Analysis and Extension of Perturbation Theories for Modelling Helmholtz Energies

A thesis accepted by the Faculty of Energy-, Process- and Bio-Engineering and the Stuttgart Center for Simulation Science (SimTech) of the University of Stuttgart in partial fulfilment of the requirements for the degree of Doctor of Engineering Sciences (Dr.-Ing.)

by

Anja Maike Reimer

born in Stuttgart

Examiner:	Prof. Dr.-Ing. Joachim Gross
Co-Examiner:	Prof. Øivind Wilhelmsen
Date of defense:	16.01.2026

Institute of Thermodynamics and Thermal Process Engineering
University of Stuttgart

2026

Contents

List of Figures

List of Tables

Abstract

Equations of state expressed in the Helmholtz energy enable accurate predictions of a broad range of fluid properties, i.e., PVT-behavior, densities, phase equilibria, or heat capacities, at low computational cost. Perturbation theories provide a powerful physical basis for developing equations of state. In this work, an equation of state for Mie ν-6 fluids is developed within the framework of the $u\nu$-theory. The equation of state is based on interpolation between two first-order perturbation theories, and reproduces the virial expansion up to third order at low densities. The new model is not only comparable in accuracy to state-of-the-art empirical models, but also offers several advantages due to its physical basis. These include a better description of the meta-stable and unstable regions, and a more systematic extension to non-spherical fluids and mixtures. As a first step towards extending the equation of state to chain fluids, an algebraic model for the second virial coefficient of Lennard-Jones chains is developed. This model combines a third-order perturbation theory with empirical terms adjusted to newly generated molecular simulation data. Although perturbation theories can be formulated rigorously for mixtures, this is rarely done in practice, as deriving algebraic models is challenging, and numerical evaluation is computationally demanding. Instead, one-fluid approximations are typically used, treating mixtures as effective pure fluids at the cost of accuracy. An algebraic correction term is developed for size-asymmetric mixtures of square-well and Lennard-Jones fluids. This correction can be applied to various perturbation-theory-based equations of state using a one-fluid approximation. The correction term improves the prediction of mixture properties without introducing (significant) additional complexity. The functional form of the correction was obtained using symbolic regression. More complex interactions, specifically Coulomb interactions are additionally addressed. Formulating analytic models for electrolytes based on perturbation theories is challenging due to non-converging integrals. A common approach is to replace terms associated with Coulomb interactions with simple Helmholtz energy expressions based on the Debye-Hückel theory, the mean spherical approximation, or the Born theory of solvation. Using molecular simulations to generate reference data for ion-related Helmholtz energy contributions, it is shown that these substitutions are not sufficiently accurate without empirical adjustments.

Kurzfassung

Zustandsgleichungen ermöglichen die präzise Vorhersage zahlreicher Stoffeigenschaften, beispielsweise des PVT-Verhaltens sowie von Phasengleichgewichten, bei geringem Rechenaufwand. Störungstheorien bilden eine solide physikalische Grundlage für die Entwicklung von Zustandsgleichungen. In dieser Arbeit wird die Anwendung von Störungstheorien zur Modellierung der Helmholtz-Energie verschiedener Fluidsysteme analysiert. Zunächst wird eine Zustandsgleichung (ZSG) für Mie-ν-6 Fluide entwickelt, die auf einer Interpolation zwischen zwei Störungstheorien erster Ordnung basiert. Die neue ZSG erreicht eine Genauigkeit, die mit den derzeit besten verfügbaren empirischen Modellen vergleichbar ist. Bedingt durch ihre physikalische Grundlage bietet sie jedoch mehrere Vorteile: Sie ermöglicht eine konsistentere Beschreibung metastabiler und instabiler Fluidregionen und lässt sich systematischer auf nicht-sphärische Moleküle und Mischungen erweitern. Zur Erweiterung der ZSG auf kettenförmige Moleküle wird ein analytisches Modell für den zweiten Virialkoeffizienten von Lennard-Jones-Ketten entwickelt. Dieses Modell kombiniert eine Störungstheorie dritter Ordnung mit empirischen Termen. Im Weiteren wird die Beschreibung von Mischungen durch Störungstheorien untersucht. Störungstheorien können grundsätzlich rigoros für Mischungen formuliert werden. In der Praxis wird dies jedoch selten umgesetzt, da die Herleitung analytischer Modelle schwierig ist und die numerische Auswertung mit hohen Rechenkosten verbunden ist. Stattdessen werden Mischungen häufig als effektive Reinstoffe angenähert und mit Reinstoffmodellen beschrieben. Solche Annäherungen führen zu einem Verlust an Modellgenauigkeit. Um diesen Verlust zu kompensieren, ohne signifikante zusätzliche Komplexität einzuführen, wird eine analytische Mischungskorrektur entwickelt. Eine geeignete funktionale Form der Korrektur wird dabei mittels symbolischer Regression ermittelt. Abschließend wird die Modellierung von Elektrolyten thematisiert. In Störungstheorien für ionische Systeme treten aufgrund langreichweitiger Coulomb-Wechselwirkungen nicht konvergente Integrale auf. Üblicherweise werden deshalb die mit Coulomb-Wechselwirkungen verbundenen Terme durch einfache Helmholtz-Energie-Modelle basierend auf der Debye-Hückel-Theorie, der Mean-Spherical-Approximation oder der Born-Theorie ersetzt. Anhand von Referenzdaten aus Molekularsimulationen wird gezeigt, dass diese Substitutionen ohne empirische Anpassungen nicht ausreichend genau sind.

Journal Publications

This thesis is based on the following publications:

- Chapter 3: A. Reimer, T. van Westen and J. Gross. Physically based equation of state for Mie ν-6 fluids. *J. Chem. Phys.*, **158**, 164506 (2023), doi:10.1063/5.0141856

  T. van Westen and J. Gross supervised this work and were involved in editing the manuscript.

- Chapter 4: A. Reimer, J. Gross and T. van Westen. Effect of Chain Conformation on the Free Energy of Dilute Polymer Solutions: Monte Carlo Simulations and Perturbation Theory for the Second Virial Coefficient of Lennard–Jones Chains. *Macromolecules*, **58**, 3, 1498–1511 (2025), doi:10.1021/acs.macromol.4c02827

  T. van Westen co-conceptualized and supervised this work. J. Gross supervised this work and was involved in editing the manuscript.

- Chapter 5: A. Reimer, T. Winkler-Markert and J. Gross. Beyond One-Fluid Approximations for the Thermodynamics of Fluid Mixtures. *J. Chem. Phys.*, **164**, 014117 (2026), doi:10.1063/5.0307122

  T. Winkler-Markert performed molecular simulations and wrote the molecular simulations section. J. Gross supervised this work and was involved in editing the manuscript.

- Chapter 6: A. Reimer, I. Reisch and J. Gross. Primitive and non-primitive model electrolytes: Comparing ion-related Helmholtz energies using molecular simulations. *J. Chem. Phys.*, **162**, 124503 (2025), doi:10.1063/5.0257401

  I. Reisch implemented the division of the Ewald summation and the Lennard-Jones interactions under my supervision. J. Gross supervised this work and was involved in editing the manuscript.

Chapters 3–6 present literal quotes of the published or submitted work respectively. Any addition with respect to the published or submitted work is marked. Any deletion or substitution is indicated with square brackets as '[...]'. Cross-references between chapters of this thesis may have been added to the published content to increase readability. The supporting information and appendices to the published and submitted work are presented in the appendix of this thesis.

1 Importance of New Helmholtz Energy Models

Equations of state (EoS) expressed in terms of the Helmholtz energy capture the full thermodynamic information of a fluid in a single analytic model. They enable thermodynamically consistent prediction of equilibrium properties such as pressure–volume–temperature behaviour, phase equilibria, and caloric properties at low computational cost. Therefore, EoS are powerful tools for numerous technical applications, including process design and optimization in the chemical and pharmaceutical industries. Over the past century, a variety of EoS have been developed, which can be roughly categorized into two groups: empirical models and physically based models.

Empirical EoS, including classical multi-parameter and machine-learned models, are usually developed for a single substance or a limited set of substances. They can achieve very high accuracy when fitted to extensive experimental or molecular simulation data, but their applicability faces several limitations: Empirical EoS can only be developed for systems with sufficient high-quality data. As they have no physical basis, they may predict unphysical behaviour, such as multiple van der Waals loops, and their predictions cannot be reliably extrapolated beyond the state points covered by the reference data. Furthermore, these models cannot be systematically extended to additional substances or mixtures. Examples of empirical EoS include the IAPWS formulation for water[1], the reference equations of Span and co-workers[2], e.g., for nitrogen[3,4] and carbon dioxide[5], and the GERG-2008 model[6], which covers 21 components, most of them present in natural gas, as well as their mixtures. More recently, machine-learned EoS have been proposed for model fluids, such as spherical Mie fluids[7] and Mie chains[8], and for real substances like polystyrene[9].

Physically based EoS are generally less accurate than empirical models but offer broader applicability. They are typically developed to describe a model fluid with well-defined inter- and intramolecular potentials, rather than to represent real substances directly. The potential parameters, such as segment size or interaction strength, are assigned by fitting to experimental data or relating them to molecular properties. The parameters can be varied to describe a broad range of substances using the same equation of state. Physically based EoS rely on fewer adjustable parameters than empirical models, enabling predictions for substances with limited experimental data. Their physical basis facilitates reliable extrapolation beyond the state points included in the experimental data.

A popular and successful pathway for developing physically based EoS are perturbation theories[10–14]. In perturbation theories, the potential energy is divided into a reference and a perturbation part, and the Helmholtz energy is expressed as a series expansion about the reference fluid. Prominent examples include variants of the statistical associating fluid theory (SAFT)[15–27], such as PCP-SAFT[28–33] and SAFT-VR Mie[34].

The currently available EoS based on perturbation theory face several practical limitations. The accuracy but also the complexity of a perturbation theory increases with its order: first-order theories rely only on pair correlations, while second-order models already require information about triplet and quadruplet correlations. Most broadly applicable EoS are based on second- or third-order perturbation theories. They are typically formulated for relatively simple model fluids, such as linear chains. Extensions to more realistic fluids, including branching, intramolecular angles, or rigidity, are possible in principle, but developing analytic expressions is tedious. Although perturbation theories allow for a rigorous mixture description, most current models rely on simplifying approximations. These approximations reduce computational cost and model complexity but introduce inaccuracies. Additional challenges arise for fluids containing charged substances. When applying standard perturbation theories to electrolytes, diverging correlation integrals occur. Electrostatic contributions are therefore typically approximated using simple Helmholtz energy expressions based on the Debye-Hückel theory[35], the mean-spherical approximation[36,37], or the Born theory of solvation[38], rather than being treated more rigorously. These approximations limit both the predictive power and the physical insight provided by these models. Existing EoS, both empirical and physically based, offer complementary strengths but also have inherent limitations. This motivates the development of new models that combine the broad applicability and theoretical foundation of physically based EoS with the accuracy of empirical EoS, while overcoming the limitations of both. The long-term objective is a physically consistent, widely applicable, and highly accurate equation of state. This thesis takes the first steps toward achieving this vision. The thesis is structured around four main studies, each tackling a distinct challenge in molecular modelling.

In chapter 2, the fundamentals of perturbation theories are reviewed, and the uv-theory[39] is introduced. The uv-theory is a physically based framework for EoS development that approximates the Helmholtz energy by interpolating between two first-order perturbation theories, one of which exactly reproduces the low-density limit. By relying solely on first-order perturbation theories, the uv-theory is more readily extensible to complex fluids than higher-order approaches, making it a promising starting point for further developments.

Chapter 3 investigates how the uv-theory can be systematically improved and whether it can reach accuracy comparable to empirical models without losing the advantages of its physical

foundation. The uv-B_3-theory, a new equation of state for Mie ν-6 fluids incorporating an analytic third virial coefficient, is developed and its performance is compared to the best available EoS for Lennard-Jones fluids ($\nu = 12$).

In chapter 4, chain molecules are studied, where intramolecular interactions and configurations must be considered. An analytic model for the second virial coefficient of Lennard-Jones chains is developed, which may serve as input for extending the uv-theory to chainlike fluids. Molecular simulations are performed to generate reference data and to investigate the low-density behaviour of chains with varying rigidity.

Standard approximations for modelling mixtures within perturbation theories are reviewed in chapter 5. A correction to the one-fluid approximation is developed by combining physics with machine learning through symbolic regression. The approach is tested for square-well, Lennard-Jones, and real mixtures and applied within different EoS, including the square-well model of Gross and Sadowski[40], the PC-SAFT equation of state[28], and the uv-theory[39].

In chapter 6, electrolytes are investigated to improve the theoretical understanding of the corresponding Helmholtz energy contributions in these systems and move toward more predictive models. Molecular simulations are performed to generate reference data for ion related Helmholtz energy contributions. State-of-the-art approximations for electrolyte modelling are systematically assessed by comparing the simulation results to predictions from classical expressions based on the Debye–Hückel theory, the mean-spherical approximation, and the Born theory of solvation.

References

[1] W. Wagner and A. Pruß. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *J. Phys. Chem. Ref. Data*, **31**(2):387–535, 2002.

[2] R. Span. *Multiparameter equations of state: an accurate source of thermodynamic property data*. Springer Science & Business Media, 2013.

[3] R. Span, E. Lemmon, R. Jacobsen, and W. Wagner. A reference quality equation of state for nitrogen. *Int. J. Thermophys.*, **19**:1121–1132, 1998.

[4] R. Span, E. W. Lemmon, R. T. Jacobsen, W. Wagner, and A. Yokozeki. A reference equation of state for the thermodynamic properties of nitrogen for temperatures from 63.151 to 1000 K and pressures to 2200 MPa. *J. Phys. Chem. Ref. Data*, **29**(6):1361–1433, 2000.

[5] R. Span and W. Wagner. A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa. *J. Phys. Chem. Ref. Data*, **25**(6):1509–1596, 1996.

[6] O. Kunz and W. Wagner. The GERG-2008 wide-range equation of state for natural gases and other mixtures: An expansion of GERG-2004. *J. Chem. Eng. Data*, **57**(11):3032–3091, 2012.

[7] G. Chaparro and E. A. Müller. Development of thermodynamically consistent machine-learning equations of state: Application to the Mie fluid. *J. Chem. Phys.*, **158**(18), 2023.

[8] K. Zhu and E. A. Müller. Generating a machine-learned equation of state for fluid properties. *J. Phys. Chem. B*, **124**(39):8628–8639, 2020.

[9] J. Hinz, D. Yu, D. S. Pandey, H. Sapkota, Q. Yu, D. Mihaylov, V. Karasiev, and S. Hu. The development of thermodynamically consistent and physics-informed equation-of-state model through machine learning. *APL Machine Learning*, **2**(2), 2024.

[10] C. G. Gray and K. E. Gubbins. *Theory of Molecular Fluids*. Oxford University Press, 1984. ISBN 9780198556022, 9780191919251. doi:10.1093/oso/9780198556022.001.0001.

[11] J.-P. Hansen and I. R. McDonald. *Theory of simple liquids: with applications to soft matter*. Academic press, 2013.

[12] S. Zhou and J. R. Solana. Progress in the Perturbation Approach in Fluid and Fluid-Related Theories. *Chem. Rev.*, **109**(6):2829–2858, 2009. doi:10.1021/cr900094p.

[13] K. E. Gubbins. The theory of non-electrolyte solutions: An historical review. *Mol. Phys.*, **111**(24):3666–3697, 2013. doi:10.1080/00268976.2013.831140.

[14] K. E. Gubbins. Perturbation theories of the thermodynamics of polar and associating liquids: A

historical perspective. *Fluid Phase Equilib.*, **416**:3–17, 2016. doi:10.1016/j.fluid.2015.12.043.

[15] W. G. Chapman, G. Jackson, and K. E. Gubbins. Phase equilibria of associating fluids: Chain molecules with multiple bonding sites. *Mol. Phys.*, **65**(5):1057–1079, 1988. doi:10.1080/00268978800101601.

[16] W. Chapman, K. Gubbins, G. Jackson, and M. Radosz. SAFT: Equation-of-state solution model for associating fluids. *Fluid Phase Equilib.*, **52**:31–38, 1989. doi:10.1016/0378-3812(89)80308-5.

[17] W. G. Chapman, K. E. Gubbins, G. Jackson, and M. Radosz. New reference equation of state for associating liquids. *Ind. Eng. Chem. Res.*, **29**(8):1709–1721, 1990. doi:10.1021/ie00104a021.

[18] S. H. Huang and M. Radosz. Equation of state for small, large, polydisperse, and associating molecules. *Ind. Eng. Chem. Res.*, **29**(11):2284–2294, 1990. doi:10.1021/ie00107a014.

[19] S. H. Huang and M. Radosz. Equation of state for small, large, polydisperse, and associating molecules: extension to fluid mixtures. *Ind. Eng. Chem. Res.*, **30**(8):1994–2005, 1991. doi:10.1021/ie00056a050.

[20] F. J. Blas and L. F. Vega. Thermodynamic behaviour of homonuclear and heteronuclear Lennard-Jones chains with association sites from simulation and theory. *Mol. Phys.*, **92**(1):135–150, 1997. doi:10.1080/002689797170707.

[21] F. J. Blas and L. F. Vega. Prediction of Binary and Ternary Diagrams Using the Statistical Associating Fluid Theory (SAFT) Equation of State. *Ind. Eng. Chem. Res.*, **37**(2):660–674, 1998. doi:10.1021/ie970449+.

[22] A. Gil-Villegas, A. Galindo, P. J. Whitehead, S. J. Mills, G. Jackson, and A. N. Burgess. Statistical associating fluid theory for chain molecules with attractive potentials of variable range. *J. Chem. Phys.*, **106**(10):4168–4186, 1997. doi:10.1063/1.473101.

[23] A. Galindo, L. A. Davies, A. Gil-Villegas, and G. Jackson. The thermodynamics of mixtures and the corresponding mixing rules in the SAFT-VR approach for potentials of variable range. *Mol. Phys.*, **93**(2):241–252, 1998. doi:10.1080/002689798169249.

[24] A. Galindo, A. Gil-Villegas, G. Jackson, and A. N. Burgess. SAFT-VRE: Phase Behavior of Electrolyte Solutions with the Statistical Associating Fluid Theory for Potentials of Variable Range. *J. Phys. Chem. B*, **103**(46):10272–10281, 1999. doi:10.1021/jp991959f.

[25] A. Gil-Villegas, A. Galindo, and G. Jackson. A statistical associating fluid theory for electrolyte solutions(SAFT-VRE). *Mol. Phys.*, **99**(6):531–546, 2001. doi:10.1080/00268970010018666.

[26] A. Lymperiadis, C. S. Adjiman, A. Galindo, and G. Jackson. A group contribution method for associating chain molecules based on the statistical associating fluid theory (SAFT-γ). *J. Chem. Phys.*, **127**(23):234903, 2007. doi:10.1063/1.2813894.

[27] V. Papaioannou, T. Lafitte, C. Avendaño, C. S. Adjiman, G. Jackson, E. A. Müller, and A. Galindo. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. *J. Chem. Phys.*, **140**(5):054107, 2014. doi:10.1063/1.4851455.

[28] J. Gross and G. Sadowski. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Ind. Eng. Chem. Res.*, **40**(4):1244–1260, 2001. doi:10.1021/ie0003887.

[29] J. Gross and G. Sadowski. Modeling Polymer Systems Using the Perturbed-Chain Statistical Associating Fluid Theory Equation of State. *Ind. Eng. Chem. Res.*, **41**(5):1084–1093, 2002. doi:10.1021/ie010449g.

[30] J. Gross and G. Sadowski. Application of the Perturbed-Chain SAFT Equation of State to Associating Systems. *Ind. Eng. Chem. Res.*, **41**(22):5510–5515, 2002. doi:10.1021/ie010954d.

[31] J. Gross and J. Vrabec. An equation-of-state contribution for polar components: Dipolar molecules. *AIChE J.*, **52**(3):1194–1204, 2006. doi:10.1002/aic.10683.

[32] J. Gross. An equation-of-state contribution for polar components: Quadrupolar molecules. *AIChE J.*, **51**(9):2556–2568, 2005. doi:10.1002/aic.10502.

