{(t)}(r_{\alpha}), \quad \tilde{c}_{ij}^{(t)} \equiv c_{ij}^{(t)}(r_{\alpha}), \quad \alpha = 1, \ldots, N \]
\[ \tilde{h}_{ij}^{(t)} \equiv h_{ij}^{(t)}(k_{\beta}), \quad \tilde{c}_{ij}^{(t)} \equiv c_{ij}^{(t)}(k_{\beta}), \quad \beta = 1, \ldots, N \]

6.34

Equation 6.29 for the TCF is approximated by truncating the integral at \( R_c \) and using the trapezoidal rule,

\[ \tilde{h}_{ij}^{(t)} = T \cdot h_{ij}^{(t)} \]

6.35

Elements of the matrix \( T \) are,

\[ T_{\beta\alpha} = 4\pi \cdot \Delta r \cdot r_{\alpha}^2 \cdot \sin k_{\beta} r_{\alpha} \cdot \left( \frac{1}{k_{\beta} r_{\alpha}} \cdot \frac{\delta_{\alpha\alpha} + \delta_{\beta\beta}}{2} \right) \]

6.36

with \( \alpha = 1, \ldots, N, \beta = 1, \ldots, N \), with \( \sin(kr)/(kr) \) being unity if either \( k \) or \( r \) is zero. Equation 6.30 for the DCF is approximated in a similar way by truncating the integral at \( K_c \),

\[ \tilde{c}_{ij}^{(t)} = U \cdot c_{ij}^{(t)} \]

6.37

with,

\[ U_{\beta\alpha} = \frac{4\pi}{(2\pi)^3} \cdot \Delta k \cdot k_{\beta}^2 \cdot \sin k_{\beta} r_{\alpha} \cdot \left( \frac{1}{k_{\beta} r_{\alpha}} \cdot \frac{\delta_{\alpha\alpha} + \delta_{\beta\beta}}{2} \right) \]

6.38

with \( \alpha \) and \( \beta \) as in Equation 6.36. As stated above, the middle step of converting \( \tilde{h}_{ij}^{(t)} \) to \( \tilde{c}_{ij}^{(t)} \) is carried out by solving the linear system of Equation 6.31 for each value of \( \beta \). If \( n \) denotes
indexing such that \( r_{ij} \leq R_{ij} \leq r_{ij} + 1 \), and \( h_{ij}^{(t)} \) and \( c_{ij}^{(t)} \) denote vectors containing the elements of \( h_{ij}^{(t)} \) and \( c_{ij}^{(t)} \), respectively, with \( n_{ij} + 1 \leq \alpha \leq N \), \( h_{ij}^{(t)} \) is updated at each iteration step according to,

\[
h_{ij}^{(t+1)} = h_{ij}^{(t)} + \Delta h_{ij}^{(t)}
\]  

Here, \( \Delta h_{ij}^{(t)} \) in Newton’s method is found by solution of the linear system,

\[
\begin{bmatrix}
J_{11}^{11} & J_{12}^{11} & J_{12}^{12} & \Delta h_{11}^{(t)} \\
J_{11}^{12} & J_{12}^{12} & J_{12}^{22} & \Delta h_{12}^{(t)} \\
J_{11}^{22} & J_{12}^{22} & J_{12}^{22} & \Delta h_{22}^{(t)}
\end{bmatrix}
\begin{bmatrix}
\Delta c_{11}^{(t)} \\
\Delta c_{12}^{(t)} \\
\Delta c_{22}^{(t)}
\end{bmatrix}
= \begin{bmatrix}
\Delta J_{ij}^{(t)} \\
\Delta J_{ij}^{(t)} \\
\Delta J_{ij}^{(t)}
\end{bmatrix}
\]  

where the right-hand side represents the difference between the approximation of the long-range DCF to be enforced and the currently computed DCF,

\[
\Delta c_{ij}^{(t)} = t_{ij}(r_{ij}) - c_{ij}^{(t)}, \quad \alpha = n_{ij} + 1, \ldots, N
\]  

The Jacobian has the elements,

\[
J_{a \beta}^{ij} = \begin{bmatrix}
\frac{\partial c_{ij}^{(t)}}{\partial h_{a\beta}^{(t)}} & \ldots & \frac{\partial c_{ij}^{(t)}}{\partial h_{a\beta}^{(t)}} \\
\frac{\partial h_{ij}^{(t)}}{\partial h_{a\beta}^{(t)}} & \ldots & \frac{\partial h_{ij}^{(t)}}{\partial h_{a\beta}^{(t)}} \\
\frac{\partial c_{ij}^{(t)}}{\partial h_{ij}^{(t)}} & \ldots & \frac{\partial c_{ij}^{(t)}}{\partial h_{ij}^{(t)}}
\end{bmatrix}
\]  

These are partial derivatives that can be expanded by the chain rule to,

\[
\frac{\partial c_{ij}^{(t)}}{\partial h_{ab,\alpha'}} = \sum_{\beta=0}^{N} \sum_{a'=0}^{N} U_{a\beta} \frac{\partial c_{ij}^{(t)}}{\partial h_{ab,\beta'}} \frac{\partial h_{ab,\beta'}}{\partial h_{ab,\alpha'}}
\]  

The last equality is due to the result,

\[
\frac{\partial \tilde{c}_{ij}^{(t)}}{\partial h_{ab,\beta'}} = \delta_{\beta \beta'} \frac{\partial \tilde{c}_{ij}^{(t)}}{\partial h_{ab,\beta}}
\]  

a consequence of Equations 6.31, 6.39, and 6.41. The partial derivatives are obtained from three linear systems derived from Equation 6.31 by differentiation with respect to \( h_{11,\beta}, h_{12,\beta} \) and \( h_{22,\beta} \).
The results are,

\[
\begin{bmatrix}
\frac{\partial\tilde{c}_{11,0}}{\partial h_{11,0}} & \frac{\partial\tilde{c}_{11,0}}{\partial h_{12,0}} & \frac{\partial\tilde{c}_{11,0}}{\partial h_{22,0}} \\
\frac{\partial\tilde{c}_{12,0}}{\partial h_{11,0}} & \frac{\partial\tilde{c}_{12,0}}{\partial h_{12,0}} & \frac{\partial\tilde{c}_{12,0}}{\partial h_{22,0}} \\
\frac{\partial\tilde{c}_{22,0}}{\partial h_{11,0}} & \frac{\partial\tilde{c}_{22,0}}{\partial h_{12,0}} & \frac{\partial\tilde{c}_{22,0}}{\partial h_{22,0}}
\end{bmatrix}
= \left[ I + \rho \mathbf{H}(\beta \cdot \Delta k) \right]^{-1}
\begin{bmatrix}
1 - x_1 \rho \tilde{c}_{11,0} & -x_2 \rho \tilde{c}_{12,0} & 0 \\
-x_1 \rho \tilde{c}_{12,0} & 1 - x_2 \rho \tilde{c}_{22,0} & 0 \\
0 & -x_1 \rho \tilde{c}_{12,0} & 1 - x_2 \rho \tilde{c}_{22,0}
\end{bmatrix}
\]

At each iteration step, these systems are solved for the partial derivatives which then are used to evaluate the Jacobians in Equation 6.44.

The short-range parts of the calculated DCFs are not used within the iteration scheme though the short-range part of the DCF obtained from the final iteration is considered in selecting the parameters \(R_{ij}\). Initially, the discretized TCFs are set to \(h_{ij}(0) = h_{MD,ij}(r_a)\) for all \(r_a\) within the sampling range for \(h_{MD,ij}(r)\), and \(h_{ij}(0) = 0\) for larger \(r\). The iteration is carried out until,

\[
\sum_{i,j} \sum_{a=a_{ij}}^{N} \left[ \Delta c_{ij,a}^{(r)} r_a^2 \right] < \eta
\]

with \(\eta = 10^{-4}\) or less. Typically, this is achieved after 5-15 iterations. For some systems, in particular those at high density where the functions \(h_{ij}(r)\) have significant structure beyond the sampling range, the tail model by Christensen et al. (Christensen et al. 2007a; Christensen et al. 2007b; Christensen et al. 2007c) was used to estimate the long-range behavior for the initial guess \(h_{ij}^{(0)}\). Using this approach, the Newton iterations have converged for all systems we have studied to date. Issues of how to select the matching distance, and the angle-averaging of potentials are discussed in detail by Wedberg (Wedberg 2011).


6.4.2 Results

We now discuss results from MD simulations that test the capabilities of the method. The KBIs are primarily verified by comparing the derivative properties obtained from the integration procedure with the same properties obtained from alternative analyses, or from simulation results in the literature. For the simulations of water/organic solvent mixtures, the derivative properties obtained by integration are also compared against values derived from correlations of experimental data. In this last case, consistency depends not only on the accuracy of the integration procedure, but also on the accuracy of the force field.

6.4.3 Model Fluids

The methodology was first tested on pure and mixed LJ and Stockmayer fluids (Wedberg et al. 2010) for several reasons. First, these fluids are well-defined, so simulation results over wide temperature and density ranges could be acquired with limited computational effort. Second, the thermodynamic derivative properties obtained from the extended pair-distribution function can be validated against data derived from correlations of previous simulations. Third, the simple form of the inter-atomic potentials allows a basic test of the assumption that the OZ equation can be resolved into isotropic and anisotropic parts. For the Stockmayer fluid, this form of the OZ relation is inexact, becoming less accurate as the (reduced squared) dipole moment, $\mu^*\sigma^2$, increases. The accuracy of the properties obtained for large $\mu^*\sigma^2$ could indicate validity of the isotropic OZ equation. In what follows, the physical quantities are in dimensionless values, where the quantities marked with an asterisk (*) have been reduced with respect to the LJ parameters $\varepsilon$ (energy), $\sigma$ (length) and the atomic mass. Here we give here later developments (Wedberg 2011; Wedberg et al. 2011b) than described by Wedberg et al. (Wedberg et al. 2010),...
which has less reliable approximations for the DCF tail.