[33] J. Vrabec and J. Gross. Vapor–Liquid Equilibria Simulation and an Equation of State Contribution for Dipole–Quadrupole Interactions. *J. Phys. Chem. B*, **112**(1):51–60, 2008. doi:10.1021/jp072619u.

[34] T. Lafitte, A. Apostolakou, C. Avendaño, A. Galindo, C. S. Adjiman, E. A. Müller, and G. Jackson. Accurate statistical associating fluid theory for chain molecules formed from Mie segments. *J. Chem. Phys.*, **139**(15):154504, 2013. doi:10.1063/1.4819786.

[35] P. Debye and E. Hückel. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Phys. Z.*, **24**(9):185–206, 1923.

[36] L. Blum. Mean spherical model for asymmetric electrolytes: I. Method of solution. *Mol. Phys.*, **30**(5):1529–1535, 1975.

[37] L. Blum and J. Høye. Mean spherical model for asymmetric electrolytes. 2. Thermodynamic properties and the pair correlation function. *J. Phys. Chem.*, **81**(13):1311–1316, 1977.

[38] M. Born. Volumen und Hydratationswärme der Ionen. *Z. Phys.*, **1**(1):45–48, 1920.

[39] T. van Westen and J. Gross. Accurate thermodynamics of simple fluids and chain fluids based on first-order perturbation theory and second virial coefficients: uv-theory. *J. Chem. Phys.*, **155**(24):244501, 2021. doi:10.1063/5.0073572.

[40] J. Gross and G. Sadowski. Application of perturbation theory to a hard-chain reference fluid: an

equation of state for square-well chains. *Fluid Phase Equilib.*, **168**(2):183–199, 2000.

2 Physically Based Approaches for Modelling Helmholtz Energies

In this chapter, fundamental concepts of statistical mechanics and perturbation theory are introduced as the basis for developing physically-based Helmholtz energy models. The treatment closely follows established references, in particular Gray and Gubbins [1], McQuarrie [2], and Tuckerman [3]. These concepts provide the theoretical basis for the uv-theory [4], which is presented in section 2.4.

Unless stated otherwise, pure fluids are considered. All derivations are restricted to homogeneous systems of molecules without intramolecular degrees of freedom. Intermolecular interactions are assumed to be spherically symmetric, i.e., depending only on the scalar distance between molecular centres.

2.1 Helmholtz Energy in the Canonical Ensemble

Consider a canonical ensemble characterized by a constant number of particles or molecules N, volume V, and temperature T. The corresponding thermodynamic potential is the Helmholtz energy $A(N,V,T)$. Knowledge of A as a function of its natural variables provides access to all other equilibrium thermodynamic properties through differentiation. For example, the pressure p, entropy S, and internal energy U,

$$p = -\left(\frac{\partial A}{\partial V}\right)_{N,T} \qquad S = -\left(\frac{\partial A}{\partial T}\right)_{N,V} \qquad U = A + TS = \left(\frac{\partial (A/T)}{\partial (1/T)}\right)_{V,N} \tag{2.1}$$

Statistical mechanics establishes the connection between macroscopic thermodynamics and molecular interactions by expressing thermodynamic potentials in terms of partition functions. In the canonical ensemble, the Helmholtz energy is related to the canonical partition function Q_{NVT} through

$$A = -k_\mathrm{B} T \ln Q_{NVT} \tag{2.2}$$

where k_B denotes Boltzmann's constant. For a system of N identical particles at positions $\mathbf{r}^N = \{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$ and with momenta $\mathbf{p}^N = \{\mathbf{p}_1, \ldots, \mathbf{p}_N\}$, the canonical partition function is

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int \int \exp\big(-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)\big) \mathrm{d}\mathbf{r}^N \mathbf{p}^N \tag{2.3}$$

with Planck's constant h, the inverse temperature $\beta = (k_\mathrm{B} T)^{-1}$, and the Hamiltonian

$$\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) = K(\mathbf{p}^N) + U(\mathbf{r}^N) \tag{2.4}$$

The kinetic energy is

$$K(\mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m} \tag{2.5}$$

where m is the particle mass. Assuming pairwise additive interactions, the potential energy can be written as

$$U(\mathbf{r}^N) = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} u(r_{ij}) \tag{2.6}$$

where $u(r_{ij})$ denotes the intermolecular pair potential at distance $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$. Integration of eq. (2.3) over the momenta of all particles yields a decomposition of the canonical partition function into an ideal-gas contribution and a residual contribution,

$$Q_{NVT} = Q_{NVT}^{\mathrm{ig}} Q_{NVT}^{\mathrm{res}} \tag{2.7}$$

with

$$Q_{NVT}^{\mathrm{ig}} = \frac{V^N}{\Lambda^{3N} N!} \qquad \text{and} \qquad Q_{NVT}^{\mathrm{res}} = \frac{1}{V^N} \int \exp\big[-\beta U(\mathbf{r}^N)\big] \, d\mathbf{r}^N \tag{2.8}$$

where Λ is the de Broglie wavelength,

$$\Lambda = \frac{h}{\sqrt{2\pi m k_\mathrm{B} T}} \tag{2.9}$$

Inserting the decomposed partition function into eq. (2.2) yields the corresponding decomposition of the Helmholtz energy,

$$A = A^{\mathrm{ig}} + A^{\mathrm{res}} \tag{2.10}$$

Applying Stirling's approximation $\ln(N!) \approx N\ln(N) - N$ the ideal gas contribution can be written as

$$A^{\mathrm{ig}} = Nk_{\mathrm{B}}T\left(\ln\left(\rho\Lambda^3\right) - 1\right) \tag{2.11}$$

with the number density $\rho = N/V$. A direct evaluation of the residual contribution

$$A^{\mathrm{res}} = -k_{\mathbf{B}}T\ln Q^{\mathrm{res}}_{NVT} \tag{2.12}$$

is not feasible due to the high $(3N)$ dimensionality of the configuration integral. Approximations are therefore required. Two common strategies are: (1) series expansions about limiting cases, such as the virial expansion about the low-density limit, and (2) perturbation theories, where the target fluid is represented by an expansion about the structural properties of a well-known reference fluid. In the following, the concepts of the virial expansion and perturbation theories are introduced.

2.2 Virial Equation of State

In its standard form[5], the virial equation of state is written as a series expansion of the compressibility factor Z about the low density limit $\rho = 0$,

$$Z = \frac{p}{\rho k_{\mathrm{B}}T} = 1 + \sum_{i=2}^{\infty} B_i(T)\rho^{i-1} \tag{2.13}$$

where $B_i(T)$ are the virial coefficients. Introducing the reduced Helmholtz energy $a = A/(Nk_{\mathrm{B}}T)$, the compressibility factor can be written as

$$Z = \rho\left(\frac{\partial a}{\partial \rho}\right)_{N,T} = 1 + \rho\left(\frac{\partial a^{\mathrm{res}}}{\partial \rho}\right)_{N,T} \tag{2.14}$$

Equating this expression with eq. (2.13) and integrating under the condition $a^{\mathrm{res}}(\rho \to 0) = 0$ gives the virial expansion of the residual Helmholtz energy,

$$a^{\mathrm{res}} = \sum_{i=2}^{\infty} \frac{B_i}{i-1}\rho^{i-1} = B_2\rho + \frac{1}{2}B_3\rho^2 + \frac{1}{3}B_4\rho^3 + \ldots \tag{2.15}$$

Expressions for the virial coefficients in terms of the Mayer function, $f(r) = \exp(-\beta u(r)) - 1$, are obtained from a cluster expansion of the configurational integral.[2] For example, the

second and third virial coefficients are

$$B_2 = -\frac{1}{2}\int f(r_{12})\mathrm{d}\mathbf{r}_{12} \tag{2.16}$$

$$B_3 = -\frac{1}{3}\iint f(r_{12})f(r_{13})f(r_{23})\mathrm{d}\mathbf{r}_{12}\mathrm{d}\mathbf{r}_{13} \tag{2.17}$$

The virial expansion is exact in the low-density limit and provides accurate descriptions of gases even when truncated after the first two terms. At liquid-like densities, however, the series may converge slowly, necessitating the inclusion of a substantial number of higher-order terms. The dimensionality of the integrals defining the virial coefficients B_i increases rapidly with i,[5] making the computation of higher-order terms demanding.

2.3 Perturbation Theory

In contrast to the virial equation of state, perturbation theories are typically developed for fluids in the liquid state. In his pioneering work, van der Waals suggested that the molecular structure of liquids is primarily determined by strong short-range repulsive interactions, while attractions and other longer-ranged interactions play a subordinate role.[6,7] Thermodynamic perturbation theories build on this picture by decomposing the total potential energy U into a repulsive reference part U_0 and a perturbation part W. The Helmholtz energy is then expressed as a series expansion about that of the reference fluid with potential energy U_0. The first statistical-mechanical formulation of a perturbation theory was proposed by Zwanzig,[8] and subsequently refined in the seminal works of Barker and Henderson[9] and Weeks, Chandler, and Andersen.[10–12]

To derive a perturbation theory, the potential energy is written in terms of a coupling parameter $0 \leq \lambda \leq 1$,

$$U_\lambda(\mathbf{r}^N) = U_0(\mathbf{r}^N) + W_\lambda(\mathbf{r}^N) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} u_0(r_{ij}) + \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} w_\lambda(r_{ij}) \tag{2.18}$$

which interpolates between the reference fluid with $U_{\lambda=0} = U_0$ and the target fluid with $U_{\lambda=1} = U$. The difference between the Helmholtz energies of the target and reference systems

can then be written as

$$A - A_0 = \int_0^1 \left(\frac{\partial A_\lambda}{\partial \lambda}\right)_{N,V,T} \mathrm{d}\lambda = \int_0^1 \frac{1}{Z_\lambda} \int \left(\frac{\partial U_\lambda}{\partial \lambda}\right)_{N,V,T} \exp\left(-\beta U_\lambda\left(\mathbf{r}^N\right)\right) \mathrm{d}\mathbf{r}^N \mathrm{d}\lambda \tag{2.19}$$

with the configuration integral

$$Z_\lambda = \int \exp\left(-\beta U_\lambda(\mathbf{r}^N)\right) \mathrm{d}\mathbf{r}^N \tag{2.20}$$

Exploiting translational invariance and introducing the radial distance $r = |\mathbf{r}| = |\mathbf{r}_2 - \mathbf{r}_1|$, eq. (2.19) can be simplified to

$$A - A_0 = \frac{1}{2}\rho N \int_0^1 \int \left(\frac{\partial w_\lambda(r)}{\partial \lambda}\right) g_\lambda(r) \mathrm{d}\mathbf{r} \mathrm{d}\lambda \tag{2.21}$$

with the pair distribution function

$$g_\lambda(r) = \frac{N(N-1)}{Z_\lambda \rho^2} \int \exp\left(-\beta U_\lambda(\mathbf{r}^N)\right) \mathrm{d}\mathbf{r}^{N-2} \tag{2.22}$$

where $\mathrm{d}\mathbf{r}^{N-2}$ denotes integration over all particle coordinates except those of particles 1 and 2. Expanding the integrand of eq. (2.21) in a Taylor series about $\lambda = 0$ and performing the integration over λ, yields a perturbation series for the Helmholtz energy,

$$A = A_0 + \sum_{i=1}^{\infty} \Delta A_i \tag{2.23}$$

with an infinite number of perturbation terms ΔA_i. Truncation of this series after the nth term defines an nth-order perturbation theory. In practice, the series is usually truncated after the first few terms (typically one to three). Higher orders improve accuracy but rapidly increase complexity, since the derivatives $(\partial^{n-1} g_\lambda(r)/\partial \lambda^{n-1})_{\lambda=0}$ involve multi-particle distribution functions. In particular, the nth-order perturbation term involves distribution functions of up to $2n$ particles.[5,13] In the following, the perturbation series is derived explicitly for two alternative definitions of $w_\lambda(r)$, leading to the u-expansion and the f-expansion.

2.3.1 u-Expansion

In the u-expansion (also referred to as the high-temperature expansion) [1,5], linear coupling is employed,

$$u_\lambda(r) = u_0(r) + \lambda w(r) \tag{2.24}$$

with $w(r) = u(r) - u_0(r)$. Substituting $w_\lambda(r) = \lambda w(r)$ into eq. (2.21) gives

$$A - A_0 = \frac{1}{2} N\rho \int_0^1 \int w(r) g_\lambda(r) \mathrm{d}\mathbf{r}\mathrm{d}\lambda \tag{2.25}$$

Expanding $g_\lambda(r)$ about $\lambda = 0$ and carrying out the integration over λ yields the perturbation series of eq. (2.23), with the first- and second-order terms

$$\Delta A_1^u = \frac{1}{2}\rho N \int w(r) g_0(r) \mathrm{d}\mathbf{r} \tag{2.26}$$

$$\Delta A_2^u = \frac{1}{4}\rho N \int w(r) \left(\frac{\partial g_\lambda(r)}{\partial \lambda} \right)_{\lambda=0} \mathrm{d}\mathbf{r} \tag{2.27}$$

where $g_0(r)$ denotes the pair distribution function of the reference fluid and the superscript u indicates the u-expansion.

The accuracy of the u-expansion for modelling Lennard-Jones fluids was investigated by Van Westen and Gross [14]. They found that a first-order u-expansion provides accurate results at high densities, where repulsive interactions dominate and the structures of the reference and target fluids are similar. At low densities, however, the accuracy decreases significantly: the series converges slowly, and higher-order terms add considerable complexity while providing only limited improvement.

2.3.2 f-Expansion

In the f-expansion [1,11,15,16] the coupling scheme is defined as

$$u_\lambda(r) = u_0(r) - k_\mathrm{B} T \ln[1 + \lambda f(w(r))] \tag{2.28}$$

where $f(w(r)) = \exp[-\beta w(r)] - 1$ is the Mayer function with $w(r) = u(r) - u_0(r)$. Substituting

$$w_\lambda(r) = -k_\mathrm{B} T \ln[1 + \lambda f(w(r))] \tag{2.29}$$

into eq. (2.21) gives

$$A-A_0=-\frac{1}{2}k_\mathrm{B}T\rho N\int_0^1\int\Delta e(r)\frac{g_\lambda(r)}{e_\lambda(r)}\mathrm{d}\mathbf{r}\mathrm{d}\lambda \tag{2.30}$$

with $\Delta e(r)=e(r)-e_0(r)$, $e_\lambda(r)=e_0(r)+\lambda\Delta e(r)$, and the Boltzmann factors $e(r)=\exp(-\beta u(r))$ and $e_0(r)=\exp(-\beta u_0(r))$. Expanding the integrand about $\lambda=0$ and integration over λ leads to the perturbation series of eq. (2.23) with the first- and second-order terms given by

$$\Delta A_1^f=-\frac{1}{2}k_\mathrm{B}T\rho N\int f(w(r))g_0(r)\mathrm{d}\mathbf{r} \tag{2.31}$$

$$\Delta A_2^f=-\frac{1}{4}k_\mathrm{B}T\rho N\int\Delta e(r)\left(\frac{\partial(g_\lambda(r)/e_\lambda(r))}{\partial\lambda}\right)_{\lambda=0}\mathrm{d}\mathbf{r} \tag{2.32}$$

A key strength of the f-expansion is its exactness in the low-density limit,[17,18] where the pair distribution function of the reference fluid reduces to $g_0(r)=e_0(r)$. In this limit, the first-order f-expansion coincides with the second-order virial expansion,

$$\lim_{\rho\to0}\left(\frac{\beta\left(A_0+\Delta A_1^f\right)}{N}\right)=a^\mathrm{ig}+B_{20}\rho-\frac{1}{2}\rho\int\left(\frac{e(r)}{e_0(r)}-1\right)e_0(r)\mathrm{d}\mathbf{r}=a^\mathrm{ig}+B_2\rho \tag{2.33}$$

with the second virial coefficient of the reference fluid

$$B_{20}=-\frac{1}{2}\int(e_0(r)-1)\mathrm{d}\mathbf{r} \tag{2.34}$$

and the reduced Helmholtz energy of the ideal gas $a^\mathrm{ig}=\beta A^\mathrm{ig}/N$.

2.3.3 Fluid Mixtures

Perturbation theories can be rigorously extended to multi-component systems. For a mixture of S species, the total potential energy can be expressed as

$$U(\mathbf{r}^N)=\frac{1}{2}\sum_{\alpha=1}^{S}\sum_{\gamma=1}^{S}\sum_{i=1}^{N_\alpha}\sum_{\substack{j=1\\ \alpha i\neq\gamma j}}^{N_\gamma}u_{\alpha\gamma}(r_{\alpha i,\gamma j}) \tag{2.35}$$

where N_α is the number of particles of species α, and $u_{\alpha\gamma}(r_{\alpha i,\gamma j})$ denotes the pair potential between particles i and j of species α and γ, separated by a distance $r_{\alpha i,\gamma j}=|\mathbf{r}_{\gamma j}-\mathbf{r}_{\alpha i}|$. The

coupling scheme is defined as

$$u_{\lambda,\alpha\gamma}(r) = u_{0,\alpha\gamma}(r) + w_{\lambda,\alpha\gamma}(r) \tag{2.36}$$

Following the same approach as for the pure fluid case, the Helmholtz energy difference between the target and reference systems is given by

$$A - A_0 = \frac{1}{2}\rho N \sum_{\alpha=1}^{S}\sum_{\gamma=1}^{S} x_\alpha x_\gamma \int_0^1 \int \left(\frac{\partial w_{\lambda,\alpha\gamma}(r_{\alpha 1,\gamma 2})}{\partial \lambda}\right) g_{\lambda,\alpha\gamma}(r_{\alpha 1,\gamma 2}) \mathrm{d}\mathbf{r}_{\alpha\mathbf{1},\gamma\mathbf{2}} \mathrm{d}\lambda \tag{2.37}$$

where $x_\alpha = N_\alpha/N$ is the mole fraction of species α and $N = \sum_{\alpha=1}^{S} N_\alpha$ is the total number of particles. The radial distribution function between a particle of species α and a particle of species γ in the mixture is defined as

$$g_{\lambda,\alpha\gamma}(r_{\alpha i,\gamma j}) = \frac{N_\alpha(N_\gamma - \delta_{\alpha\gamma})}{Z_\lambda \rho_\alpha \rho_\gamma} \int \exp\left(-\beta U_\lambda(\mathbf{r}^N)\right) \frac{\mathrm{d}\mathbf{r}^N}{\mathrm{d}\mathbf{r}_{\alpha i}\mathrm{d}\mathbf{r}_{\gamma j}} \tag{2.38}$$

where $\rho_\alpha = N_\alpha/V$ and the Kronecker delta $\delta_{\alpha\gamma}$ excludes self-interactions. The first-order perturbation term of the u-expansion follows as

$$\Delta A_1^u = \frac{1}{2}\rho N \sum_{\alpha=1}^{S}\sum_{\gamma=1}^{S} x_\alpha x_\gamma \int w_{\alpha\gamma}(r) g_{0,\alpha\gamma}(r) \mathrm{d}\mathbf{r} \tag{2.39}$$

where $w_{\alpha\gamma}(r) = u_{\alpha\gamma}(r) - u_{0,\alpha\gamma}(r)$, $r = |\mathbf{r}_{\alpha\mathbf{1},\gamma\mathbf{2}}|$, and $g_{0,\alpha\gamma}(r)$ is the pair distribution function of species α and γ in the reference fluid.

The extension of perturbation theories to mixtures is conceptually straightforward. Their practical evaluation, however, is already challenging at first order and becomes increasingly complex at higher orders. Evaluating eq. (2.39) requires models or data for all distribution functions $g_{0,\alpha\gamma}(r)$ as functions of density, temperature, and composition. Such information is rarely available. Even when it is available, solving the integrals for all species pairs within the double summation is computationally expensive. Consequently, equations of state based on perturbation theory typically rely on simplifying approximations to model fluid mixtures.