### 6.4.4 LJ Fluids

Figure 6.6 compares the isothermal compressibilities obtained from the method with those obtained from the EOS of Mecke et al. for LJ fluids (Mecke et al. 1996). At all four temperatures, our results are qualitatively consistent with the EOS, with differences in the range of 1-5%. The comparisons are not as good at the two lowest temperatures, where the differences are as high as 6%.

![Insert Figure 6.6]

The greatest disagreement is seen when $T^* = 1.5$ and $\rho^*$ is 0.3 or 0.4, which are the state points closest to the critical point of $\rho_{c}^* = 0.304$ and $T_{c}^* = 1.316$ (Smit 1992). At those conditions, the differences are 7% and 8%, respectively. This disagreement is not surprising considering that the reduced bulk modulus ($\rho k_B T \kappa_T = 1 - C$) can be very small in this region. We conclude that our method is best suited for systems at liquid density ($\rho > 2\rho_c$). For $\rho^* = 0.7$ and $\rho^* = 0.8$, the agreement with the EOS was better at higher temperature (1% at $T^* = 2.5$) than at lower temperature (6% at $T^* = 0.85$). It is possible that derivatives of the Mecke EOS are less accurate at lower temperatures since the EOS does not reproduce the simulation pressures very well under these conditions and at higher temperatures. Nevertheless, though low temperatures seem to offer more of a challenge, the results obtained under those conditions may still be considered satisfactory. The standard error in $T$ was less than 0.5 %, indicating that the calculations were well converged.
6.4.5 LJ/Stockmayer Mixtures

LJ/Stockmayer mixtures include “Stockmayer” atoms (2) with finite dipole moments and “LJ” atoms (1) with zero dipole moment. LJ-LJ and LJ-Stockmayer interactions thus follow the LJ potential, while Stockmayer-Stockmayer interactions include dipole-dipole interactions. The EOS of Gross and Vrabec (Gross and Vrabec 2006) describes mixtures of fluid particles with different dipole moments and thus can be used to obtain the isothermal compressibility. Values of the reduced bulk modulus obtained from our method are compared in Figure 6.7 with the Gross/Vrabec EOS. For $\mu^* = 1$ (a), the Verlet values agrees very well with the EOS; the differences are 1–1.5%. For the higher dipole moments, the agreement is still good when $x_2$ is small but deteriorates as $x_2$ increases (b) and(c). This becomes more pronounced for $\mu^* = 3$ where the discrepancies are as large as 11% at Stockmayer-rich compositions. Since the Gross/Vrabec EOS did not reproduce simulation pressures very well (Wedberg 2011) with errors increasing with larger $\mu^*$ and $x_2^*$, the discrepancy may not be only in the simulation results.

[Insert Figure 6.7]

6.4.6 Comparison with existing approaches

The truncation (Weerasinghe and Smith 2003b) and distance-shifting (Hess and van der Vegt 2009) methods were also employed to calculate $\rho k_B T \tau_r$. Simple truncation requires averaging $H(R_{\text{lim}})$ over a specific interval, where $H(R_{\text{lim}})$ is the numerical TCFI as a function of the upper integration limit $R_{\text{lim}}$. It is not obvious how to choose these values. Here, the TCFIs have been averaged with $R_{\text{lim}}$ in the interval $[2\sigma, 3\sigma]$, where $\sigma$ roughly corresponds to the oscillation period of $h_{ij}(r)$. With the Hess method, the scaling factor $\alpha_{ij}$ was evaluated with $R = 4\sigma$. The RDFs were
re-scaled, but the integrals still did not converge within the sampled range. The integration of the re-scaled RDFs was thus carried out as with the truncation method, but using larger truncation radii, based on the idea that corrected RDFs are more reliable at large separations. The TCFIs were averaged using truncation radii in the interval \([3.5\sigma, 4.5\sigma]\). The values obtained for the isothermal compressibility (Figure 6.7) demonstrates the limitations on simple truncation and distance-shifting methods. Truncation overestimates the compressibility by 10–15% while distance-shifting underestimates it by 10–40%. These are sensitive to the choice of truncation radii, so it is possible that better results could have been obtained with other radii. The methods should also perform better with significantly larger simulation systems, but this was not tried here. The Verlet method yielded accurate results for LJ/Stockmayer mixtures as indicated by comparisons with benchmark values. While the method achieves better accuracy than simpler integration approaches, caution is advised regarding activity coefficient derivatives when a system is nearly ideal or when the mole fraction of a component is less than approximately 15%, as discussed by Wedberg et al (Wedberg et al. 2011a).