One of the most notable approximations is the one-fluid approximation, which states that a pair distribution function, once written in dimensionless variables is approximately uniform for all pairs $\alpha\gamma$ and equal to that of a pure fluid. The argument of eq. (2.39) can be cast in dimensionless form using $r^* = r/\sigma_{\alpha\gamma}$ and $w^*(r^*) = w(r^*)/\varepsilon_{\alpha\gamma}$, where $\sigma_{\alpha\gamma}$ and $\varepsilon_{\alpha\gamma}$ are the size

and energy parameters defining the pair interactions, yielding

$$\Delta A_1^u = \frac{1}{2}\rho N \sum_{\alpha=1}^{S}\sum_{\gamma=1}^{S} x_\alpha x_\gamma \varepsilon_{\alpha\gamma}\sigma_{\alpha\gamma}^3 \int w^*(r^*) g_{0,\alpha\gamma}(r^*) \mathrm{d}\mathbf{r}^* \tag{2.40}$$

$$\approx \frac{1}{2}\rho N \sum_{\alpha=1}^{S}\sum_{\gamma=1}^{S} \left(x_\alpha x_\gamma \varepsilon_{\alpha\gamma}\sigma_{\alpha\gamma}^3\right) \int w^*(r^*) g_0(r^*) \mathrm{d}\mathbf{r}^* \tag{2.41}$$

Equation (2.41) introduces the one-fluid approximation, namely that $g_{0,\alpha\gamma}(r^*)$ is approximately independent of individual pairs and equal to $g_0(r^*)$ of a pure fluid, $g_{0,\alpha\gamma}(r^*) \approx g_0(r^*)$. A slightly less strict definition of the one-fluid approximation is the statement that the entire integrals in eqs. (2.40) and (2.41) are approximately equal.

2.3.4 Electrolytes

Modelling electrolyte systems with perturbation theories is particularly challenging due to the long-ranged nature of electrostatic interactions. The main difficulties when using standard reference fluids, such as the hard-sphere fluid, are (1) the poor convergence of the perturbation series, and (2) the divergence of correlation integrals.[19]

Ion–ion interactions are described by the Coulomb potential,

$$u_{\alpha\gamma}(r) = \frac{q_\alpha q_\gamma}{r} \tag{2.42}$$

with charges q_α and q_γ of ionic species α and γ. The slow decay of the potential ($\sim 1/r$) causes electrostatic effects to persist over long distances. These effects lead to structural correlations that differ strongly from those of non-electrostatic reference fluids and thus to poor convergence of perturbation expansions. Insertion of the Coulomb potential into eq. (2.16) yields a divergent second virial coefficient. Divergent integrals also occur in the perturbation terms of both the f- and u-expansions for perturbation potentials $w(r) \sim 1/r$. Special treatments are therefore required in perturbation theories of electrolytes. Theiss and Gross[19] proposed the use of damped Coulomb potentials. A widely used empirical approach is to replace the perturbation terms by analytic expressions based on the Debye-Hückel theory[20], the mean-spherical approximation[21,22], or the Born theory of solvation[23]

2.4 Equation of State Based on First-Order Perturbation Theories: The uv-Theory

A central framework for developing equations of state in this work is the uv-theory.[4] Conventional perturbation-theory based approaches, typically truncate the series of eq. (2.23) at finite order and approximate the required correlation integrals with analytic expressions. In contrast, the uv-theory relies solely on first-order perturbation theory. It combines two distinct first-order expansions within an interpolation scheme: a standard u-expansion, which is accurate at high densities, and a modified u-expansion that recovers the exact second virial coefficient at low densities.

The uv-theory is derived by expressing the Helmholtz energy A as an interpolation between a lower bound A^{lb} and an upper bound A^{ub},

$$A = A^{\mathrm{lb}} + \phi\left(A^{\mathrm{ub}} - A^{\mathrm{lb}}\right) \tag{2.43}$$

where $A^{\mathrm{lb}} \leq A \leq A^{\mathrm{ub}}$ and $0 \leq \phi \leq 1$ is an interpolation function that generally depends on density and temperature.

According to the Gibbs-Bogoliubov inequality[24], the Helmholtz energy of a first-order u-expansion satisfies

$$A - A_0 \leq \Delta A_1^u \tag{2.44}$$

which makes

$$A^{\mathrm{ub}} = A_0 + \Delta A_1^u \tag{2.45}$$

a rigorous and convenient choice for the upper bound. No rigorous lower bound of comparable simplicity is available. To overcome this, van Westen and Gross[4,17] introduced effective lower bounds which reproduce the exact low-density behaviour of the fluid. In an initial formulation,[17] the first-order f-expansion was adopted,

$$A^{\mathrm{lb}} = A_0 + \Delta A_1^f \tag{2.46}$$

leading to the uf-theory,

$$A = A_0 + \phi\,\Delta A_1^u + (1-\phi)\Delta A_1^f \tag{2.47}$$

This approach was later refined by replacing the f-expansion with a tighter effective bound that still reduces to the second-order virial description at low densities. In the uv-theory,[4] the lower bound is defined as

$$A^{\mathrm{lb}} = A_0 + \Delta A_1^u + N k_\mathrm{B} T (\Delta B_2 - \Delta B_{21}^u)\rho \tag{2.48}$$

with $\Delta B_2 = B_2 - B_{20}$ where B_2 and B_{20} are the second virial coefficients of the target fluid and the reference fluid, respectively, and $\Delta B_{21}^u = \left(\partial \Delta a_1^u / \partial \rho\right)_{\rho \to 0}$. In the limit of zero density this bound reduces to,

$$\lim_{\rho \to 0} \frac{A^{\mathrm{lb}}}{N k_\mathrm{B} T} = B_{20}\rho + \Delta B_{21}^u \rho + (\Delta B_2 - \Delta B_{21}^u)\rho = B_2 \rho \tag{2.49}$$

Insertion of the upper and lower bounds into interpolation scheme of eq. (2.43) and restructuring leads to the uv-theory,

$$A = A_0 + \Delta A_1^u + N k_\mathrm{B} T \, (1 - \phi) \left(\Delta B_2 - \Delta B_{21}^u\right) \rho \tag{2.50}$$

Here, ϕ will approach unity at high densities, where the u-expansion is accurate,[14,17] and must vanish at low densities, where the lower bound becomes exact. With these constraints, suitable ansatz functions for ϕ can be chosen and fitted to experimental or molecular simulation data.

The uv-theory combines the strengths of the u- and f-expansions: it reproduces the exact second virial coefficient in the low-density limit while preserving the high-density accuracy of the u-expansion. Relying solely on first-order perturbation terms, it requires only relatively simple inputs and can be applied to complex systems more easily and more rigorously than higher-order perturbation theories.

References

[1] C. G. Gray and K. E. Gubbins. *Theory of Molecular Fluids*. Oxford University Press, New York, 1st edition, 1984.

[2] D. A. McQuarrie. *Statistical Mechanics*. University Science Books, Sausalito, CA, 2000. ISBN 1-891389-15-7.

[3] M. E. Tuckerman. *Statistical mechanics: theory and molecular simulation*. Oxford University Press, 2023.

[4] T. van Westen and J. Gross. Accurate thermodynamics of simple fluids and chain fluids based on first-order perturbation theory and second virial coefficients: uv-theory. *J. Chem. Phys.*, **155**(24):244501, 2021. doi:10.1063/5.0073572.

[5] J.-P. Hansen and I. R. McDonald. *Theory of simple liquids: with applications to soft matter*. Academic press, 2013.

[6] J. D. Van Der Waals and J. S. Rowlinson. *On the continuity of the gaseous and liquid states*. Courier Corporation, 2004.

[7] D. Chandler, J. D. Weeks, and H. C. Andersen. Van der Waals picture of liquids, solids, and phase transformations. *Science*, **220**(4599):787–794, 1983.

[8] R. W. Zwanzig. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.*, **22**(8):1420–1426, 1954.

[9] J. A. Barker and D. Henderson. Perturbation Theory and Equation of State for Fluids. II. A Successful Theory of Liquids. *J. Chem. Phys.*, **47**(11):4714–4721, 1967. doi:10.1063/1.1701689.

[10] D. Chandler and J. D. Weeks. Equilibrium Structure of Simple Liquids. *Phys. Rev. Lett.*, **25**(3):149–152, 1970. doi:10.1103/physrevlett.25.149.

[11] H. C. Andersen, J. D. Weeks, and D. Chandler. Relationship between the Hard-Sphere Fluid and Fluids with Realistic Repulsive Forces. *Phys. Rev. A*, **4**(4):1597–1607, 1971. doi:10.1103/physreva.4.1597.

[12] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.*, **54**(12):5237–5247, 1971. doi:10.1063/1.1674820.

[13] F. Drunsel. *Perturbation Theory and Molecular Simulation of Nonprimitive Model Electrolyte Solutions*. Ph.D. thesis, Universität Stuttgart, Stuttgart, Germany, 2021.

[14] T. Van Westen and J. Gross. A critical evaluation of perturbation theories by Monte Carlo simulation of the first four perturbation terms in a Helmholtz energy expansion for the Lennard-

Jones fluid. *J. Chem. Phys.*, **147**(1), 2017.

[15] K. Gubbins, W. Smith, M. Tham, and E. Tiepel. Perturbation theory for the radial distribution function. *Mol. Phys.*, **22**(6):1089–1105, 1971.

[16] J. A. Barker. Cluster integrals and statistical mechanics of solutions. *Proc. R. Soc. Lond. A*, **241**(1227):547–553, 1957.

[17] T. van Westen and J. Gross. Accurate first-order perturbation theory for fluids: uf-theory. *J. Chem. Phys.*, **154**(4):041102, 2021. doi:10.1063/5.0031545.

[18] J. E. Mayer and M. G. Mayer. *Statistical mechanics*. John Wiley & Sons New York, 7th edition, 1940.

[19] M. Theiss and J. Gross. Perturbation approaches for describing dipolar fluids and electrolyte solutions. *J. Chem. Phys.*, **153**(4), 2020.

[20] P. Debye and E. Hückel. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Phys. Z.*, **24**(9):185–206, 1923.

[21] L. Blum. Mean spherical model for asymmetric electrolytes: I. Method of solution. *Mol. Phys.*, **30**(5):1529–1535, 1975.

[22] L. Blum and J. Høye. Mean spherical model for asymmetric electrolytes. 2. Thermodynamic properties and the pair correlation function. *J. Phys. Chem.*, **81**(13):1311–1316, 1977.

[23] M. Born. Volumen und Hydratationswärme der Ionen. *Z. Phys.*, **1**(1):45–48, 1920.

[24] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Elsevier, 2023.

3 Physically Based Equation of State for Mie ν-6 Fluids

The content of this chapter is a literal quote of the publication:

A. Reimer, T. van Westen and J. Gross. Physically Based Equation of State for Mie ν-6 Fluids. *J. Chem. Phys.*, **158**, 164506 (2023), doi: 10.1063/5.0141856

3.1 Abstract

We develop a physically based equation of state that describes Mie ν-6 fluids with an accuracy comparable to that of state-of-the-art empirical models. The equation of state is developed within the framework of the *uv*-theory [J. Chem. Phys. **155**, 244501 (2021)], which is modified by incorporating the third virial coefficient B_3 in the low-density description of the model. The new model interpolates between a first-order Weeks-Chandler-Andersen (WCA) perturbation theory at high densities, and a modified first-order WCA theory that recovers the virial expansion up to B_3 at low densities. A new algebraic equation for the third virial coefficient of Mie ν-6 fluids is developed - other inputs are taken from previous work. Predicted thermodynamic properties and phase equilibria are compared to a comprehensive data-base of molecular simulation results from the literature, including Mie fluids of repulsive exponents $9 \leq \nu \leq 48$. The new equation of state is applicable to states with densities up to $\rho^*(T^*) \lessapprox 1.1 + 0.12\sqrt{T^*}$ and temperatures $T^* > 0.3$. For the Lennard-Jones fluid ($\nu = 12$), the performance of the model is comparable to that of the best empirical equations of state available. The physical basis of the new model provides several advantages however: (1) the new model is applicable to Mie fluids of repulsive exponents $9 \leq \nu \leq 48$ instead of only $\nu = 12$, (2) the model leads to a better description of the meta-stable and unstable region (which is important for describing interfacial properties by classical Density Functional Theory), and (3) being a first-order perturbation theory, the new model (potentially) allows an easier and more rigorous extension to non-spherical (chain) fluids and mixtures.

3.2 Introduction

Equations of state (EoS) expressed in the Helmholtz energy summarize the full thermodynamic information of a system in a single analytic model. They are a powerful tool for predicting fluid properties at very low computational effort. Among equation-of-state models, a rough distinction can be made between fully empirical approaches and physically based approaches[1]. Fully empirical models (recently reviewed by Huber *et al.*[2]) typically contain a large number of parameters which are fitted to reference data obtained from either experiments or molecular simulations. Such models can provide a very accurate description of fluid properties and phase behaviour; however, extrapolations into the meta-stable region often fail and the applicability to mixtures or different molecule types is not granted. Physically based models are often less accurate, but much more versatile: (1) they allow a description of a wider range of different molecule types, varying in size, shape and interaction strength, (2) their extension to mixtures can be done more rigorously, and (3) they lead to a better (qualitative) description of the meta-stable region, which is relevant for describing the properties of inhomogeneous fluids (e.g. surface tensions) via classical density functional theory (DFT).

Popular examples of physically based EoS are those developed based on perturbation theory.[3–7] In such theories, the intermolecular potential is divided into a reference and a perturbation part; the Helmholtz energy can then be formulated as a series expansion about the reference fluid, where each coefficient in the series can be calculated based on structural properties of the reference fluid only. Though the accuracy of a perturbation theory increases with its order, so does its complexity.[8]

While first-order perturbation theories merely require information on pair correlations, higher-order theories also need correlation functions of triplets, quadruplets and so on, which are not easily obtained nor extended to mixtures. For this reason, first-order perturbation theories are superior when it comes to the benefits of physically based models discussed above:

Their extension to mixtures can be done rigorously, based on the pair-correlation function of mixtures or based on unambiguous approximations made at the level of the pair-correlation function.[9] Also, they are less prone to numerical artifacts (e.g. oscillating isotherms in the meta-/unstable vapour-liquid region) as compared to models based on higher-order perturbation theories.[10]

Many EoS are a combination of the two approaches, comprising contributions with a strong physical background as well as empirical parts. However, in mixing the two, some of the benefits of physically based models typically become compromised. In this paper, we aim to marry both approaches, that is, to develop an equation of state with a comparable level of

accuracy as fully empirical models, without sacrificing the benefits of physically based EoS, particularly those of first-order perturbation theories. The framework we use for this is the *uv*-theory.[11]

The *uv*-theory is based on an interpolation of the Helmholtz energy between a first-order perturbation theory (*u*-expansion) at high densities, and a modified first-order perturbation theory that recovers the exact second virial coefficient at low densities.

The development of the *uv*-theory can be placed in the context of a rich history of approaches aimed at combining accurate low-density and high-density descriptions of fluids in a single equation of state.[12–18] Unique to the *uv*-theory, in this respect, is that it maintains all benefits of first-order perturbation theory but can reach much higher accuracy due to including the exact second virial coefficient. In previous work, the *uv*-theory was successfully applied to develop very accurate EoS for Lennard-Jones (LJ) fluid mixtures[11], Mie fluids[11], LJ dimers[11] and LJ-spline fluids[19].

In this work, we further refine the *uv*-theory for Mie ν-6 fluids, with the aim of reaching an accuracy comparable to the best empirical models for the LJ fluid[17,20–23].

While in most regions of the fluid-phase diagram the original *uv*-theory provides an excellent description of Mie ν-6 fluids, the analysis of this work showed that slight improvements to the original model are required to reach the desired accuracy; particularly for fluids at very low density ($\rho^* \lessapprox 0.05$) and temperatures below the critical point, and fluids at very high density (roughly $\rho^* \gtrapprox 1$). To improve the accuracy of the *uv*-theory at those conditions, we incorporate the third virial coefficient in its low-density description and suggest a slightly modified description of the reference fluid.

Compared to the LJ potential (ν=12), the Mie ν-6 potential allows the repulsiveness of the intermolecular potential ν to be treated as an additional model parameter of the equation of state. This has been shown beneficial for describing fluid-phase properties[24–27]. We keep the attractive exponent fixed at a value of 6, which has some theoretical justification in the decay of London dispersion forces of simple molecules/atoms.[28,29] Other state-of-the art models for describing Mie ν-6 fluids are the SAFT-VR-Mie[30] equation of state (which can be applied to attractive exponents other than 6), and the *uf*-theory[9].

3.3 Equation of State

3.3.1 Molecular Model

The equation of state presented in this work describes fluids comprising molecules that interact by a Mie ν-6 pair potential

$$u(r) = \mathcal{C}\varepsilon\left[\left(\frac{\sigma}{r}\right)^{\nu} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{3.1}$$

where r is the distance between the two molecules, σ is the size parameter, ε is the energy parameter, ν is the repulsive Mie exponent, and the prefactor $\mathcal{C}$ ensures $u(r_\mathrm{m}) = -\varepsilon$ as the minimum value of the potential,

$$\mathcal{C} = \frac{\nu}{\nu-6}\left(\frac{\nu}{6}\right)^{\frac{6}{\nu-6}} \tag{3.2}$$

with

$$r_\mathrm{m} = \sigma\left(\frac{\nu}{6}\right)^{\frac{1}{(\nu-6)}} \tag{3.3}$$

as the distance at which the potential of eq. (3.1) is at its minimum.

We divide the Mie potential as proposed by Weeks, Chandler and Anderson (WCA)[31–33] into the reference potential

$$u_0(r) = \begin{cases} u(r) + \varepsilon & \text{for } r < r_\mathrm{m} \\ 0 & \text{for } r \geq r_\mathrm{m} \end{cases} \tag{3.4}$$

and the perturbation potential

$$w(r) = u(r) - u_0(r) \tag{3.5}$$

In the remainder, all thermodynamic variables and properties are made dimensionless using σ and ε as unit length and energy, and ε/k_B, where k_B is Boltzmann's constant, as unit temperature. Dimensionless quantities are denoted by an asterisk: e.g. temperature $T^* = k_\mathrm{B}T/\varepsilon$, pressure $p^* = p\sigma^3/\varepsilon$, density $\rho^* = \rho\sigma^3$, internal energy per molecule $u^* = U/(N\varepsilon)$, and specific isochoric heat capacity per molecule $c_\mathrm{V}^* = C_V/(Nk_\mathrm{B})$.

3.3.2 Original uv-Theory

The idea of the uv-theory[11] is to express the Helmholtz energy as an interpolation between an upper bound (superscript ub) and a lower bound (superscript lb) according to

$$a = \frac{A}{N k_B T} = a^{\mathrm{lb}} + \phi^u (a^{\mathrm{ub}} - a^{\mathrm{lb}}) \tag{3.6}$$

with Boltzmann's constant k_B, $a^{\mathrm{lb}} \leq a \leq a^{\mathrm{ub}}$ at each state point, and the interpolation function $0 \leq \phi^u \leq 1$, referred to as u-fraction. The u-fraction depends on the thermodynamic state, which is defined by the temperature T and the density $\rho = N/V$, where V denotes the volume and N the number of particles. The upper bound is described rigorously[4,34], based on a first-order u-expansion[3,4],

$$a^{\mathrm{ub}} = a_0 + \Delta a_1^u \tag{3.7}$$

where a_0 is the Helmholtz energy of the reference fluid of intermolecular potential $u_0(r)$ defined in eq. (3.4), and Δa_1^u is the corresponding first-order perturbation term

$$\Delta a_1^u = 2\pi \rho^* \beta^* \int_0^\infty g_0(r^*) w^*(r^*) r^{*2} \mathrm{d}r^* \tag{3.8}$$

with $r^* = r/\sigma$, $\beta^* = 1/T^*$, the pair-correlation function of the reference fluid $g_0(r^*)$ and the dimensionless perturbation potential $w^*(r^*) = w(r^*)/\varepsilon$ defined in eq. (3.5). The lower bound is approximated by the upper bound, eq. (3.7), modified such that it recovers the exact second virial coefficient B_2,

$$a^{\mathrm{lb,\,eff}} = a_0 + \Delta a_1^u + (\Delta B_2^* - \Delta B_{21}^{u*}) \rho^* \tag{3.9}$$

with

$$\Delta B_{21}^{u*} = \lim_{\rho^* \to 0} \left(\frac{\partial \Delta a_1^u}{\partial \rho^*} \right)_{N, T^*} \tag{3.10}$$

and $\Delta B_2^* = B_2^* - B_{20}^*$, where B_{20}^* is the second virial coefficient of the reference fluid. Insertion of eqs. (3.7) and (3.9) into the interpolation scheme, eq. (3.6), leads to the uv-theory

$$a = a_0 + \Delta a_1^u + (1 - \phi^u)(\Delta B_2^* - \Delta B_{21}^{u*}) \rho^* \tag{3.11}$$

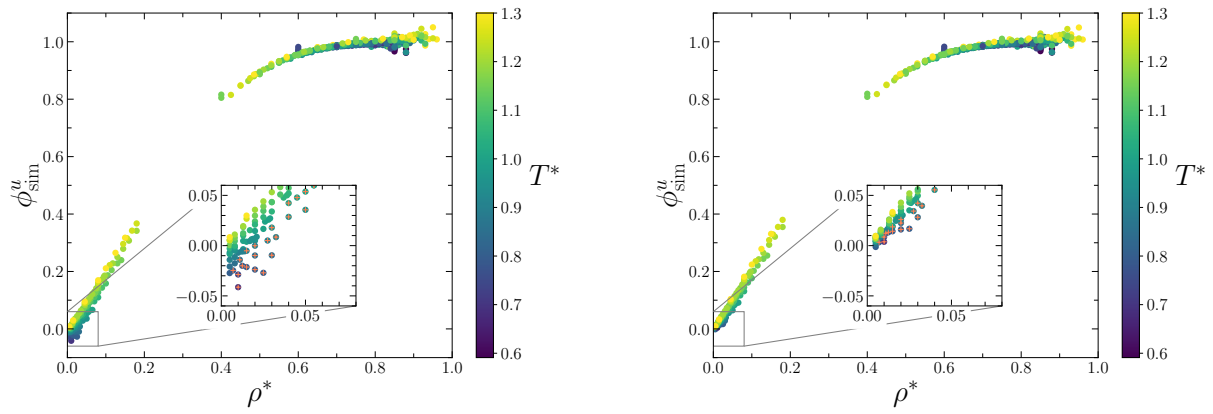
Details on calculating the different inputs of this equation are given by van Westen and Gross.[11]

The u-fraction ϕ^u is the only empirical contribution to eq. (3.11). Nevertheless, the (limiting) behaviour of ϕ^u is for a very large part dictated by the characteristics of the upper and lower bounds of eqs. (3.7) and (3.9): (1) ϕ^u must become zero for $\rho \to 0$ to recover the second virial coefficient, (2) ϕ^u must be close to unity for high densities, since the accuracy of a first-order u-expansion grows when density increases[3,35], (3) $\phi^u < 1$ because a first-order u-expansion is a rigorous upper bound to the Helmholtz energy, (4) $0 < \phi^u$, if the effective lower bound $a^{\text{lb, eff}}$ is a true lower bound of the Helmholtz energy. An ansatz function for ϕ^u that meets the mentioned requirements can be defined and fitted to reference data from simulations or experiments.
In the original uv-theory for Mie fluids[11], the u-fraction was fitted to vapour-liquid equilibria, pressure and internal energy. When using a WCA division of the intermolecular potential into reference and perturbation parts, the temperature dependence of the u-fraction was shown to be small, and an ansatz function that depends only on density could be used.
Although the original uv-theory leads to a very accurate equation of state for Mie fluids, especially in the vapour phase the uv-theory is less accurate than the best fully empirical models from the literature. The slight inaccuracies in the vapour-phase description are caused by the effective lower bound of eq. (3.9), which is no actual lower bound of the Helmholtz energy for Mie fluids[9,11]. We illustrate this in figure 3.1(a), where we show the u-fraction ϕ^u_{sim}, obtained by inserting simulation results for the Helmholtz energy of a LJ fluid into eq. (3.11). The other contributions to eq. (3.11) are obtained from the uv-WCA-theory for Mie fluids by van Westen and Gross[11]. The negative values of the u-fraction for low densities and temperatures indicate that the actual Helmholtz energy is smaller than the effective lower bound $a^{\text{lb, eff}}$. The reason for this is the strongly negative third virial coefficient of the LJ fluid at low temperatures, which can't be captured by a first-order u-expansion. For a u-fraction constrained between zero and unity, this behaviour can not be described.

3.3.3 Increasing the Accuracy of the uv-Theory for Mie ν-6 Fluids: uv-B_3-Theory

We propose a modified effective lower bound of the Helmholtz energy that incorporates the third virial coefficient B_3:

$$a^{\text{lb, eff}} = a_0 + \Delta a_1^u + \left\{ (\Delta B_2^* - \Delta B_{21}^{u*})\rho^* + \frac{1}{2}\alpha(\Delta B_3^* - \Delta B_{31}^{u*})\rho^{*2} \right\} \tag{3.12}$$



(a) ϕ^u_{sim} calculated from eq. (3.11). (b) ϕ^u_{sim} calculated from eq. (3.14).

Figure 3.1: u-fraction ϕ^u_{sim} for the LJ fluid at subcritical temperatures calculated from the uv-theory (eq. (3.11)) and the uv-B_3-theory (eq. (3.14)) using simulation data for the Helmholtz energy from Stephan *et al.* [1]. Inputs of the uv-theory were obtained from van Westen and Gross, [11] whereas all inputs of the uv-B_3 theory are listed in the Appendix. The plus symbols on the insets mark data points within the meta-/unstable vapour region.

where $\Delta B^*_3 = B^*_3 - \tilde{B}^*_{30}$, $\tilde{B}^*_{30}$ defines the third virial coefficient of the WCA reference fluid predicted by our description of a_0 (see appendix A.1), ΔB^{u*}_{31} denotes the third virial coefficient obtained from the first-order perturbation contribution $\Delta B^{u*}_{31} = \lim_{\rho^*\to 0}\left(\partial^2 \Delta a^u_1/\partial \rho^{*2}\right)_{N,T^*}$, and α is an empirical damping function with two main purposes:

- α ensures that the virial term $\{\dots\}$ in eq. (3.12) is negative for all state points, and thus that the effective lower bound is always smaller than the upper bound. This empirical adjustment is necessary because $\Delta B^*_3 - \Delta B^{u*}_{31}$ becomes positive for elevated temperatures, while $\Delta B^*_2 - \Delta B^{u*}_{21}$ is always negative (a corollary of eq. (3.7) being a rigorous upper bound to the Helmholtz energy[4,34]).

- α is chosen such that the model recovers the full third virial coefficients at zero density but dampens the very large negative values of $(\Delta B^*_3 - \Delta B^{u*}_{31})$ at temperatures approaching zero for $\rho^* > 0$. This prevents large gradients and improves the qualitative behaviour of the model at very low temperatures, which is discussed in section 3.4.3.

We find that the following ansatz function meets these requirements

$$\alpha(\rho^*, T^*) = \exp\left(-\frac{c_0}{\nu}\left(1 + \frac{c_1}{T^{*2}}\right)\rho^*\right) \tag{3.13}$$

where c_0 and c_1 are empirical model constants.

Insertion of the new effective lower bound eq. (3.12) into the interpolation scheme eq. (3.6) leads to the new equation of state, which we refer to as *uv*-B_3-theory:

$$a = a_0 + \Delta a_1^u + (1-\phi^u)\left\{(\Delta B_2^* - \Delta B_{21}^{u*})\rho^* + \frac{1}{2}\alpha(\Delta B_3^* - \Delta B_{31}^{u*})\rho^{*2}\right\} \tag{3.14}$$

Algebraic expressions for the model inputs a_0, Δa_1^u, ΔB_{21}^{u*}, and ΔB_{31}^{u*} are presented in the appendix A.1. The expressions are the same as in the original *uv*-theory[11] with the exception of a_0, which was newly parametrized. Elliott *et al.* [18] proposed a semi-empirical series approximation for the second virial coefficient of LJ fluids. This model was recently slightly revised and extended to Mie fluids by van Westen[36], referred to as Revised-series-approximation (RSAP). We use the RSAP based on the WCA division of the Mie-potential to model ΔB_2. An algebraic expression for ΔB_3 is developed in the next section (section 3.3.4).

We define the *u*-fraction as

$$\phi^u(\rho^*, T^*) = 1 - \exp\left(-\left(c_2 + \frac{c_3}{c_4 + \nu}\right)\rho^{*2} - \omega\right) \tag{3.15}$$

including the low-temperature correction

$$\omega(\rho^*, T^*) = \exp\left(-c_5 T^* (\rho^* - c_6)^2\right)\left(\frac{1}{\tanh(c_7 T^*)} - 1\right)^2 \tag{3.16}$$

and model constants $c_2, \ldots, c_7 \geq 0$. The low-temperature correction ω is introduced to prevent the occurrence of multiple van der Waals loops of pressure isotherms in the meta-stable region for temperatures $T^* \leq 0.4$. The constants c_5, c_6 and c_7 are chosen such that ω is negligible for moderate and high temperatures and takes on a value that leads to a *u*-fraction of unity for $T^* \to 0$. As shown in section 3.4.3, this limiting behaviour of ω suppresses artifacts such as multiple (>2) zero-crossings of pressure isotherms.

The 5 empirical constants $c_0, \ldots, c_4$ describing the *u*-fraction of eq. (3.15) and the damping function α of eq. (3.13) were fitted simultaneously to simulation results for the internal energy[20,37], pressure[20,37], saturated vapour pressure,[9,38] and to critical points from the correlation by Werth *et al.* [38] of Mie fluids with $8 \leq \nu \leq 27$. A total of 1343 reference data points was used for the fit. The constants c_5, c_6 and c_7 describing the low temperature correction ω of eq. (3.16) were chosen afterwards such that ω is negligible for moderate and high temperatures and large for $T^* \to 0$. The correlated constants are given in table 3.1.

Table 3.1: Model constants for eqs. (3.13), (3.15), and (3.16)

c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7
26.454	1.8045	1.7997	161.96	11.605	12	0.4	2

In figure 3.1(b), we show the u-fraction ϕ^u_{sim}, calculated by inserting molecular simulation data[1] for the Helmholtz energy of a LJ fluid in eq. (3.14). As a result of the new effective lower bound, almost all values of ϕ^u_{sim} are positive. The uv-B_3-theory is therefore capable to reproduce the gas phase behaviour of the LJ fluid at low temperatures with a u-fraction $0 \leq \phi^u \leq 1$.

The new parametrization for a_0 leads to a better description of the Helmholtz energy in the high density region but causes unphysical behaviour when extrapolated to extremely high packing fractions. The maximal density at which the model doesn't show these unphysical artifacts can be roughly estimated by

$$\rho^*_{\text{max}}(T^*) \approx 1.1 + 0.12\sqrt{T^*} \tag{3.17}$$

We restrict the applicability of the uv-B_3-theory to densities $\rho \leq \rho^*_{\text{max}}$. Figure 3.3 visualizes the maximum density $\rho^*_{\text{max}}(T^*)$. For temperatures $T^* \leq 10$, ρ^*_{max} specifies densities that are significantly higher than the saturated liquid densities.

3.3.4 Algebraic Third Virial Coefficient B_3 for Mie ν-6 Fluids

We develop an algebraic expression for the third virial coefficient of the Mie ν-6 fluid based on the work of Elliott *et al.* [18] and van Westen [36]. Elliott, Schultz, and Kofke proposed the following expansion of the perturbation contribution to the dimensionless third virial coefficient in terms of the inverse dimensionless temperature $\beta^* = 1/T^*$,

$$B^*_3 = B^*_{30} + \sum_{i=1}^{\infty} B^*_{3i} \beta^{*i} \tag{3.18}$$

with the series coefficients $B^*_{3i} = (1/i!)(\partial^i \Delta B^*_3 / \partial \beta^{*i})_{\beta^*=0}$ and $B^*_3 = B_3/\sigma^6$.

For a LJ potential divided according to the scheme of WCA, Elliott, Schultz and Kofke showed that the third virial coefficient of the reference fluid can be described quite accurately by the following approximate mapping onto the third virial coefficient of hard spheres of diameter

$q^* = q/\sigma$,

$$B_{30}^* = B_3^{\mathrm{hs},q*} = 10\left(\frac{\pi}{6}q^{*3}\right)^2 \tag{3.19}$$

where $B_3^{\mathrm{hs},q}$ is the third virial coefficient of hard spheres, and $q(T^*)$ defines the hard-sphere diameter that exactly maps the second virial coefficient of the WCA reference fluid onto hard spheres.[36] Here, we apply the mapping of Eq.(3.19) to Mie fluids. Details on the calculation of $q(T^*)$ can be found in the appendix A.1.

For describing higher order coefficients, Elliott *et al.*[18] proposed the empirical ansatz function

$$B_{3i}^*(\nu = 12) = k_{0i} + \frac{k_{1i}}{(k_{2i} + T^*)^{k_{3i}}} \tag{3.20}$$

where k_{ni} are model constants. Eq. (3.20) was shown to provide a good description of B_{3i} obtained from Mayer-sampling Monte Carlo simulations.

We choose to correlate all model constants k_{ni} simultaneously, to simulation data for B_3^*. Though this approach is less rigorous as compared to that of Elliott *et al.*[18], we show below that it leads to a very accurate model for B_3. We truncate eq. (3.18) at fourth order, leading to

$$B_3^* \approx B_{30}^* + B_{31}^*\beta^* + B_{32}^*\beta^{*2} + B_{33}^*\beta^{*3} + B_{34}^*\beta^{*4} \tag{3.21}$$

For the LJ fluid we fitted the 16 model constants to 388 data points with $0.39 \leq T^* \leq 1000$ for B_3^* from Stephan *et al.*[1]. The correlated model constants are provided in table 3.2. We extend the approach to different repulsive Mie exponents by defining

$$B_{3i}^*(\nu) = k_{0i} + \frac{\tilde{k}_{1i}(\nu)}{(k_{2i} + T^*)^{\tilde{k}_{3i}(\nu)}} \tag{3.22}$$

with the ν-dependent coefficients

$$\tilde{k}_{1i}(\nu) = k_{1i}\left(1 + p_i\frac{\nu - 12}{\nu - l_i}\right) \tag{3.23}$$

and

$$\tilde{k}_{3i}(\nu) = k_{3i}\left(1 + m_i\frac{\nu - 12}{\nu - 6}\right) \tag{3.24}$$

The 12 additional coefficients p_i, l_i and m_i were fitted to results for B_3^* calculated numerically via fast Fourier transforms (FFT) using the webapp of Schultz[39] with $10 \leq \nu \leq 24$ and $0.3 < T^* \leq 10$. The correlated constants are provided in table 3.3.

In Figure 3.2 the results of eqs. (3.19)–(3.24) are compared to reference data for B_3^* from FFT for different repulsive exponents. Our correlation is in very good agreement with the FFT data within the entire range of inverse temperature that is shown. The absolute deviations of our model from the reference data are all below 0.2. Additional analysis on eqs. (3.19)–(3.24) is provided in the appendix A.1.

Table 3.2: Model constants k_{ji} for the series elements $B_{31}, \ldots, B_{34}$ of the LJ fluid as defined in eq. (3.20).

i	k_{0i}	k_{1i}	k_{2i}	k_{3i}
1	-3.9806	79.565	0.5489	5.3632
2	1.4245	57.292	0.0	1.0031
3	-39.755	-81.213	0.6987	30.156
4	-23.692	-85.006	0.7762	12.798

3.4 Results

3.4.1 Lennard-Jones Fluid

Stephan *et al.* [1] recently published a comprehensive and detailed review of the performance of 20 LJ EoS. The authors compare the results of the EoS with data from molecular simulations, which is collected in the database of Stephan *et al.* [37]. They conclude that the EoS of Thol *et al.* [20], Kolafa and Nezbeda [17], Gottschalk [23] and Mecke *et al.* [21,22] are overall the most accurate. The equation of state of Thol *et al.* [20] is fully empirical whereas the other three models contain both physically based and empirical parts. The *uv*-theory [11] - and the related *uf*-theory [9] were not considered in the review of Stephan *et al.* [1].
In the following, we compare the performance of the *uv*-B_3-theory (*uv*-B_3) to that of the

Table 3.3: Model constants $\tilde{k}_{ji}$ for the series elements $B_{31}, \ldots, B_{34}$ of Mie v-6 fluids as defined in eqs. (3.22)–(3.24)

i	p_i	l_i	m_i
1	0.80844	4.9485	0.11853
2	-0.095409	-21.300	0.078556
3	0.47525	7.0000	-0.55039
4	-2.8328	3.2162	0.0091630

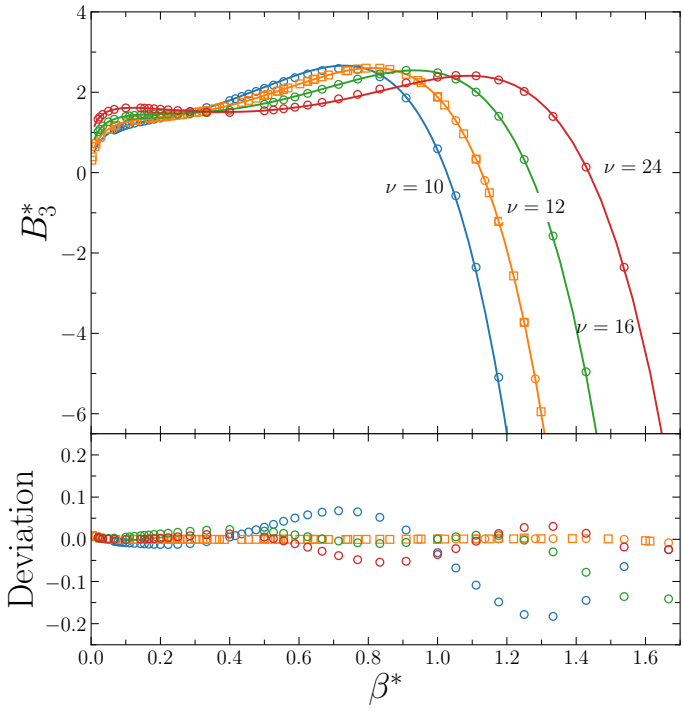


Figure 3.2: Third virial coefficient B_3^* for various repulsive exponents ν as a function of the inverse temperature β^*. Comparison of our model, eq. (3.21) (lines), with reference data obtained by numerical integration of B_3 using FFT[39] (circles) and from Stephan *et al.*[1] (squares). Deviations from the reference data are defined as $B_3^{*\mathrm{Model}}(\beta^*) - B_3^{*\mathrm{Ref}}(\beta^*)$.

original *uv*-theory[11] (*uv*), and the EoS of Kolafa and Nezbeda[17] (Ko & Ne) and Thol *et al.*[20] (Thol). First, the vapour-liquid equilibrium (VLE) is considered. Second, the performance of the EoS in homogeneous fluid states is assessed using the same systematic procedure as was used by Stephan *et al.*[1]. Therefore the LJ phase diagram is divided into different fluid regions, which is shown graphically in Figure 3.3. For homogeneous fluid states, we consider the vapour phase (V), the liquid phase (L), the high density liquid phase (HD-L), the meta-stable or unstable region (MU), the critical region (C), the supercritical region (Su) and the high density super critical region (HD-Su). The extreme-temperature region (Ex-T) with $T^* > 6$ is considered separately.