### 6.4.7 Aqueous Alcohol Mixtures

The major goal is to establish an integration method that accurately predicts activity coefficient derivatives, partial molar volumes, and isothermal compressibilities from simulations of molecular mixtures with atom-atom interaction models. This section focuses on such applications with results compared to values derived from correlations of experimental data. It should be noted that the accuracy also depends on the validity of the molecular force fields and reliability of experimental data. As with the analysis of the LJ/Stockmayer mixtures, the simple
truncation (Weerasinghe and Smith 2003b) and distance-shifting methods (Hess and van der Vegt 2009) were employed to evaluate the same properties. Simple truncation averaging of the integral varied in the interval for $R_{\text{lim}}$ from 1.0 to 1.5 nm. With the distance-shifting method, the scaling factors $\alpha_{ij}$ were evaluated from the calculated RDFs with the parameter $R = 2.0$ nm.

Numerical integration of the re-scaled RDFs did not converge within the sampling range, so the integrals of the rescaled TCFs were evaluated by truncation using intervals of 1.4–1.9 nm. The truncation radii employed for integration of the re-scaled TCFs were larger than those used with the simple truncation approach since the rescaled TCFs probably were more accurate than the original TCFs for large $r$, as discussed in Section 6.3. For comparison, isothermal compressibilities were evaluated via the fluctuations of the simulation box volume. The results are shown in Figure 6.8. The Verlet method reproduced the fluctuation formula results to within 5%, while the simple truncation and distance-shifting methods were greatly in error. In fact, distance-shifting yielded negative compressibilities. It is likely that reliable results by these methods require simulations of larger systems.

In order to validate the partial molar volumes obtained by the different integration methods, the excess molecular volume, $V^E$, was evaluated for the simulations at each composition according to,

$$V^E_m(x_1) = V_m(x_1) - x_1V_1^\circ - x_2V_2^\circ$$

where $V_m(x_1)$ denotes the average molecular volume obtained at the composition $x_1$, the mole fraction of water. Also, $V_1^\circ$ and $V_2^\circ$ denote the average molar volumes of the corresponding pure
components, obtained from separate simulations. The polynomial model of Handa and Benson (Handa and Benson 1979),

\[ V_m^E(\text{model}) = x_1x_2[a_0 + a_1(x_2 - x_1) + a_2(x_2 - x_1)^2] \]

was fitted to the calculated values of the excess molar volume. Reduced partial molar volumes were evaluated by analytical differentiation of the model according to,

\[ \rho \bar{V}_1 = \rho V_m^E + \rho \left( \frac{\partial(NV_m^E)}{\partial N_1} \right)_{T,p,N_2} \]

In Figure 6.9, the results for \( \rho \bar{V}_1 \) are compared with TCFIs calculated by our Verlet, simple truncation, and distance-shifting methods. The partial molar volumes obtained from the correlations of simulation volumes are in very good agreement with those obtained from experimental correlations (Handa and Benson 1979). The results obtained via the three TCFI calculation methods agreed very well with both correlations. Furthermore, the three methods yielded similar results, though for dilute water systems, simple truncation underestimates the values relative to the other methods and experimental data.

[Insert Figure 6.9]

Figure 6.10 shows activity coefficient derivatives over the whole composition range for experiment from three correlations, and the Verlet method. A procedure for experimental data analysis was described by Wooley and O’Connell (Wooley and O’Connell 1991), in which one extracts the isothermal compressibility, partial molar volumes, and activity coefficient derivatives from experimental data. The activity coefficient derivatives are obtained by fitting mixture vapor-liquid equilibrium data to obtain parameters for at least two different \( G^E \) models. Wooley and O’Connell employed the Wilson, non-random two liquid (NRTL) and modified Margules (mM) models. Partial molar volumes are obtained from correlations of mixture
densities (Handa and Benson 1979). Isothermal compressibilities are either taken from measurements or estimated with the correlation of Huang and O’Connell (Huang and O’Connell 1987). Figure 6.10 also shows the relation \( y = -1/x_1 \); for complete miscibility, the activity coefficient derivative must always lie above this curve. The simulation value at \( x_1 = 0.1 \) may be unreliable due to diluteness. The results from the NRTL and mM correlations show immiscibility, which is not observed, while the simulations suggest complete miscibility. The Wilson correlation cannot give two liquid phases so it is more consistent with the simulations at dilute alcohol concentrations.

[Insert Figure 6.10]

As described in Chapter 1, neither TCFIs nor DCFIs can be measured directly in experiments, though they can be derived from correlations of experimental data for other thermodynamic properties or integrals of x-ray or neutron scattering measurements. Figure 6.11 shows TCFIs from the correlated results of Figures 6.9 and 6.10 along with results from our Verlet method.

[Insert Figure 6.11]

The TCFIs obtained by our Verlet method apparently converged at all compositions, indicating phase stability over the whole composition. They compare favorably with those from simple truncation (Weerasinghe and Smith 2003b), and the distance-shifting (Hess and van der Vegt 2009) methods. When a simulated system is sufficiently large, the three methods can be expected to yield similar results, but the Verlet method is superior for smaller systems and when the RDFs have significant structure beyond the sampling limit. This is an important result, since the Verlet method might allow thermodynamic derivative properties to be accurately obtained from
simulations of complex systems with relatively low computational effort.