The performance of each of the aforementioned EoS in describing a thermodynamic property j within a fluid-phase region k is quantified by the mean absolute relative deviation (MARD)

with respect to simulation data

$$\mathrm{MARD}_{jk} = \frac{1}{N_{jk}} \sum_{i=1}^{N_{jk}} \left| \frac{Y_{jk,i}^{\mathrm{EoS}} - Y_{jk,i}^{\mathrm{Sim}}}{Y_{jk,i}^{\mathrm{Sim}}} \right| \cdot 100\% \tag{3.25}$$

where N_{jk} is the number of simulation data points for property j in region k, $Y_{jk,i}^{\mathrm{Sim}}$ is one simulation result and $Y_{jk,i}^{\mathrm{EoS}}$ is the corresponding result of the EoS. The overall performance of the EoS for a certain property j is quantified by the mean absolute relative deviation,

$$\mathrm{MARD}_{j} = \frac{1}{N_{j}} \sum_{i=1}^{N_{j}} \sum_{k=1}^{N_{K}} \left| \frac{Y_{jk,i}^{\mathrm{EoS}} - Y_{jk,i}^{\mathrm{Sim}}}{Y_{jk,i}^{\mathrm{Sim}}} \right| \cdot 100\% \tag{3.26}$$

where $N_j = \sum_{k=1}^{N_K} N_{jk}$ is the total number of data points for property j and N_K is the total number of fluid regions.

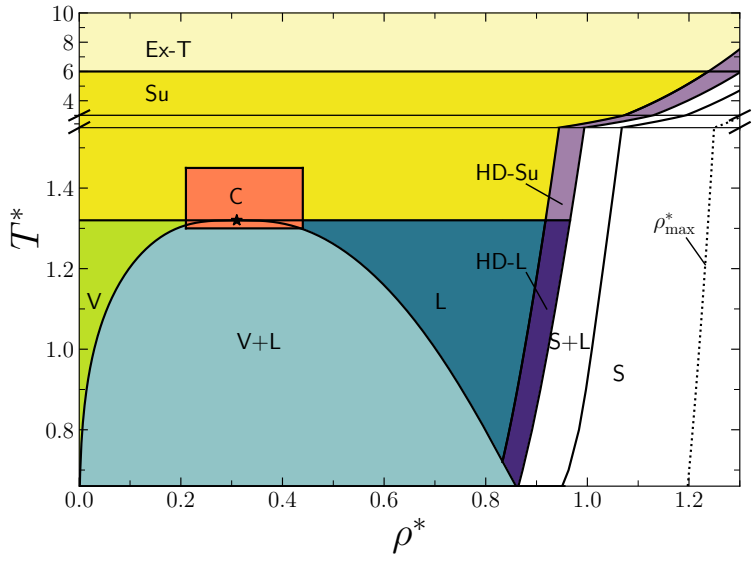


Figure 3.3: Division of the LJ phase diagram into different regions as defined by Stephan *et al.* [1].The dotted line indicates the maximal density of eq. (3.17).

3.4.1.1 Vapour-Liquid Equilibrium of the LJ Fluid

In figure 3.4, we compare the percentage deviations of results from three equation-of-state models for several VLE properties with respect to data from molecular simulations. The *uv*-B_3-theory leads to a very accurate description, and is of comparable accuracy as the empirical models. The simulation data for the liquid density ρ^{l}, the vapour density ρ^{v} and the saturated

vapour pressure p^s is from van Westen and Gross[9], the data for the enthalpy of vaporization $\Delta^{lv}h$ is from Stephan *et al.*[1]. The overall MARD values for each property are summarized in table 3.4.

Potoff and Panagiotopoulos[40] performed Monte Carlo simulations with histogram reweighting methods in the grand canonical ensemble to obtain liquid-vapour coexistence curves and critical points of the LJ fluid. They corrected the critical data for finite-size effects using mixed-field finite-size scaling analysis and concluded that the critical point of the LJ fluid can be estimated to be at the critical density $\rho_c^* = 0.316 \pm 0.001$, the critical temperature $T^* = 1.3120 \pm 0.0007$ and the critical pressure $p_c^* = 0.1279 \pm 0.0006$. The critical point obtained from the *uv*-B_3-theory lies at $\rho_c^* = 0.31425$, $T_c^* = 1.3133$ and $p_c^* = 0.12860$. The absolute relative deviations to the simulation results are 0.55% for the critical density, 0.10% for the critical temperature and 0.55% for the critical pressure.

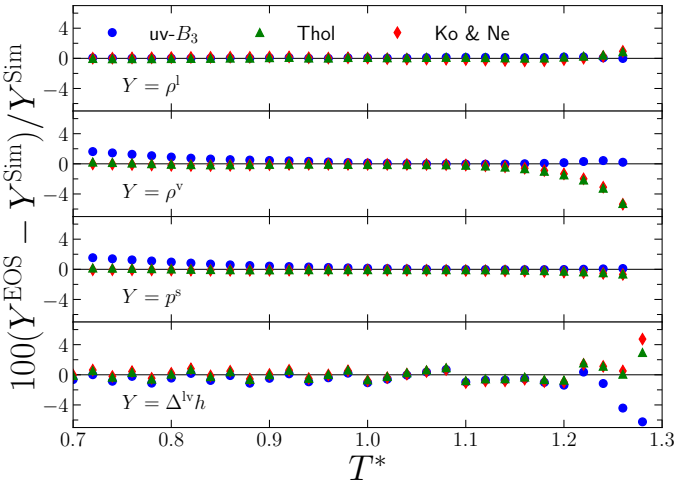


Figure 3.4: Deviations of predicted VLE properties of the LJ fluid with respect to molecular simulation results of van Westen and Gross[9] and Stephan *et al.*[37]. Results are shown for the *uv*-B_3-theory (circles) and the EoS of Thol *et al.*[20] (triangles) and Kolafa and Nezbeda[17] (diamonds).

3.4.1.2 Homogeneous Fluid States of the LJ Fluid

We use the same reference data as was used in the review of Stephan *et al.*[1]. As shown in table 3.5, the *uv*-B_3-theory performs very well within all considered fluid regions. All considered properties are described with an overall percentage error that is of the same order of magnitude as that of the more empirical models of Thol *et al.*[20] and Kolafa and Nezbeda[17]. Only the Helmholtz energy in the supercritical and high-density supercritical

Table 3.4: LJ fluid: Percentage deviations (MARD, eq. (3.26)) of results obtained from the *uv*-B_3-theory (*uv*-B_3), the original *uv*-theory[11] (*uv*), the equation of state of Thol *et al.*[20] (Thol), and the equation of state of Kolafa and Nezbeda[17] (Ko & Ne) with respect to molecular simulation data for several VLE properties. The column # lists the number of data points.

			MARD / %			
	Ref.	#	*uv*-B_3	*uv*	Thol	Ko & Ne
Saturated vapour pressure p^s	9	28	0.4	1.0	0.1	0.2
Saturated liquid density ρ^l	9	28	0.1	0.6	0.1	0.2
Saturated vapour density ρ^v	9	28	0.4	1.2	0.6	0.6
Enthalpy of vaporization $\Delta^{lv}h$	37	31	0.9	0.9	0.6	0.7

regions and the chemical potential are described slightly less accurately by the *uv*-B_3-theory. The largest differences occur for the chemical potential in the high-density regions. These higher percentage errors occur at densities where the chemical potential crosses zero. The *uv*-B_3-theory describes all properties in the gas phase more accurately than the original *uv*-theory. This improvement can be attributed to the incorporation of the third virial coefficient and the more accurate model for the second virial coefficient.

The high-temperature ($T^* > 6$) region is analysed in table 3.6. In this region of high temperature, the *uv*-B_3-theory provides the most accurate description of thermodynamic properties, followed by the *uv*-theory.

Table 3.5: Percentage deviations (MARD, eqs. (3.25) and (3.26)) of results obtained from the *uv*-B_3-theory (*uv*-B_3), the original *uv*-theory[11] (*uv*), the equation of state of Thol *et al.*[20] (Thol), and the equation of state of Kolafa and Nezbeda[17] (Ko & Ne) with respect to molecular simulation data for several thermodynamic properties in different fluid-phase regions (cf.figure 3.3), rounded to one digit. The extreme temperature region ($T^* > 6$) is considered separately, in table 3.6. The column # denotes the number of data points.

		MARD / %						MARD / %			
	#	*uv*-B_3	*uv*	Thol	Ko & Ne		#	*uv*-B_3	*uv*	Thol	Ko & Ne
Helmholtz energy a^{res}						Thermal expansion coefficient α					
V	95	0.1	0.5	0.1	0.1	V	86	0.3	0.8	0.3	0.3
L	136	0.4	0.6	0.3	0.3	L	126	2.5	2.4	2.2	2.2
HD-L	49	1.1	1.1	0.8	0.8	HD-L	44	3.5	2.2	1.7	1.3
MU	155	0.2	0.6	0.8	0.2	MU	7	35.5	33.8	34.0	34.5
C	34	0.3	2.0	0.1	0.1	C	84	29.2	31.9	35.8	33.6
Su	224	1.4	2.8	0.5	0.7	Su	525	3.1	3.8	3.1	3.1

Table 3.5 *(continued)*

		MARD / %						MARD / %			
	#	*uv*-B_3	*uv*	Thol	Ko & Ne		#	*uv*-B_3	*uv*	Thol	Ko & Ne
HD-Su	25	3.5	3.6	2.1	2.1	HD-Su	28	1.1	5.4	1.1	1.8
Overall	718	0.8	1.5	0.5	0.5	**Overall**	900	5.4	6.1	5.9	5.7
Pressure p						Isothermal compressibility β					
V	169	0.1	0.3	0.1	0.1	V	92	0.3	0.7	0.3	0.2
L	537	3.5	6.5	2.9	3.4	L	159	3.6	3.6	2.3	2.6
HD-L	233	4.7	5.0	3.2	3.4	HD-L	46	3.6	3.1	1.0	0.8
MU	495	6.0	12.7	528.7	7.1	MU	23	12.5	12.6	18.2	12.7
C	225	3.0	3.2	2.3	2.1	C	108	34.1	36.5	44.8	38.1
Su	1715	0.9	1.1	0.5	0.5	Su	935	2.9	3.4	2.6	2.6
HD-Su	152	0.9	0.7	0.2	0.2	HD-Su	43	0.9	2.6	1.1	1.0
Overall	3526	2.4	3.9	75.3	2.1	**Overall**	1406	5.3	5.9	5.8	5.2
Internal energy u^{res}						Thermal pressure coefficient γ					
V	174	0.3	2.3	0.3	0.5	V	113	0.2	1.1	0.1	0.2
L	438	0.5	0.3	0.1	0.1	L	177	1.4	1.7	0.5	0.6
HD-L	212	0.6	0.5	0.3	0.3	HD-L	52	0.8	1.3	0.6	0.6
MU	307	1.0	2.4	0.5	0.6	MU	50	2.1	3.8	3.9	2.0
C	185	1.1	1.2	1.0	1.0	C	131	1.0	2.3	2.8	1.6
Su	1492	0.5	0.9	0.2	0.3	Su	1111	0.9	0.9	0.4	0.4
HD-Su	139	0.5	3.6	0.3	0.9	HD-Su	54	0.5	2.6	0.4	1.2
Overall	2947	0.6	1.2	0.3	0.4	**Overall**	1688	0.9	1.3	0.7	0.6
Isochoric Heat capacity c_v						Joule-Thomson coefficient μ_{JT}					
V	151	1.4	3.7	0.4	1.1	V	100	0.7	2.1	0.5	0.6
L	205	0.7	0.8	0.4	0.4	L	138	2.9	3.3	2.8	3.3
HD-L	57	0.8	0.6	0.5	0.7	HD-L	45	1.8	1.2	1.3	1.1
MU	113	5.1	7.3	16.6	4.6	MU	7	3.1	7.1	90.3	4.5
C	152	10.9	16.9	42.9	10.1	C	126	5.5	7.2	7.0	5.5
Su	1160	0.7	1.2	0.4	0.4	Su	898	4.9	5.3	4.5	4.5
HD-Su	53	0.4	0.8	0.3	0.6	HD-Su	44	0.9	3.7	0.9	1.6
Overall	1891	1.8	3.0	4.8	1.5	**Overall**	1358	4.2	4.8	4.5	4.0
Isobaric Heat capacity c_p						Grüneisen parameter Γ					
V	92	0.6	2.0	0.3	0.5	V	90	0.6	1.6	0.1	0.5
L	156	1.6	1.6	1.4	1.4	L	154	1.1	2.1	0.5	0.7

Table 3.5 *(continued)*

		MARD / %						MARD / %		
	#	*uv*-B_3	*uv*	Thol	Ko & Ne	#	*uv*-B_3	*uv*	Thol	Ko & Ne
HD-L	45	0.9	0.7	0.9	0.7	HD-L	43	0.4	0.9	0.4	0.4
MU	6	27.0	30.5	68.7	29.2	MU	15	5.5	7.7	7.0	5.4
C	101	27.0	29.5	30.2	29.9	C	39	8.5	17.3	4.4	8.0
Su	934	2.0	2.4	2.0	2.0	Su	419	1.0	1.4	0.3	0.4
HD-Su	43	0.7	2.7	0.6	1.2	HD-Su	36	0.3	2.2	0.3	0.9
Overall	1377	3.8	4.4	4.1	4.0	**Overall**	796	1.4	2.5	0.7	0.9
Chemical potential μ						Speed of sound w					
V	105	0.1	0.7	0.1	0.2	V	117	0.1	0.1	0.1	0.1
L	205	4.4	4.0	1.4	1.5	L	182	1.2	1.6	0.3	0.4
HD-L	103	9.6	8.2	2.7	2.9	HD-L	50	1.0	1.2	0.1	0.2
MU	153	0.3	0.9	0.4	0.3	MU	49	1.5	2.0	1.5	1.4
C	39	0.4	1.2	0.2	0.3	C	131	6.6	11.7	8.9	5.5
Su	206	1.8	2.3	0.2	0.3	Su	1111	0.8	1.1	0.4	0.4
HD-Su	26	1.2	1.1	0.4	0.3	HD-Su	43	0.3	0.5	0.2	0.3
Overall	837	2.9	2.9	0.8	0.9	**Overall**	1683	1.3	1.9	1.1	0.8

Table 3.6: Same as table 3.5 but for the high temperature region with $T^* > 6.0$ and $\rho^* \leq \rho^*_{\mathrm{max}}(T^*)$.

		MARD / %			
	#	*uv*-B_3	*uv*	Thol	Ko & Ne
Helmholtz energy a^{res}	212	0.8	0.8	0.6	0.8
Pressure p	527	5.1	4.7	5.6	10.8
Internal energy u^{res}	510	5.1	5.2	4.3	7.8
Isochoric Heat capacity c_v	274	0.5	1.1	0.3	0.8
Isobaric Heat capacity c_p	249	0.9	1.8	0.5	1.7
Chemical potential μ	212	0.8	0.8	1.1	1.7
Thermal expansion coefficient α	238	1.8	3.3	2.0	5.9
Isothermal compressibility β	249	1.2	1.8	2.9	7.8
Thermal pressure coefficient γ	270	1.1	2.0	1.0	2.8
Joule-Thomson coefficient μ_{JT}	249	2.0	3.1	1.4	3.2
Grüneisen parameter Γ	275	0.7	1.2	1.2	2.2
Speed of sound w	270	0.4	0.5	1.2	2.2

3.4.2 Mie Fluids

As shown in figures 3.5 and 3.6, the *uv*-B_3-theory leads to a highly accurate description of the VLE of Mie fluids. Results are compared to those obtained based on the SAFT-VR-Mie equation of state[30], which is a third-order Barker-Henderson[41] perturbation theory with a rigorously calculated first-order perturbation term, and effective second and third-order perturbation terms that were correlated to molecular simulation results for vapour-liquid equilibria of Mie fluids. Though the performance of both models is very similar, the *uv*-B_3 theory proves slightly more accurate, especially for the more repulsive (larger ν) Mie fluids.

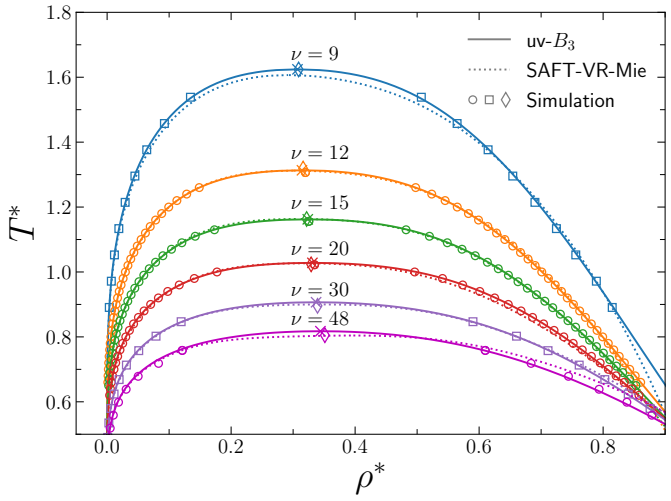


Figure 3.5: VLE of Mie ν-6 fluids obtained with the *uv*-B_3-theory (solid lines), the SAFT-VR-Mie equation of state[30] (dashed lines) and molecular simulations from van Westen and Gross[9] (circles) and Werth *et al.*[38] (squares). Critical points were calculated with the correlation of Werth *et al.*[38] (diamonds) and the *uv*-B_3-theory (crosses).

In figure 3.7, we show the isochoric heat capacity and the pressure for two super-critical temperatures and repulsive Mie exponents $\nu = \{12, 18, 25\}$. For the higher temperature, $T^* = 2.0$, both EoS describe the simulation results very well. Larger deviations with respect to simulations can be observed for $T^* = 1.4$. Especially for $\nu = 12$, where $T^* = 1.4$ is close to the critical temperature both EoS underestimate the heat capacity. Nevertheless, the *uv*-B_3-theory is considerably more accurate, and qualitatively capable to reproduce the local maximum.

3.4.3 Low Temperature Behaviour

In figure 3.8, we compare equation-of-state results for three isotherms in a pressure-density diagram to simulation data from Stephan *et al.*[1]. The diagram shows isotherms for sub-critical

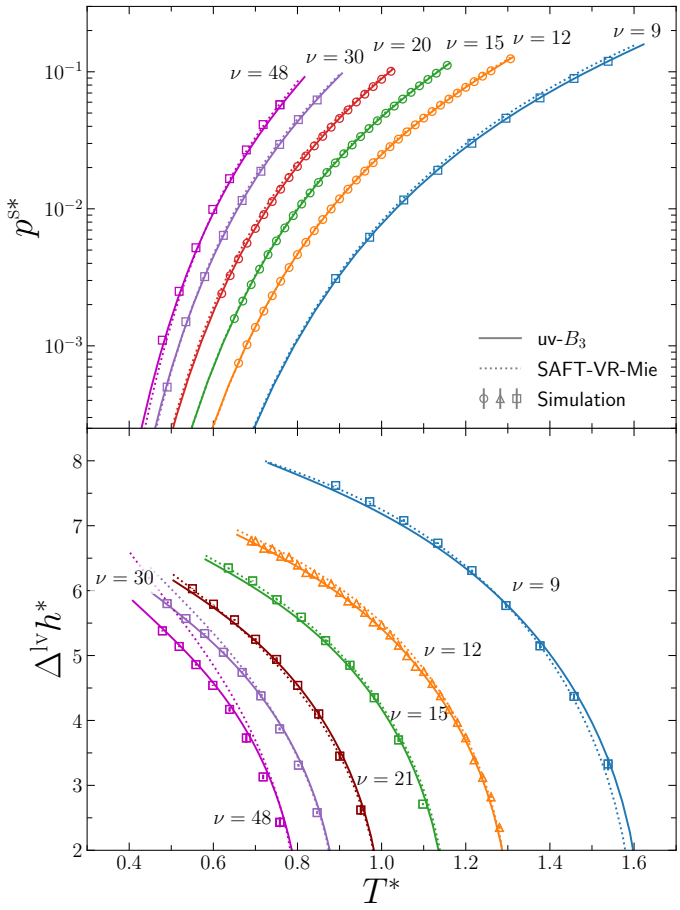


Figure 3.6: Saturated vapour pressure and enthalpy of vaporization of repulsive exponents $9 \leq \nu \leq 48$ obtained with the *uv*-B_3-theory (solid lines), the SAFT-VR-Mie equation of state[30] (dashed lines) and molecular simulations from van Westen and Gross[9] (circles), Werth *et al.*[38] (squares) and Stephan *et al.*[1] (triangles).

temperatures, including unstable regions of density. The lowest temperature is below the triple-point temperature, so that at high densities a meta-stable fluid is regarded. For many (semi-)empirical LJ EoS, an oscillating behaviour of the pressure can be observed in the unstable and meta-stable regimes[1]. An example thereof is the equation of state of Thol *et al.*[20] (dashed line), which shows strong oscillations and multiple (>2) zero-crossings of the pressure for all three temperatures in figure 3.8. This behaviour is undesired because (1) it can lead to small "false" stable regions (where $(\partial p/\partial \rho)_T > 0$) of positive pressure within the unstable region (false roots), which is detrimental for phase-equilibrium solvers, and (2) an incorrect shape of the isotherms in the meta-stable region leads to problems when calculating vapour-liquid interfacial properties from density functional theory[42] (DFT) or

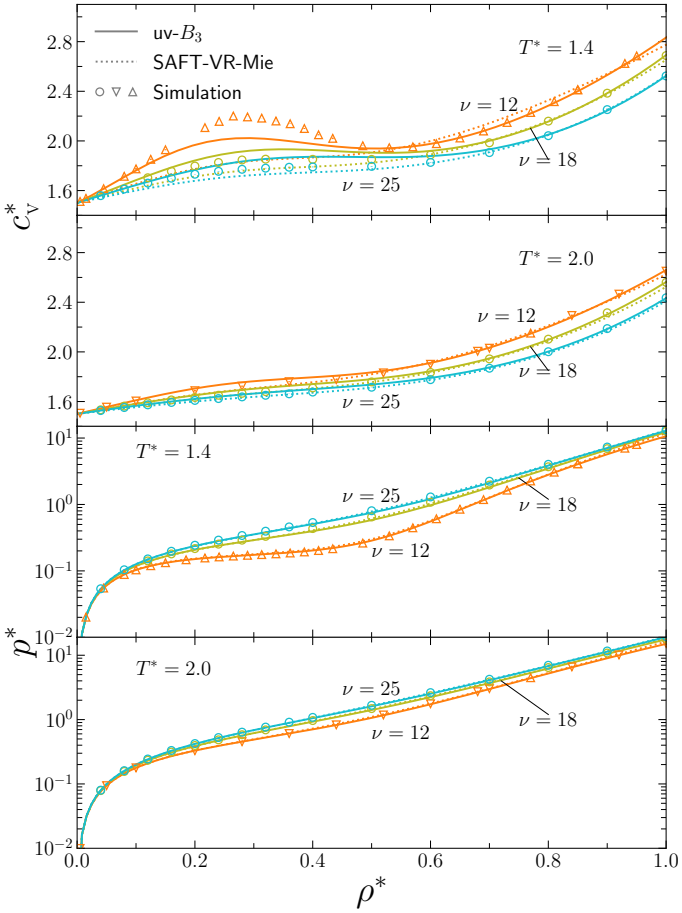


Figure 3.7: Isotherms of the isochoric heat capacity and the pressure at super-critical temperatures $T^* = 1.4$ and $T^* = 2.0$. Comparison of results of the *uv*-B_3-theory (solid lines), the SAFT-VR-Mie equation of state[30] (dotted lines) and molecular simulations by Stephan *et al.*[37] (triangles), Thol *et al.*[20] (triangles upside-down) and van Westen and Gross[9] (circles).

density gradient theory[43] (DGT). We consider EoS as well-behaved, if (1) the number of zero-crossings of the pressure isotherms is less than three, (2) the number of zero-crossings of $(\partial p/\partial \rho)_T$ is less than three, and (3) the number of zero-crossings of $(\partial^2 p/\partial \rho^2)_T$ is less than two. According to this, the equation of Kolafa and Nezbeda[17] (dotted line) is well-behaved for temperatures $T^* \geq 0.8$ and the *uv*-B_3-theory is well-behaved for temperatures down to $T^* = 0.6$. Moreover, the *uv*-B_3-theory fulfils the first two conditions down to temperatures of $T^* = 0.3$.

As mentioned in section 3.3, we applied a low-temperature correction $\omega(\rho^*, T^*)$, eq. (3.16), to the *u*-fraction in order to enforce a better qualitative description of low-temperature

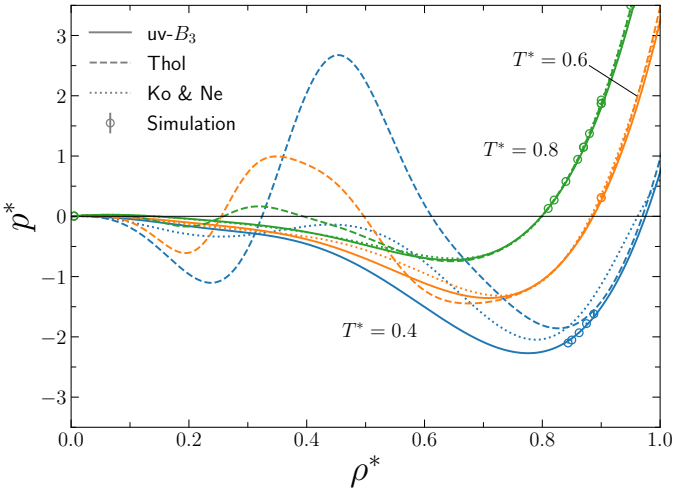


Figure 3.8: Pressure-density results for isotherms obtained from the *uv*-B_3-theory (solid line), the EoS of Thol *et al.* [20] (dashed line) and the EoS of Kolafa and Nezbeda [17] (dotted line) and simulation data (circles) from Stephan *et al.* [1] for the LJ fluid at subcritical temperatures $T^* = [0.4, 0.6, 0.8]$.

isotherms. To appreciate the benefits (and limitations) of this, we show pressure isotherms and the corresponding curves of $\omega(\rho^*, T^*)$ for $T^* = [0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]$ in figure 3.9. The correction ω is negligibly small for moderate and high temperatures ($T^* \geq 0.8$) and increases steeply for temperatures approaching zero. The maximum of ω at a density $\rho^* = 0.4$ corrects for the tendency of the pressure from the *uv*-B_3-theory to oscillate close to this density. For large values of ω (at $T^* \to 0$) the *u*-fraction approaches unity and the fluid is described solely based on the (well-behaved) first-order *u*-expansion. A side-effect of the strong increase of ω at low temperature is a crossing of isotherms in the unstable vapour region at negative pressures for temperatures $T^* < 0.3$. Apparently our chosen functional form for ω combined with the *uv*-B_3 framework is not capable (yet) to provide an equation of state that is completely free of artifacts. Although these side-effects are unwanted, we emphasize they are not detrimental to the application of the equation-of-state model in general.

In summary, we evaluated the *uv*-B_3 theory for temperatures down to $T^* = 0.01$ (figure 3.9 and figure 3.10). The empirical low-temperature correction ω ensures a single Van-der-Waals loop of the pressure isotherms, even at these extremely low temperatures. However, for temperatures $T^* \leq 0.3$ and negative pressures, crossings of isotherms occur in the unstable vapour region, which is why we suggest $T^* > 0.3$ as lower temperature limit for the model.

The discussion above, on the effect of the low-temperature correction ω for the LJ fluid is transferable to other Mie ν-6 fluids. This is visualized in figure 3.10, where we show results

of the *uv*-B_3 theory for the pressure isotherms of Mie fluids of repulsive exponents $\nu = 10$ and $\nu = 30$ at temperatures $0.01 \leq T^* \leq 0.8$. The shape of the isotherms is similar to the shape of the isotherms of the LJ fluid in figure 3.9.

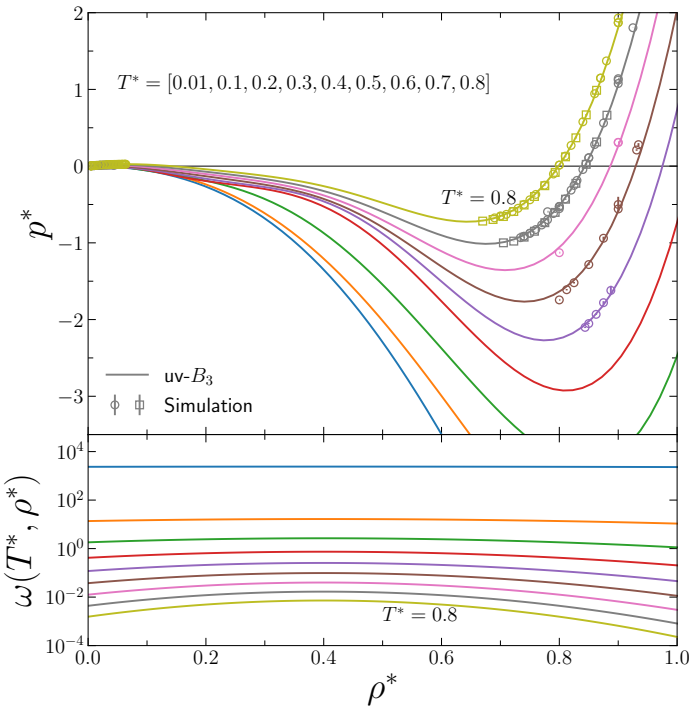


Figure 3.9: Isotherms of the pressure calculated based on the *uv*-B_3-theory, simulation data from Stephan *et al.* [1] (circles) and the NIST Standard Reference Database[44] (squares) and isotherms of the low temperature correction ω for the LJ fluid at temperatures $T^* = [0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]$.

3.5 Conclusion and Outlook

The *uv*-B_3 theory for Mie ν-6 fluids developed in this work combines the quantitative accuracy of fully empirical thermodynamic models with the benefits of a physically based approach:

- The repulsive exponent of the Mie potential ν is a generic model parameter that can be varied between $9 \leq \nu \leq 48$. The scope of our model is therefore wider than that of pure LJ EoS.
- Being a first-order perturbation theory, the *uv*-B_3-theory maintains a direct connection to molecular distribution functions. This will facilitate the application to inhomogeneous

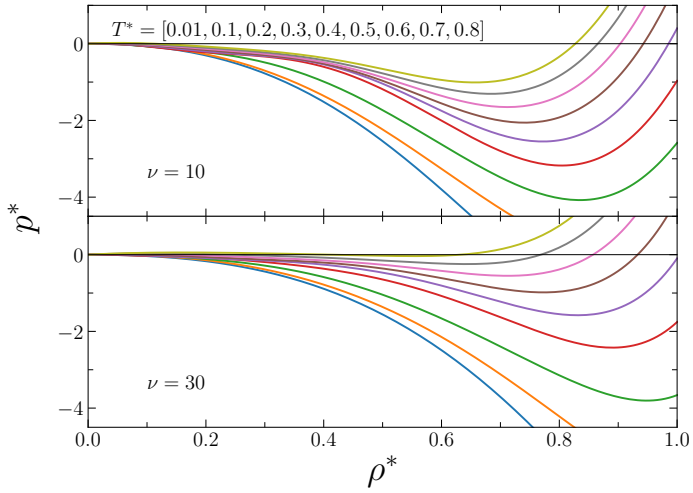


Figure 3.10: Pressure isotherms of Mie ν-6 fluids calculated based on the *uv*-B_3-theory with repulsive exponents $\nu = 10$ and $\nu = 30$ at temperatures $T^* = [0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]$.

systems via density functional theory (DFT) or density gradient theory (DGT).

- The extension of the model to mixtures is relatively straightforward. Most of the required equations are already provided by van Westen and Gross[8].
- The extension of the model to non-spherical (chain) fluids is easier than for more empirical models. The original *uv*-theory has already been applied to LJ dimers with considerable success[11].
- By incorporating a low temperature correction in the model, we prevent common numerical artifacts in the meta-stable and unstable regions. The model shows no false roots or multiple (>2) zero-crossings of the pressure isotherms and can even be used for extrapolations to temperatures far below the triple point. We suggest $T^* > 0.3$ as lower temperature limit for the model.

For now, the applicability of the model is restricted to densities below some maximum value. In ongoing work we will develop a more robust description of Helmholtz energy of the WCA reference fluid to eliminate this restriction.

References

[1] S. Stephan, J. Staubach, and H. Hasse. Review and comparison of equations of state for the Lennard-Jones fluid. *Fluid Phase Equilibr.*, **523**:112772, 2020. doi:10.1016/j.fluid.2020.112772.

[2] M. L. Huber, E. W. Lemmon, I. H. Bell, and M. O. McLinden. The NIST REFPROP Database for Highly Accurate Properties of Industrially Important Fluids. *Ind. Eng. Chem. Res.*, **61**(42):15449–15472, 2022. doi:10.1021/acs.iecr.2c01427.

[3] C. G. Gray and K. E. Gubbins. *Theory of Molecular Fluids*. Oxford University Press, 1984. ISBN 9780198556022, 9780191919251. doi:10.1093/oso/9780198556022.001.0001.

[4] J.-P. Hansen and I. R. McDonald. *Theory of Simple Liquids*. Elsevier, 2013. ISBN 9780123870322. doi:10.1016/c2010-0-66723-x.

[5] S. Zhou and J. R. Solana. Progress in the Perturbation Approach in Fluid and Fluid-Related Theories. *Chem. Rev.*, **109**(6):2829–2858, 2009. doi:10.1021/cr900094p.

[6] K. E. Gubbins. The theory of non-electrolyte solutions: An historical review. *Mol. Phys.*, **111**(24):3666–3697, 2013. doi:10.1080/00268976.2013.831140.

[7] K. E. Gubbins. Perturbation theories of the thermodynamics of polar and associating liquids: A historical perspective. *Fluid Phase Equilibr.*, **416**:3–17, 2016. doi:10.1016/j.fluid.2015.12.043.

[8] T. van Westen and J. Gross. A critical evaluation of perturbation theories by Monte Carlo simulation of the first four perturbation terms in a Helmholtz energy expansion for the Lennard-Jones fluid. *J. Chem. Phys.*, **147**(1):014503, 2017. doi:10.1063/1.4991008.

[9] T. van Westen and J. Gross. Accurate first-order perturbation theory for fluids: uf-theory. *J. Chem. Phys.*, **154**(4):041102, 2021. doi:10.1063/5.0031545.

[10] N. M. Alsaifi and J. R. Elliott. Avoiding Artifacts in Noncubic Equations of State. *Ind. Eng. Chem. Res.*, **61**(42):15661–15677, 2022. doi:10.1021/acs.iecr.2c01923.

[11] T. van Westen and J. Gross. Accurate thermodynamics of simple fluids and chain fluids based on first-order perturbation theory and second virial coefficients: uv-theory. *J. Chem. Phys.*, **155**(24):244501, 2021. doi:10.1063/5.0073572.

[12] H. C. Andersen and D. Chandler. Optimized cluster expansion for classical fluids. I. General theory and variational formulation of the mean spherical model and hard sphere Percus-Yevick equations. *J. Chem. Phys.*, **57**:1918–1929, 1972.

[13] H. C. Andersen, D. Chandler, and J. D. Weeks. Optimized cluster expansion for classical fluids. III. Applications to ionic solutions and simple liquids. *J. Chem. Phys.*, **57**:2626–2631, 1972.

[14] H. C. Andersen, D. Chandler, and J. D. Weeks. Roles of repulsive and attractive forces in liquids:

the equilibrium theory of classical fluids. *Advances in Chemical Physics*, **34**:105, 1976.

[15] R. L. Cotterman, B. J. Schwarz, and J. M. Prausnitz. Molecular thermodynamics for fluids at low and high densities. Part I: pure fluids containing small or large molecules. *AIChE J.*, **32**:1787–1798, 1986.

[16] Y. Song and E. A. Mason. Statistical-mechanical theory of a new analytical equation of state. *J. Chem. Phys.*, **91**:7840–7853, 1986.

[17] J. Kolafa and I. Nezbeda. The Lennard-Jones fluid: An accurate analytic and theoretically-based equation of state. *Fluid Phase Equilibr.*, **100**:1–34, 1994. doi:10.1016/0378-3812(94)80001-4.

[18] J. R. Elliott, A. J. Schultz, and D. A. Kofke. Combined temperature and density series for fluid-phase properties. II. Lennard-Jones spheres. *J. Chem. Phys.*, **151**(20):204501, 2019. doi:10.1063/1.5126281.

[19] T. van Westen, M. Hammer, B. Hafskjold, A. Aasen, J. Gross, and Ø. Wilhelmsen. Perturbation theories for fluids with short-ranged attractive forces: A case study of the Lennard-Jones spline fluid. *J. Chem. Phys.*, **156**(10):104504, 2022. doi:10.1063/5.0082690.

[20] M. Thol, G. Rutkai, A. Köster, R. Lustig, R. Span, and J. Vrabec. Equation of State for the Lennard-Jones Fluid. *J. Phys. Chem. Ref. Data*, **45**(2):023101, 2016. doi:10.1063/1.4945000.

[21] M. Mecke, A. Müller, J. Winkelmann, J. Vrabec, J. Fischer, R. Span, and W. Wagner. An accurate Van der Waals-type equation of state for the Lennard-Jones fluid. *Int. J. Thermophys.*, **17**(2):391–404, 1996. doi:10.1007/bf01443399.

[22] M. Mecke, A. Müller, J. Winkelmann, J. Vrabec, J. Fischer, R. Span, and W. Wagner. An accurate Van der Waals-type equation of state for the Lennard-Jones fluid. *Int. J. Thermophys.*, **17**(2):391–404, 1996. doi:10.1007/bf01443399.

[23] M. Gottschalk. An EOS for the Lennard-Jones fluid: A virial expansion approach. *AIP Adv.*, **9**(12):125206, 2019. doi:10.1063/1.5119761.

[24] T. Lafitte, D. Bessieres, M. M. Piñeiro, and J.-L. Daridon. Simultaneous estimation of phase behavior and second-derivative properties using the statistical associating fluid theory with variable range approach. *J. Chem. Phys.*, **124**(2):024509, 2006. doi:10.1063/1.2140276.

[25] T. Lafitte, M. M. Piñeiro, J.-L. Daridon, and D. Bessières. A Comprehensive Description of Chemical Association Effects on Second Derivative Properties of Alcohols through a SAFT-VR Approach. *J. Phys. Chem. B*, **111**(13):3447–3461, 2007. doi:10.1021/jp0682208.

[26] A. de Villiers, C. Schwarz, A. Burger, and G. Kontogeorgis. Evaluation of the PC-SAFT, SAFT and CPA equations of state in predicting derivative properties of selected non-polar and hydrogen-bonding compounds. *Fluid Phase Equilibr.*, **338**:1–15, 2013. doi:10.1016/j.fluid.2012.09.035.

[27] A. Hemmen and J. Gross. Transferable Anisotropic United-Atom Force Field Based on the Mie Potential for Phase Equilibrium Calculations: N-alkanes and n-Olefins. *J. Phys. Chem. B*, **119**(35):11695–11707, 2015. doi:10.1021/acs.jpcb.5b01354.

[28] R. Eisenschitz and F. London. Über das Verhältnis der van der Waalsschen Kräfte zu den homöpolaren Bindungskräften. *Z. Phys.*, **60**:491, 1930.

[29] A. Stone. *The theory of intermolecular forces*. Oxford University Press, Oxford, second edition, 2013.

[30] T. Lafitte, A. Apostolakou, C. Avendaño, A. Galindo, C. S. Adjiman, E. A. Müller, and G. Jackson. Accurate statistical associating fluid theory for chain molecules formed from Mie segments. *J. Chem. Phys.*, **139**(15):154504, 2013. doi:10.1063/1.4819786.

[31] D. Chandler and J. D. Weeks. Equilibrium Structure of Simple Liquids. *Phys. Rev. Lett.*, **25**(3):149–152, 1970. doi:10.1103/physrevlett.25.149.

[32] H. C. Andersen, J. D. Weeks, and D. Chandler. Relationship between the Hard-Sphere Fluid and Fluids with Realistic Repulsive Forces. *Phys. Rev. A*, **4**(4):1597–1607, 1971. doi:10.1103/physreva.4.1597.

[33] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.*, **54**(12):5237–5247, 1971. doi:10.1063/1.1674820.

[34] A. Isihara. The Gibbs-Bogoliubov inequality dagger. *J. Phys. A: Gen. Phys.*, **1**(5):539–548, 1968. doi:10.1088/0305-4470/1/5/305.

[35] R. W. Zwanzig. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.*, **22**(8):1420–1426, 1954.