6.5 Future Applications

Classical thermodynamics can only provide relations among properties; values must be found by experiment or computation. Given our current techniques, future applications of TCF-integrations for properties may be achieved in the following areas.

6.5.1 Enzyme solutions

Water activity is an important element of non-aqueous biocatalytic systems. Recently, we have explored different approaches to this property via MD simulation (Wedberg, Abildskov, and Peters 2012). Two main strategies to study how protein properties depend on water activity are termed “real-time” control and “a posteriori” analysis. The former comprises simulations of the protein in a non-aqueous medium in which the number of water molecules is adjusted to maintain a desired water activity. In the latter strategy, conventional MD simulations are carried out, but the water activity is calculated through post-analysis of the simulations. The study of Branco et al. (Branco et al. 2009) is apparently the only work that explicitly considers water activity as a variable. However, their medium was assumed to be an ideal mixture. The greater challenge of nonideal media, such as aqueous organic solutions, has been addressed (Wedberg, Abildskov, and Peters 2012). Much more work needs to be done before establishing a standard method.
6.5.2 Diffusion and Reaction

The current works on properties have involved only equilibrium properties. Computed RDFs for homogeneous, but non-equilibrium, states could lead to local chemical potential gradients for diffusional driving forces and chemical reaction driving forces. Such computations would be unique and powerful for both thermodynamic and transport phenomena.

In particular, the Stefan-Maxwell constitutive equation for multicomponent diffusion in non-ideal solutions (Curtiss and Bird 1999; Wheeler and Newman 2004a, 2004b) has driving forces derived from chemical potential gradients of all but one component. The non-ideality adjustment has been obtained for mutual diffusion in binary mixtures using FST (Jolly and Bearman 1980; Schoen and Hoheisel 1984; Chitra and Smith 2001c). Simulations based on the methods described here could describe higher multicomponent systems, which are of significant interest.

6.6 Conclusions

The successes described in this Chapter for both model and real mixtures indicate that molecular simulation methods for FST should now be ready for greater implementation and extension.

Investigations to refine the various methods to compute KBIs are still ongoing. At this point, our extended Verlet method appears to be the most general and reliable for obtaining thermodynamic properties, especially for dense systems. Its advantages of minimal computational effort and limited need for case-by-case judgment in analysis indicate its efficiency and robustness.
Figure 6.1 (left) RDFs from simulation of water (1) + t-butanol (2) at $x_1 = 0.65$, at 323 K and 1 atm; (right) Numerically evaluated (non-converged) integrals $H_{ij}(R_{lim})$ from Equation 6.9.

Figure 6.2. (a) Distribution function of benzene (1) / ethanol (2), $g_{22}(r)$, at $x_2 = 0.5$; (b) Differential contributions to the $G_{22}$ integral, $dG_{22}$, at $x_2 = 0.5$. (Reprinted with permission from S. Christensen, G. H. Peters, F. Y. Hansen, J. P. O’Connell, and J. Abildskov. 2007. Generation of thermodynamic data for organic liquid mixtures from molecular simulations. Molecular Simulation. 33, 449.)

Figure 6.3. Spatial RDF – Blocks for integrating $h(r)$. Direct from simulation; End-of-Box for matching simulation to fitted function such as Equation 6.18; Long Distance from fitted function.


Figure 6.5. Activity coefficients (based on simulations) of (1) ethanol ($\square, \bullet$) / (2) HFC-227ea ($\Delta, \triangle$) at 283.17K (——) and 343.13 K (– - -). (Reprinted with permission from Modified from S. Christensen, G. H. Peters, F. Y. Hansen, J. P. O’Connell, and J. Abildskov. 2007. State conditions
transferability of vapor-liquid equilibria via fluctuation solution theory with correlation function integrals from molecular dynamics simulation. *Fluid Phase Equilibria*. 260, 169)

**Figure 6.6.** Values (x) for results (left) and relative residuals (right) from calculations of the reduced bulk modulus, \( \rho k_B T \kappa_T \), for the pure LJ fluid at the reduced temperatures \( T^* = 0.85, T^* = 1.0, T^* = 1.5 \) and \( T^* = 2.5 \). Lines derived from EOS of Mecke, *et al.* (Mecke *et al.* 1996) are also shown.

**Figure 6.7.** Values of \( \rho k_B T \kappa_T \) for LJ/Stockmayer mixtures vs. the mole fraction of Stockmayer particles, \( x_2 \), for dipole moments of (a) \( \mu^* = 1 \), (b) \( \mu^* = 2 \) and (c) \( \mu^* = 3 \), derived from the Gross/Vrabec EOS (Gross and Vrabec 2006) (—), compared with results from our MD-Verlet (□), truncation (Weerasinghe and Smith 2003b) (○), and distance-shifting (Hess and van Der Vegt 2009) (Δ) methods.

**Figure 6.8.** Isothermal compressibilities for the mixture water (1) / t-butanol (2). Results from our Verlet (□), truncation (Weerasinghe and Smith 2003b) (○), and distance-shifting (Hess and van Der Vegt 2009) (Δ), methods compared with values obtained from the fluctuation formula (—x—).
Figure 6.9. Relative partial molar volumes ($\rho \bar{V}_1$) for water (1) / t-butanol (2). The results from the Verlet (○), truncation (Weerasinghe and Smith 2003b) (●), and distance-shifting (Hess and van Der Vegt 2009) (Δ) methods for obtaining the TCFIs compared with results from full simulations smoothed with a quadratic polynomial (----) and with smoothed experimental data (-----).

Figure 6.10. Composition derivative of the activity coefficient for water vs. the water mole fraction $x_1$ of water (1) / t-butanol (2). The Verlet method (A) is compared with experimental data smoothed using the Wilson (----), NRTL (---) and mM (•••••••) models. For phase stability, activity coefficient derivatives must everywhere lie above the curve $y = -1/x_1$ (●●●).

Figure 6.11. TCFIs vs. mole fraction, $x_1$, for water (1) / t-butanol (2): (a) $H_{11}$, (b) $H_{12}$, and (c) $H_{22}$, obtained from simulations using the Verlet method (+), compared with TCFIs obtained from experimental data using the procedure of Wooley and O’Connell (Wooley and O’Connell 1991), based on the Wilson (----), NRTL (---) or mM (―) models. The NRTL and mM models approach infinity since they falsely predict a phase split.
Center-to-center distance, $r$, Å

Differential contribution to TCFI, $dG$

$\langle t \rangle^{\Omega \Phi}_{\text{RDF}}$
Direct End-of-box Long Distance

RDF, $g_{22}(r)$

Center-to-center distance, $r$, Å
Activity coefficient, $\ln \gamma_i$

Mole fraction ethanol, $x_1$
\[ T^* = 1.00 \]
\[ T^* = 2.50 \]
\[ T^* = 0.85 \]
\[ T^* = 1.50 \]

\[ \rho^* \] vs. \[ \frac{bk}{T K^T} \] for different reduced temperatures. The graphs show the relationship between reduced density and the reduced temperature, with each graph illustrating the behavior at specific values of reduced temperature. The curves and data points indicate the deviation of the reduced density from an ideal gas behavior as the temperature is varied.
Mole fraction water, $x_1$
Mole fraction water, $x_1$
Mole fraction water, $x_1$
### Symbol Description

#### Greek

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Isobaric thermal expansion coefficient (Equation 1.6)</td>
</tr>
<tr>
<td>Γ&lt;sub&gt;23&lt;/sub&gt;</td>
<td>Preferential binding parameter (Equation 1.86)</td>
</tr>
<tr>
<td>γ&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Mean ion molal activity coefficient (Equation 1.92)</td>
</tr>
<tr>
<td>γ&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Lewis-Randall/rational/mole fraction activity coefficient (Equation 1.19)</td>
</tr>
<tr>
<td>γ&lt;sup&gt;m&lt;/sup&gt;</td>
<td>Molar activity coefficient</td>
</tr>
<tr>
<td>Δ</td>
<td>Isothermal-isobaric partition function (Equation 1.28)</td>
</tr>
<tr>
<td>ΔG&lt;sub&gt;ij&lt;/sub&gt;</td>
<td>$G_i + G_{ij} - 2G_{ij}$ (Equation 1.93)</td>
</tr>
<tr>
<td>ζ&lt;sub&gt;2&lt;/sub&gt;</td>
<td>$1 + c_i G_{ii} + c_i G_{ij} + c_i c_j (G_{ii} G_{jj} - G_{ij}^2)$ (Equation 1.66)</td>
</tr>
<tr>
<td>η&lt;sub&gt;12&lt;/sub&gt;</td>
<td>$c_i + c_j (G_{ii} - G_{ij}^2 - 2G_{ij})$ (Equation 1.66)</td>
</tr>
<tr>
<td>κ&lt;sub&gt;T&lt;/sub&gt;</td>
<td>Isothermal compressibility (Equation 1.5)</td>
</tr>
<tr>
<td>Λ&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Thermal de Broglie wavelength of specie i</td>
</tr>
<tr>
<td>λ&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Absolute activity of i</td>
</tr>
<tr>
<td>µ&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Chemical potential of component i</td>
</tr>
<tr>
<td>µ&lt;sub&gt;ij&lt;/sub&gt;</td>
<td>Chemical potential derivative (Equation 1.1)</td>
</tr>
<tr>
<td>ν</td>
<td>Number of cations/anions, $ν = ν_+ + ν_-$</td>
</tr>
<tr>
<td>Ξ</td>
<td>Grand canonical partition function (Equation 1.28)</td>
</tr>
<tr>
<td>π</td>
<td>Osmotic pressure</td>
</tr>
<tr>
<td>ρ</td>
<td>Mass or total number density</td>
</tr>
<tr>
<td>ρ&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Number density of $i$, $= N_i/V$; see also $c_i$</td>
</tr>
<tr>
<td>φ&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Volume fraction of $i$, $= ρ_i V_i$</td>
</tr>
<tr>
<td>φ&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Fugacity coefficient of $i$ (Equation 1.23)</td>
</tr>
<tr>
<td>Ω</td>
<td>Microcanonical partition function (Equation 1.28)</td>
</tr>
</tbody>
</table>