[36] T. van Westen. Algebraic second virial coefficient of the Mie m— 6 intermolecular potential based on perturbation theory. *J. Chem. Phys.*, **154**(23):234502, 2021. doi:10.1063/5.0050659.

[37] S. Stephan, M. Thol, J. Vrabec, and H. Hasse. Thermophysical Properties of the Lennard-Jones Fluid: Database and Data Assessment. *J. Chem. Inf. Model.*, **59**(10):4248–4265, 2019. doi:10.1021/acs.jcim.9b00620.

[38] S. Werth, K. Stöbener, M. Horsch, and H. Hasse. Simultaneous description of bulk and interfacial properties of fluids by the Mie potential. *Mol. Phys.*, **115**(9-12):1017–1030, 2016. doi:10.1080/00268976.2016.1206218.

[39] A. J. Schultz. Reduced Second Virial Coefficient from the Lennard-Jones Potential. <https://www.etomica.org/apps/virial/b3>, 2005. doi:10.1017/cbo9780511815928.046.

[40] J. J. Potoff and A. Z. Panagiotopoulos. Critical point and phase behavior of the pure fluid and a

Lennard-Jones mixture. *J. Chem. Phys.*, **109**(24):10914–10920, 1998. doi:10.1063/1.477787.

[41] J. A. Barker and D. Henderson. Perturbation Theory and Equation of State for Fluids. II. A Successful Theory of Liquids. *J. Chem. Phys.*, **47**(11):4714–4721, 1967. doi:10.1063/1.1701689.

[42] R. Evans. The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids. *Adv. Phys.*, **28**(2):143–200, 1979. doi:10.1080/00018737900101365.

[43] J. W. Cahn and J. E. Hilliard. Free Energy of a Nonuniform System. I. Interfacial Free Energy. *J. Chem. Phys.*, **28**(2):258–267, 1958. doi:10.1063/1.1744102.

[44] D. Siderius. NIST Standard Reference Simulation Website - SRD 173. 2017. doi:10.18434/mds2-232.

4 Effect of Chain Conformation on the Free Energy of Dilute Polymer Solutions: Monte Carlo Simulations and Perturbation Theory for the Second Virial Coefficient of Lennard-Jones Chains

The content of this chapter is a literal quote of the publication:

A. Reimer, J. Gross and T. van Westen. Effect of Chain Conformation on the Free Energy of Dilute Polymer Solutions: Monte Carlo Simulations and Perturbation Theory for the Second Virial Coefficient of Lennard–Jones Chains. *Macromolecules*, **58**, 3, 1498–1511 (2025), doi:10.1021/acs.macromol.4c02827

4.1 Abstract

The free energy of chain molecules in solution, and therefore polymer-solvent phase equilibria, are generally believed to be strongly connected to changes in chain conformation. In this [chapter], we employ Monte Carlo simulations to analyse this connection. Specifically, we calculate the osmotic second virial coefficient B_2 and several single-chain properties for 3-dimensional, off-lattice chains comprising up to 256 segments interacting by a Lennard-Jones potential of mean force. Our results indicate that: (1) The temperature for single-chain collapse (T_θ), the Boyle temperature (T_B), and the upper critical solution temperature for polymer-solvent phase separation (T_c) asymptotically converge to the same value for long chains, consistent with Flory-Huggins mean-field predictions for polymers on a lattice. (2) The asymptotic scaling of the second virial coefficient with chain length in the poor solvent regime is exponential. (3) The emergence of the three scaling regimes for B_2 (i.e., the good solvent regime, theta solvent regime and poor solvent regime), the scaling of B_2 with chain length in those regimes, and—to lesser extent—the actual value of B_2, are unaffected by single-chain collapse. The third point suggests that the residual free-energy of dilute polymer-solvent systems is insensitive to changes in chain conformation, implying a simplified route to developing thermodynamic models for describing polymer-solvent phase equilibria based on molecular models that do not exhibit single-chain collapse. Based on our simulation data and

liquid-state perturbation theory, we develop an analytic model for the second virial coefficient of Lennard-Jones chains that might further benefit the development of such thermodynamic models.

4.2 Introduction

The conformation of chain molecules depends sensitively on the temperature, density, and composition of their environment.[1–5] For chains in a solvent, this dependence is usually summarized in terms of the "quality" of the solvent —a measure for the average interaction, or potential of mean force, between the segments of a chain molecule in solution. In good solvents, the segments of a chain preferably interact with solvent molecules, resulting in an effective repulsion between chain segments and an expanded coil-like structure of the chain. Conversely, in poor solvents, a chain's segments tend to favour self interaction, causing a chain to collapse into a more compact, globular structure. The boundary between these solvent regimes, known as the theta region, is characterized by an overall balance of the repulsive and attractive forces between the chain's segments, yielding chain conformations akin to ideal chains, where interactions between the segments of a chain (separated by a great distance along the chain contour) are absent.[2,4,5]

The interactions between chain molecules in a solvent, and consequently phase equilibria such as polymer-solvent phase separation, are believed to be strongly coupled to changes in chain conformation. For a large part, this expectation is rooted in the mean-field theories of Flory[1] and Flory-Huggins,[6–9] which predict that the temperature for single-chain collapse $T_\theta(m)$, the Boyle temperature $T_\mathrm{B}(m)$(defined by a second virial coefficient equal to zero) and the upper critical solution temperature for polymer-solvent phase separation $T_\mathrm{c}(m)$ converge to the same value (i.e. the theta temperature θ) if the number of segments per chain m increases to values typical for real polymers.[5,10] Some theoretical and simulation studies have suggested similar relationships for the lower critical solution temperature in a polymer-solvent system.[11,12] Despite these connections, our quantitative understanding of the coupling between chain conformation and phase behaviour remains limited, especially at low temperatures where most significant conformational changes occur. Largely, this is caused by difficulties in experiments used to track single-chain collapse,[10] but sampling issues in molecular simulations at low temperatures and the inadequacy of mean-field theories in the poor solvent regime[5] further complicate matters.

In this [chapter], we analyse the relationship between chain conformation and the free energy of fluid mixtures containing chain molecules. Our ultimate goal is to determine if —and to

what extent —changes in chain conformation need to be incorporated in theoretical models aimed at predicting polymer(-solvent) properties and phase equilibria.

We focus our analysis on the simplest possible quantity relating chain conformation, the intramolecular and intermolecular potential energy, and the free energy of polymer-solvent systems: the osmotic second virial coefficient B_2.[13] The scaling of this coefficient with the number of segments per polymer chain, and how this scaling is affected by changes in chain conformation is studied in detail by means of Monte Carlo (MC) simulations of a simple model system for chain molecules in implicit solvent, using a Lennard-Jones (LJ) 12-6 pair potential for describing the effective interaction between chain segments. This effective interaction is thus a potential of mean force, describing the interaction between chain segments averaged over all solvent degrees of freedom. Strictly speaking, the potential of mean force depends on the thermodynamic state and differs for each segment on a chain molecule,[14] but this is neglected throughout this work. As shown by McMillan-Mayer theory[13] (see also Hill[15] and Vafaei *et al*[16]), the osmotic virial expansion around a pure solvent is equivalent to the standard virial expansion around an ideal gas, if the latter is expressed in terms of a potential of mean force instead of an intermolecular potential energy. Therefore in the remainder we will drop the word "osmotic", and simply refer to "virial expansion" and "second virial coefficient".

The second virial coefficient of 3D off-lattice chains with attractive segment-segment interactions has been studied by numerous authors.[17–26] However, simulation data in the low-temperature (poor solvent) regime is scarce and studies that connect this data to single-chain characteristics are absent. The only data showing some results below the Boyle temperature are those of Harismiadis and Szleifer[20] and those of Withers *et al.*[24] both for chains with segments that interact by a LJ truncated and shifted pair potential. The results of both studies are contradictory however, showing qualitatively different trends with chain length. In this work we resolve these differences based on several theoretical arguments and a comparison to our new MC data.

Single-chain properties of 3D off-lattice chains have also been studied extensively. Results of particular relevance to this study concern studies on freely-jointed chains of segments with square-well interactions[27,28] and Lennard-Jones interactions[12,14,17,25,26,29,30]. Notable is the work of Taylor, who derived exact results for the single-chain properties of isolated LJ chains,[30] and LJ chains in explicit solvent,[14] using a cluster expansion of the single-molecule partition function. Also relevant to the present work, are the recent molecular simulation studies of Schultz and Kofke,[26] who analysed the Boyle temperature and theta temperature of Lennard-Jones chains comprising up to 512 segments, and that of Schnabel and Janke,[25] who used a special Monte Carlo algorithm to study the single-chain properties of Lennard-Jones

chains (with different bond-length than the chains analysed in this work) of up to 32768 segments.

Given our interest in developing thermodynamic models for chain molecules,[31,32] we pay particular attention to developing an analytic model able to describe our Monte Carlo simulation results for the second virial coefficient of freely-jointed Lennard-Jones chains. We develop this model using liquid-state perturbation theory, which differs from typical perturbation theories used in polymer physics[2] in terms of the chosen reference fluid, which in this work comprises chain molecules with repulsive excluded volume interactions instead of ideal chains. The developed model for the second virial coefficient is expected to benefit the development of thermodynamic models for polymer melts and concentrated polymer solutions. Fluid theories (e.g. Flory-Huggins theories,[6–9] generalized Flory-dimer theory[33], Wertheim's TPT,[34,35] and SAFT[36–39]) aiming to describe such systems can usually not reliably predict intramolecular structure of larger chain molecules, because this structure is determined by many-body correlation functions. Especially at low density, where single-chain transitions are expected to occur, the adequacy of these approximations might be questioned. The virial expansion offers a potential solution, because its coefficients establish a direct link between the intra- and intermolecular interactions of chains and the free energy.[40–42] Although the virial expansion is applicable primarily to fluids at low density (or low concentration), the lower-order coefficients can enhance fluid theories (which in principle are applicable at any density) by ensuring the correct low-density behaviour. Our recent endeavours in developing uf-theory[43] and uv-theory,[32,44,45] which augment standard first-order perturbation theory with an accurate low-density description, exemplify such approaches.

This [chapter] is organized as follows. In Sec. 4.3 we define the molecular model. Relevant theory is discussed in Sec. 4.4, and simulation details are provided in Sec. 4.5. Simulation results are presented and discussed in Sec. 4.6, whereas the model for the second virial coefficient is developed and analysed in Sec. 4.7. In Sec. 4.8 we present our conclusions.

4.3 Molecular Model

We consider linear chains of m identical spherical segments bonded tangentially, at a fixed distance σ. The full configuration of a chain is defined by the chain's centre-of-mass position $\boldsymbol{R}$ and conformation $\boldsymbol{\Omega}$. The conformation $\boldsymbol{\Omega}$ is specified by the set of $m-1$ unit vectors that describe the orientation of the bonds, $\hat{\boldsymbol{\omega}}_1$, $\hat{\boldsymbol{\omega}}_2$, …, $\hat{\boldsymbol{\omega}}_{m-1}$, where $\boldsymbol{\omega}_\alpha = \mathbf{r}_{\alpha+1} - \mathbf{r}_\alpha$, and $\mathbf{r}_\alpha$ is the position vector of a segment of index $\alpha \in [1, m]$. The full configuration of a chain of index i is denoted by the short-hand notation $\boldsymbol{i} = (\boldsymbol{R}_i, \boldsymbol{\Omega}_i)$.

Two different chain types are studied: chains of freely-jointed segments, characterized by an intramolecular potential energy without any bond-bending or torsion potentials, and linear-rigid chains, where all bond vectors are equal.

The segments of different chains and those within the same chain separated by at least two bonds are assumed to interact through a LJ 12-6 pair potential

$$u(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{4.1}$$

where r is the scalar distance distance between two segments, and σ and ϵ define the size parameter and the well depth. The intramolecular potential energy of a chain of index 1 can thus be written

$$U^{\text{intra}}(\mathbf{1}) = \sum_{\alpha_1=1}^{m-2} \sum_{\alpha_1'=\alpha_1+2}^{m} u(r_{\alpha_1\alpha_1'}) \tag{4.2}$$

whereas the intermolecular potential energy of two chains of index 1 and 2 (with $m_1 = m_2 = m$) follows from a summation over all m^2 segment-segment interactions

$$U^{\text{inter}}(\mathbf{12}) = \sum_{\alpha_1=1}^{m} \sum_{\alpha_2=1}^{m} u(r_{\alpha_1\alpha_2}) \tag{4.3}$$

In the remainder, properties are made dimensionless by the size parameter σ and the energy parameter ϵ of the LJ potential whenever appropriate. Dimensionless properties are denoted by an asterisk, e.g. $T^* = k_B T/\epsilon$ denotes dimensionless temperature and $B_2^* = B_2/\sigma^3$ is the dimensionless second virial coefficient.

4.4 Theory

4.4.1 Second Virial Coefficient

The second virial coefficient of two chain molecules 1 and 2 is defined as a weighted average of the Mayer-f function

$$f_{\mathrm{M}}(\mathbf{12}) = \exp\left(-\beta U^{\mathrm{inter}}(\mathbf{12})\right) - 1 \tag{4.4}$$

where $\beta = 1/(k_{\mathrm{B}}T)$, according to[40–42]

$$\begin{aligned} B_2 &= -\frac{1}{2}\int f_{\mathrm{M}}(\mathbf{12}) f(\mathbf{\Omega_1}) f(\mathbf{\Omega_2}) \mathrm{d}\mathbf{\Omega_1}\,\mathrm{d}\mathbf{\Omega_2}\mathrm{d}\mathbf{r} \\ &\equiv -\frac{1}{2}\int \langle f_{\mathrm{M}}(\mathbf{12})\rangle\,\mathrm{d}\mathbf{r} \end{aligned} \tag{4.5}$$

Here, $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is the vector between the centres of mass of the chains, and $f(\mathbf{\Omega}_i)$ is the Boltzmann probability density for finding a single, *isolated* chain of index i in a conformation $\mathbf{\Omega}_i$

$$f(\mathbf{\Omega}_i) = \frac{\exp\left(-\beta U^{\mathrm{intra}}(\mathbf{\Omega}_i)\right)}{\int \exp\left(-\beta U^{\mathrm{intra}}(\mathbf{\Omega}_i)\right)\mathrm{d}\mathbf{\Omega}_i} \tag{4.6}$$

which integrates to unity as $\int f(\mathbf{\Omega}_i)\mathrm{d}\mathbf{\Omega}_i = 1$.

The notation $\langle X(1\ldots n)\rangle$ introduced in Eq. (4.5) will be used throughout this work; it means an average of property $X(1\ldots n)$ over the conformations of n *isolated* chains.

4.4.2 Perturbation Contributions to the Second Virial Coefficient

At high temperatures ($k_{\mathrm{B}}T >> \epsilon$) the effect of attractive intermolecular interactions on the Mayer-f function $f_{\mathrm{M}}(12)$ becomes small. The second virial coefficient can then be described accurately by expanding about the second virial coefficient of two reference molecules with

only repulsive intermolecular interactions. The expansion coefficients (or "perturbation terms") provide a convenient starting point to develop a model for the second virial coefficient and provide additional insight into its behaviour.

We define the reference part of the intra- and intermolecular segment-segment potential $u(r)$ as

$$u_0(r) = \begin{cases} u(r) - u(r_s) & \text{for } r < r_s \\ 0 & \text{for } r \geq r_s \end{cases} \tag{4.7}$$

and the perturbation part as

$$w(r) = u(r) - u_0(r) \tag{4.8}$$

where $r_s = \sigma + \gamma(r_{min} - \sigma)$, $r_{min} = 2^{1/6}\sigma$ is the distance at the potential well, and the parameter γ is used to distinguish between the Barker-Henderson (BH)[46] division ($\gamma = 0$) and the Weeks-Chandler-Andersen (WCA)[47–50] division ($\gamma = 1$). We will only show results for the BH division in this [chapter]; results obtained using the WCA division are included in the SM.

The contribution to B_2 due to the total (i.e., intramolecular and intermolecular) perturbation energy $W(\mathbf{12}) = U(\mathbf{12}) - U_0(\mathbf{12})$ with $U(\mathbf{12}) = U^{\text{intra}}(\mathbf{1}) + U^{\text{intra}}(\mathbf{2}) + U^{\text{inter}}(\mathbf{12})$ is derived by introducing the following coupling parameter scheme

$$U_\lambda(\mathbf{12}) = U_0(\mathbf{12}) + \lambda W(\mathbf{12}) \quad \lambda = [0, 1] \tag{4.9}$$

and integrating the second virial coefficient from a coupling parameter $\lambda = 0$ to $\lambda = 1$, leading to

$$\Delta B_2 = \int_0^1 \left(\frac{\partial B_{2,\lambda}}{\partial \lambda} \right) \mathrm{d}\lambda \tag{4.10}$$

We find that the derivative $(\partial B_{2,\lambda} / \partial \lambda)$ decomposes into an intermolecular *and* intramolecular

contribution, leading to

$$\Delta B_2 = \Delta B_2^{\text{inter}} + \Delta B_2^{\text{intra}} \tag{4.11}$$

where the intermolecular part takes on a form analogous to that of simple spherical fluids (besides the integration over conformation)

$$\Delta B_2^{\text{inter}} = \frac{\beta}{2} \int_0^1 \int \left\langle \exp\left[-\beta U_\lambda^{\text{inter}}(\mathbf{12})\right] W^{\text{inter}}(\mathbf{12}) \right\rangle_\lambda \mathrm{d}\mathbf{r}\mathrm{d}\lambda \tag{4.12}$$

and the intramolecular part can be written as

$$\Delta B_2^{\text{intra}} = \frac{\beta}{2} \int_0^1 \int \left\langle \exp\left[-\beta U_\lambda^{\text{inter}}(\mathbf{12})\right] \left(W^{\text{intra}}(\mathbf{12}) - \left\langle W^{\text{intra}}(\mathbf{12}) \right\rangle_\lambda\right) \right\rangle_\lambda \mathrm{d}\mathbf{r}\mathrm{d}\lambda \tag{4.13}$$

where we introduced

$$W^{\text{intra}}(\mathbf{12}) \equiv W^{\text{intra}}(\mathbf{1}) + W^{\text{intra}}(\mathbf{2}) \tag{4.14}$$

The subscript λ in the averages $\langle \ldots \rangle_\lambda$ of Eqs. (4.12)-(4.13) denotes an average over the conformations of two isolated chains with segment-segment interactions governed by the potential $u_0(r) + \lambda w(r)$ of Eqs. (4.7)-(4.8).

The intramolecular contribution $\Delta B_2^{\text{intra}}$ arises from the derivative of the intramolecular distribution function $f_\lambda(\mathbf{\Omega}_i)$ to the coupling parameter λ; it thus captures the contribution to B_2 due to conformational changes induced by attractions between the segments within a chain. In the remainder we neglect the intramolecular contribution $\Delta B_2^{\text{intra}}$ and, in line with this assumption, we assume that the intermolecular contribution $\Delta B_2^{\text{inter}}$ of Eq. (4.12) can be approximated by averaging over the conformations of the repulsive reference molecules $\langle \ldots \rangle_0$ instead of over those of the target molecules $\langle \ldots \rangle_\lambda$,

$$\begin{aligned}\Delta B_2^{\text{intra}} &\approx 0 \\ \Delta B_2^{\text{inter}} &\approx \frac{\beta}{2}\int_0^1 \int \left\langle \exp\left[-\beta U_\lambda^{\text{inter}}(\mathbf{12})\right] W^{\text{inter}}(\mathbf{12}) \right\rangle_0 \mathrm{d}\mathbf{r}\mathrm{d}\lambda\end{aligned} \tag{4.15}$$

In the results section of this [chapter] we show these assumptions hold surprisingly well, even for temperatures quite far below the Boyle temperature.