#### Mathematical

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;X&gt;</td>
<td>Ensemble or time average</td>
</tr>
<tr>
<td>[X]</td>
<td>Set notation, ${X_1, X_2, \ldots}$</td>
</tr>
<tr>
<td>I</td>
<td>Unit matrix</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Latin

+ Cation
- Anion
∞ Pure
Infinitely dilute (limiting)
1 Solvent
1/k_BT β
2 Solute
3 Cosolvent/cosolute/additive
ΔG Reaction Gibbs energy (function)
ΔH Reaction enthalpy
ΔS Reaction entropy
δx_{ji} Preferential solvation parameter (j surrounding i)
δx'_{ji} Corrected referential solvation parameter
μVT Grand canonical ensemble
\hat{X} Fourier transformed X
\overline{X}_j Partial molar property of X
a_± Mean activity of electrolyte in solution
A Helmholtz energy (function)
A_i Aggregate/Multimer of i monomers
a_i Activity of i
aq Aqueous solution
c_i Molarity of i, see also number density, ρ_i
C_{ij} ρ_j \int c_i(r)dr, DCFI, elements of the \mathbf{C} matrix (Equation 1.39)
c_{ij}(r) Direct correlation function
C_p Constant pressure heat capacity (Equation 1.7)
D Activity derivative, concentration fluctuation term (Equation 1.73)
f_i Fugacity of a substance i in a gaseous mixture (Equation 1.21)
G Gibbs energy (function)
g_{ij} Radial (pair) distribution function, RDF
G_{ij} Kirkwood-Buff integral, KBI
H Enthalpy
h Planck’s constant
H_{ij} ρ_j \int h_i(r)dr = ρ G_{ij}, TCF (see below Equation 1.38), or Henry’s law constant
h_i(r) Total correlation function, TCF, g_{ij}(r)-1
id Ideal (mole fraction scale)
K Equilibrium constant
k Rate constant
k_B Boltzmann constant
\( k_H \)  Henry’s law constant, see also \( H_{ij} \)

\( M \)  Monomer

\( m \)  \( m \)-value for protein denaturation (see Equation 1.99)

\( m_{c} = c_{c}/c_{1} \)  (Dimensionless) molality

\( \text{mix} \)  Mixing process

\( n \)  Number of monomers in an aggregate

\( n_{c} \)  Number of components in the system

\( N_{A} \)  Avogadro’s number

\( N_{i} \)  Number of entities (usually molecules, atoms, or ions)

\( N_{ij} \)  Excess coordination number

\( N_{pT} \)  Isothermal-isobaric (Gibbs) ensemble

\( N_{VE} \)  Microcanonical ensemble

\( N_{VT} \)  Canonical ensemble

\( p \)  Pressure

\( Q \)  Canonical partition function (Equation 1.28)

\( r \)  \( |r_{1} - r_{2}| \), distance between COM of molecules

\( R \)  Gas constant

\( R_{cor} \)  Correlation radius (see \( V_{cor} \))

\( S \)  Entropy

\( T \)  Temperature (thermodynamic)

\( T_{m} \)  Melting temperature

\( \text{trs} \)  Transfer between two phases

\( U \)  Internal energy

\( V \)  Volume

\( V_{cor} \)  Correlation volume (see Equation 1.81)

\( X^{*} \)  Reduced or characteristic quantity \( X \)

\( X_{c} \)  Critical \( X \) (\( X \) is pressure or temperature)

\( X^{E} \)  Excess of \( X \)

\( X^{r} \)  Residual of quantity \( X \)

\( x_{i} \)  Liquid phase mole fraction composition

\( X_{m} \)  Molar quantity

\( y_{i} \)  Gas phase mole fraction composition, or solute solubility

\( z_{+/-} \)  Charge of cation/anion

**Acronyms**

- **COM**  Center of mass
- **DCF**  Direct correlation function
- **DCFI**  Direct correlation function integral
- **EOS**  Equation of state
- **FF**  Force Field
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>FST</td>
<td>Fluctuation solution theory</td>
</tr>
<tr>
<td>FT</td>
<td>Fluctuation theory</td>
</tr>
<tr>
<td>GD</td>
<td>Gibbs-Duhem</td>
</tr>
<tr>
<td>IG</td>
<td>Ideal gas</td>
</tr>
<tr>
<td>KB</td>
<td>Kirkwood-Buff</td>
</tr>
<tr>
<td>KBFF</td>
<td>Kirkwood-Buff Force Field</td>
</tr>
<tr>
<td>KBI</td>
<td>Kirkwood-Buff integral</td>
</tr>
<tr>
<td>LJ</td>
<td>Lennard-Jones</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MDF</td>
<td>Molecular Distribution Function</td>
</tr>
<tr>
<td>MM</td>
<td>McMillan-Mayer</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular Weight</td>
</tr>
<tr>
<td>NRTL</td>
<td>Non-Random Two Liquid</td>
</tr>
<tr>
<td>OSA</td>
<td>Osmotic stress analysis</td>
</tr>
<tr>
<td>OZ</td>
<td>Ornstein-Zernike</td>
</tr>
<tr>
<td>PF</td>
<td>Partition function</td>
</tr>
<tr>
<td>PI</td>
<td>Preferential Interaction</td>
</tr>
<tr>
<td>PMF</td>
<td>Potential of mean force</td>
</tr>
<tr>
<td>PS</td>
<td>Preferential Solvation</td>
</tr>
<tr>
<td>PY</td>
<td>Percus-Yevick</td>
</tr>
<tr>
<td>RDF</td>
<td>Radial distribution function</td>
</tr>
<tr>
<td>RK</td>
<td>Redlich-Kister</td>
</tr>
<tr>
<td>RISM</td>
<td>Reference interaction-site model</td>
</tr>
<tr>
<td>SAFT</td>
<td>Statistical associated-fluid theory</td>
</tr>
<tr>
<td>SANS</td>
<td>Small-angle neutron scattering</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small-angle X-ray scattering</td>
</tr>
<tr>
<td>SI</td>
<td>Symmetric ideal</td>
</tr>
<tr>
<td>SPT</td>
<td>Scaled particle theory</td>
</tr>
<tr>
<td>TCF</td>
<td>Total correlation function</td>
</tr>
<tr>
<td>TCFI</td>
<td>Total correlation function integral</td>
</tr>
<tr>
<td>UNIFAC</td>
<td>UNIversal Functional Activity Coefficient</td>
</tr>
<tr>
<td>UNIQUAC</td>
<td>UNIversal QUAsiChemical</td>
</tr>
<tr>
<td>VDW</td>
<td>van der Waals</td>
</tr>
</tbody>
</table>


Brelvi, S. W., and J. P. O’Connell. 1972. Corresponding states correlations for liquid compressibility and partial molal volumes of gases at infinite dilution in liquids. American Institute of Chemical Engineers Journal. 18, 1239.


solubility at sub- and near-critical conditions. *American Institute of Chemical Engineers Journal.* 42, 571.


Gmehling, J. 1995. From UNIFAC to modified UNIFAC to PSRK with the help of DDB. *Fluid Phase Equilibria.* 107, 1.


Edited by J. Gormally and E. Wyn-Jones. Amsterdam: Elsevier Scientific Publishing Co.


Jonah, D. A. 1986. On a method of numerical differentiation and its application to the evaluation
of limiting activity-coefficients and their 1st derivatives. **Chemical Engineering Science.** 41, 2261.


Kan, A. T., and M. B. Tomson. 1996. UNIFAC prediction of aqueous and nonaqueous solubilities of chemicals with environmental interest. **Environmental Science and Technology.** 30, 1369.


Kežić, B., and A. Perera. 2011. Towards a more accurate reference interaction site model integral


Kumar, S. K., and K. P. Johnston. 1988. Modelling the solubility of solids in supercritical fluids with density as the independent variable. Journal of Supercritical Fluids. 1, 15.


Lin, C.-L., and R. H. Wood. 1996. Prediction of the free energy of dilute aqueous methane, ethane, and propane at temperatures from 600 to 1200 degrees C and densities from 0 to 1 g cm⁻³ using molecular dynamics simulations. *Journal of Physical Chemistry.* 100, 16399.


Liu, Y., and D. Bolen. 1995. The peptide backbone plays a dominant role in protein stabilization
by naturally occurring osmolytes. *Biochemistry.* 34, 12884.


Majer, V., L. Hui, R. Crovetto, and R. H. Wood. 1991. Volumetric properties of aqueous 1-1 electrolyte-solutions near and above the critical-temperature of water. 1. Densities and apparent molar volumes of NaCl(aq) from 0.0025 mol·kg\(^{-1}\) to 3.1 mol·kg\(^{-1}\), 604.4-K to 725.5-K, and 18.5-MPa to 38.0-MPa. *Journal of Chemical Thermodynamics.* 23, 213.


Marcus, Y. 2002a. 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. *Physical Chemistry Chemical Physics.* 4, 4462.


Munoz, F., T. W. Li, and E. H. Chimowitz. 1995. Henry's law and synergism in dilute near-


solutions. *Journal of the American Chemical Society.* 129, 4476.


Vlcek, L., A. A. Chialvo, and D. R. Cole. 2011. Optimized unlike-pair interactions for water-
carbon dioxide mixtures described by the SPC/E and EPM2 models. Journal of Physical Chemistry B. 115, 8775.


Zeck, S., and H. Knapp. 1985. Solubilities of ethylene, ethane, and carbon-dioxide in mixed-