Substituting Eq. (4.15) in Eq. (4.11), expanding the integrand about $\lambda = 0$ and performing the (trivial) integration over λ leads to the following perturbation series

$$\Delta B_2 \approx \sum_{n=0}^{\infty} \Delta\tilde{B}_{2n} \tag{4.16}$$

where a perturbation contribution of order n is given as

$$\Delta\tilde{B}_{2n} = (-1)^{n+1}\frac{\beta^n}{2n!}\int \left\langle \exp\left[-\beta U_0^{\text{inter}}(\mathbf{12})\right]\left(W^{\text{inter}}(\mathbf{12})\right)^n \right\rangle_0 \mathrm{d}\mathbf{r} \tag{4.17}$$

and the tilde denotes that these perturbation contributions are subject to the approximations of Eq. (4.15). The equations for the exact perturbation contributions ΔB_{2i}, derived without those simplifying approximations, are provided in the SM. In case all intramolecular segment-segment interactions would be considered in the reference fluid, the above perturbation series is exact.

4.4.3 Single-chain Properties

To track the effect of changes in chain conformation on the second virial coefficient we compute various single-chain properties.

As a measure for the average size and conformation of a chain, we compute the mean-squared radius of gyration,

$$\left\langle R_g^2 \right\rangle = \frac{1}{m} \sum_{\alpha=1}^{m} \left\langle |\boldsymbol{r}_\alpha - \boldsymbol{R}|^2 \right\rangle \tag{4.18}$$

and the mean-squared end-to-end length,

$$\left\langle R_e^2 \right\rangle = \left\langle |\boldsymbol{r}_m - \boldsymbol{r}_1|^2 \right\rangle \tag{4.19}$$

To probe single-chain phase transitions it is also useful to calculate the isochoric heat capacity of a chain, which is obtained from the variance of the dimensionless intramolecular energy $\beta U^{\text{intra}}(\mathbf{1})$, according to

$$\frac{C_v}{k_B} = \beta^2 \left(\left\langle U^{\text{intra}}(\mathbf{1}) \right\rangle^2 - \left\langle U^{\text{intra}}(\mathbf{1})^2 \right\rangle \right) \tag{4.20}$$

Conformational phase transitions will show a peak of C_v at the phase transition temperature.

As a measure for the structural order within a chain one can define the average segment-segment distance fluctuation,

$$\Delta_B = \frac{1}{m-1} \frac{1}{m-2} \sum_{\alpha<\alpha'}^{m} \frac{\left[\left\langle r_{\alpha\alpha'}^2 \right\rangle - \langle r_{\alpha\alpha'} \rangle^2 \right]^{\frac{1}{2}}}{\langle r_{\alpha\alpha'} \rangle} \tag{4.21}$$

For clusters of LJ spheres, a value of $\Delta_B \leq 0.1$ suggests solid-like ordering.[51,52]

4.5 Simulation Details

Computation of the averages presented in Sec. 4.4 requires generating a large number of independent single-chain conformations. We do that using three different MC methods:

1. $\boldsymbol{m \leq 16}$: Growing each chain from scratch, using the Rosenbluth scheme.[53–56] Since this scheme does not by design generate chain conformations that obey Boltzmann statistics, each chain must be given an appropriate weight factor (the Rosenbluth weight W) in

computing the averages.

2. $m \leq 32$: Canonical NVT-MC simulations of a single chain, or of a system of N chains with a density low enough such that chain conformations of different chains can be considered independent. The latter was verified by comparing sampled single-chain properties to those sampled using the Rosenbluth method (see Fig. B.1).

3. $m \leq 80$: Mayer-sampling (MS) MC simulations for the second virial coefficient, using the Zeno software package. [57,58] In these simulations, two chains are simulated using importance sampling in the argument of the integral of Eq. (4.5). By design, this method favours chain conformations that have the largest contribution to the second virial coefficient. The bias introduced by this is automatically removed.

For methods 1 and 2, the second virial coefficient (Eq. (4.5)) is integrated numerically over a grid in centre-of-mass distance of chains 1 and 2. We do this for all N_{pair} independent chain pairs along several directions and take the average. For method 1, each chain pair must be given the appropriate weight factor, defined by the product of the two Rosenbluth weights W_1 and W_2. The statistical uncertainties in method 1 are estimated by calculating the standard deviation of five independent simulations, while for method 2 we estimate the standard error based on block averaging with five blocks. For details on the estimation of the uncertainty of the MSMC simulations (method 3) we refer to Bansal *et al.*[58]. More details for each MC methods are provided in the SM.

The single-chain properties of short LJ chains computed by methods 1 and 2 were validated to the exact cluster-expansion results of Taylor.[30] All results match to within the statistical uncertainty of the simulations. As shown in Figs. B.2-B.3, our results for B_2 obtained using methods 1, 2 and 3 typically agree to within the estimated statistical uncertainties. A comparison to the literature data of Chiew[23] is also given there.

4.6 Results and Discussion

4.6.1 Single-Chain Properties

Fig. 4.1 shows our MC simulation results for the radius of gyration, internal energy, isochoric heat capacity and average segment-segment distance fluctuation of chains with up to 256 Lennard-Jones segments. The radius of gyration displays familiar behaviour, starting from a relatively high value at elevated temperatures and decreasing to a relatively constant small value when temperature is decreased. The drop in $\left\langle R_g^2 \right\rangle$ is associated with the transition from

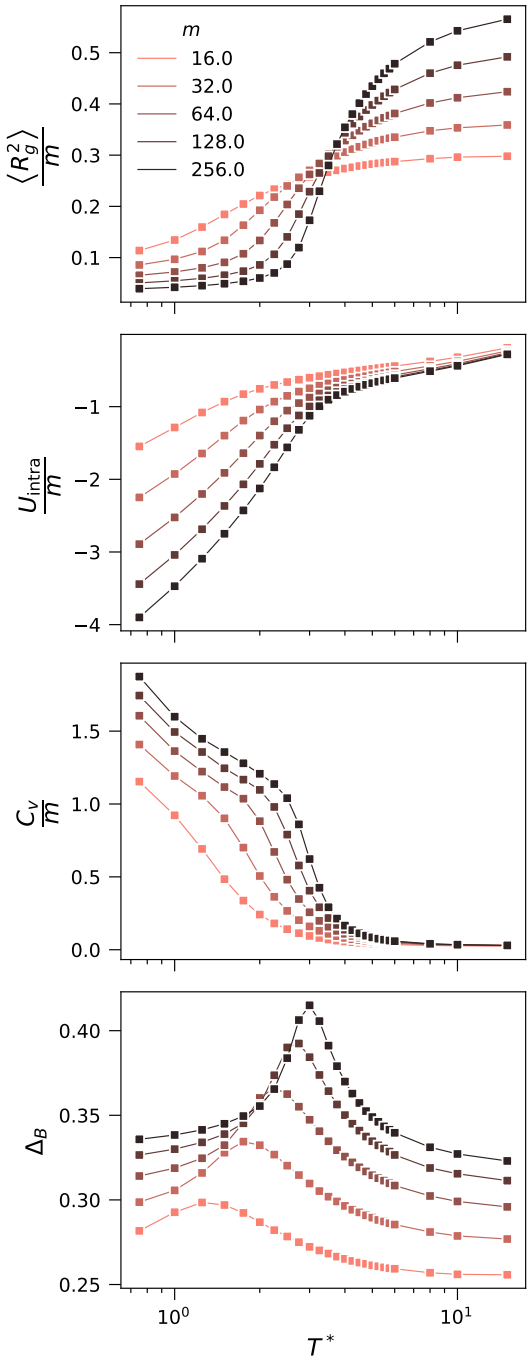


Figure 4.1: MC simulation results for the radius of gyration, intramolecular potential energy, heat capacity and segment-segment distance fluctuation of a single chain of m freely-jointed Lennard-Jones segments. Estimated standard errors are smaller than the symbol size. Lines are a guide for the eye.

an expanded chain conformation (expanded coil) at high temperatures, to a more compact, folded chain conformation (globule) at low temperatures. For a polymer in solution, this

transition corresponds to going from a good solvent to a poor solvent. The transition to the folded state is driven by the attractive interactions between the segments of the chain, which favour a compact conformation that minimizes the chain's internal energy. Only below a certain temperature (the coil-globule transition temperature) the decrease in internal energy due to folding into a globule can actually off-set the increase in free energy associated with the decrease in conformational entropy ($A = U - TS$), leading to a phase transition.

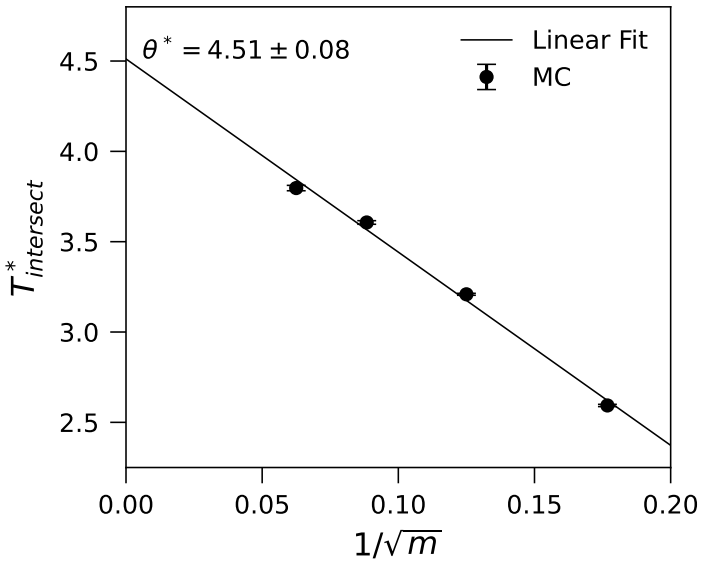


Figure 4.2: Temperature at the intersection of the $\left\langle R_g^2 \right\rangle$-vs-$T^*$ curves of freely-jointed LJ chains of $m = \{16, 32\}$, $m = \{32, 64\}$, $m = \{64, 128\}$, and $m = \{128, 256\}$. Extrapolation to infinite chain length to estimate the theta temperature $\theta^* = 4.51 \pm 0.08$. Error bars shown in this diagram correspond to one standard deviation.

As shown in Fig. 4.1 (b-d), the coil-globule transition is accompanied by a change in slope of the internal energy, an increase in the heat capacity and a peak in the average segment-segment distance fluctuation Δ_B. The absence of a clear peak in the heat capacity at the coil-globule transition might be caused by the onset of another peak associated with a freezing transition at lower temperatures, which overlaps with part of the peak at the coil-globule transition. Such behaviour has been observed for chains comprising segments that interact by a square-well potential.[29] The freezing transition takes place at lower temperatures than those considered in our simulations; adequate sampling at such low temperatures requires more advanced simulation methods than used in this work, see e.g. the work of Taylor,[28,59] where Wang-Landau sampling is used for this purpose.

At the coil-globule transition temperature, the effect of the attractive (energetic) and repulsive (entropic) contributions of the segment-segment interactions to the free energy of a single chain

cancel each other. As a result, the equilibrium properties of the chain can be described by an ideal chain, which has known[5] scaling $\left\langle R_g^2 \right\rangle \sim m$. At the transition temperature, the quantity $\left\langle R_g^2 \right\rangle / m$ is thus expected to approach a constant value upon increasing the chain length. In the diagram shown in Fig. 4.1 (a), this implies that the temperature at the intersection of the lines corresponding to $m = \{16, 32\}$, $m = \{32, 64\}$, $m = \{64, 128\}$, and $m = \{128, 256\}$ should approach a constant value $T^*_{\text{intersect}} \to \theta^*$. We calculated the intersections by equating fifth-order polynomials fitted to the radius of gyration results of chains with specified length in the range $T^* = [2, 6]$. We repeated this procedure 500 times with each data point given a random perturbation according to a Gaussian distribution characterized by a width equal to the estimated standard error of that data point. The standard deviation of the intersection temperature was then calculated as the root-mean-squared deviation with respect to the unperturbed intersection temperature. To extrapolate to infinite chain length, the theta temperature θ^* and constant a of the following scaling relation were estimated based on a least-square regression of our results,

$$\theta^* - T^*_{\text{intersect}}(m) \sim a m^{-1/2} \tag{4.22}$$

Results of the fit are shown in Fig. 4.2. We estimate a theta temperature $\theta^* = 4.49 \pm 0.025$, which is approached from below, with $a < 0$. The reported statistical uncertainty is a standard error, estimated from the covariance matrix of the regressed constants.

Another way to estimate the theta temperature is to find the temperature at which $\eta = \left\langle R_e^2 \right\rangle / (6 \left\langle R_g^2 \right\rangle) = 1$, which can be derived for chains that obey ideal-chain statistics.[5] The dependence of η on temperature is shown in the left diagram of Fig. 4.3 for chains of 16 to 256 freely-jointed LJ segments. At high temperatures, the effect of the attractive interactions between chain segments diminishes and the conformations of the chains should have statistics similar to a self-avoiding walk. In this limit, our simulations predict $\eta = 1.06$, irrespective of chain length, which agrees with other estimates for athermal chains from theory, simulations and experiments, which lead to values of η in the range 1.05-1.07 (see Luna-Bárcenas *et al.* [12] and references therein). At low temperatures, where the chains fold into a globule, η seems to approach a value slightly smaller than 0.4, whereas for a completely spherical globule of uniform segment density, geometrical arguments predict $\eta = 1/3$.[3] The deviation between both values is likely caused by the fact that the globules in our simulations are not perfect spheres, not fully folded and of non-uniform segment density. The temperature for which $\eta = 1$ is denoted as T_{ideal}. To find the theta temperature, we solved $\eta(T^*_{\text{ideal}}) = 1$ for each m by fitting a fourth order polynomial in T^* to our simulation results for η in the range

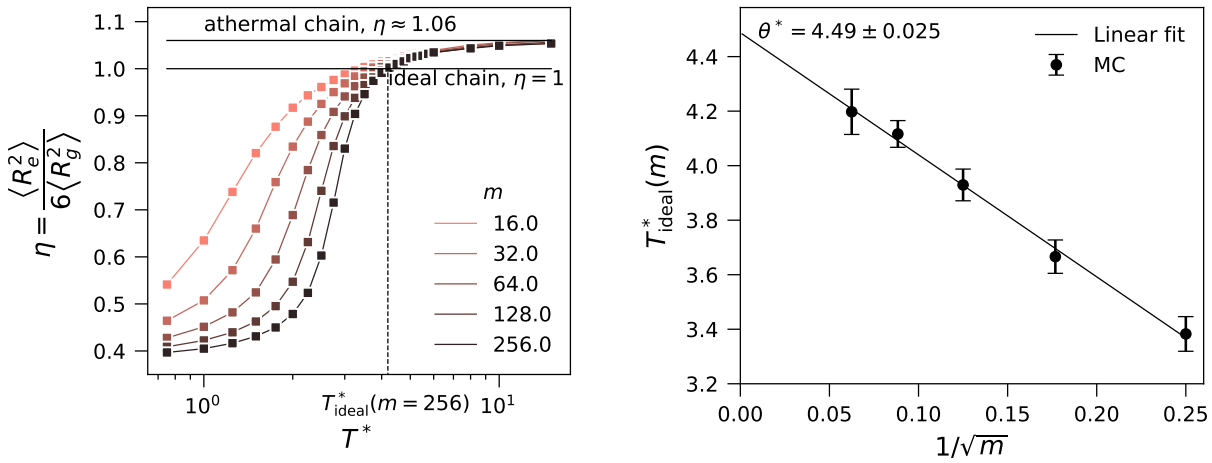


Figure 4.3: Left: Temperature dependence of η for chains of m freely-jointed LJ segments. At the temperature where $\eta = 1$, chains obey ideal-chain statistics. Standard deviations (estimated by linear error propagation) are smaller than the size of the symbols; lines are a guide for the eye. Right: Temperature at which $\eta = \eta_{\mathrm{ideal}} = 1$. Extrapolation to infinite chain length to estimate the theta temperature $\theta^* = 4.49 \pm 0.025$. Error bars shown in this diagram correspond to one standard deviation.

$\eta = [0.9, 1.015]$. To estimate the standard deviation of the calculated temperature, this procedure was repeated 250 times with each measured data-point for the radius of gyration and the end-to-end length given a random perturbation according to a Gaussian distribution of width equal to the estimated standard error of that data-point. Calculated results for T^*_{ideal} were extrapolated to infinite chain length based on a least-squares regression of Eq. (4.22). The results are shown in Fig. 4.3. We obtain $\theta^* = 4.51 \pm 0.08$, which agrees to our previous estimate $\theta^* = 4.49 \pm 0.025$ to within the estimated standard errors.

4.6.2 Second Virial Coefficient

4.6.2.1 Scaling Regimes

In Fig. 4.4, we show a subset of our simulation results for the second virial coefficient per segment segment interaction B_2/m^2 versus temperature. For visual clarity, chains of more than 16 segments are not included. For freely-jointed chains (left diagram) the results for different chain lengths are consistent with the known scaling regimes at high temperatures, where $B_2/m^2 \sim m^{-\nu}$, and at temperatures slightly below the Boyle temperature T^*_{B}, defined by $B_2(T^*_{\mathrm{B}}) = 0$, we find $B_2/m^2 \sim 1$. At low temperatures, our simulation results suggest another regime where B_2/m^2 scales very strongly with m. For polymers in solution, these three scaling

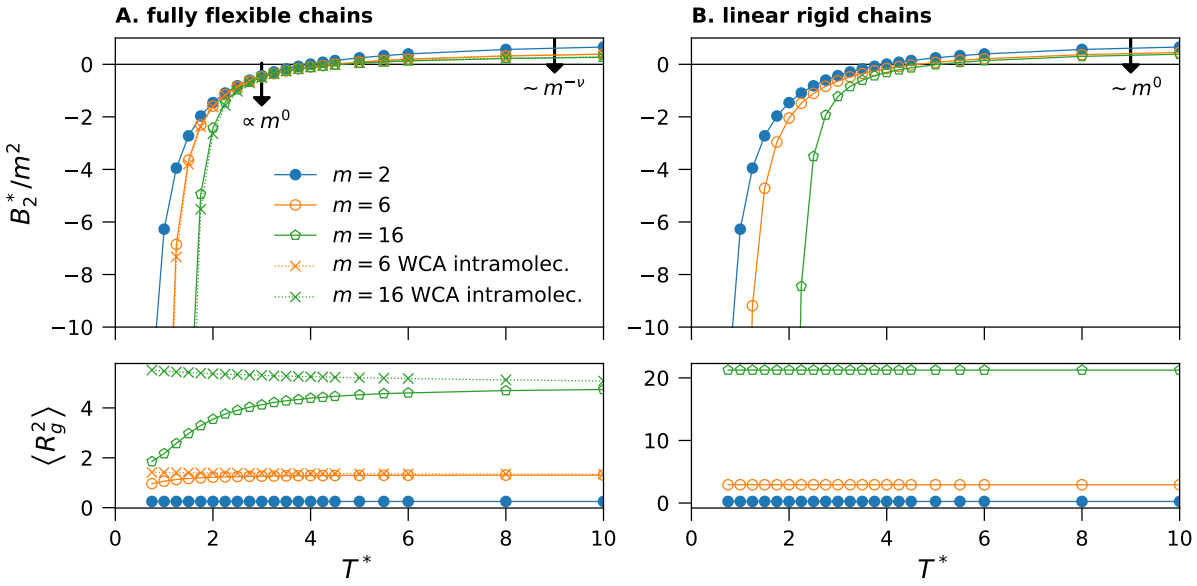


Figure 4.4: Second virial coefficient per segment-segment interaction (top) and radius of gyration (bottom) of freely-jointed (left) and linear, rigid (right) chains of m Lennard-Jones segments. In the left figure, we compare to results for chains where the intramolecular segment-segment potential is replaced by a repulsive WCA[47–50] potential (crosses). Known scaling regimes are shown by the arrows; $\nu \approx [0.2-0.25]$.[60] Estimated standard errors are smaller than the symbol size;
lines are a guide for the eye.

regimes correspond to the good solvent regime, the theta solvent regime, and the poor solvent regime, respectively.

For chains of segments arranged in a rigid, linear conformation (right diagram) qualitatively different scaling regimes are observed. The scaling in the good solvent regime changes to $B_2/m^2 \sim m^0$ (which is known exactly[61–63]), whereas the scaling in the poor solvent regime appears even stronger than for flexible chains. The theta regime disappears altogether.

The scaling of the second virial coefficient in the poor solvent regime $T < \theta$ is further analysed in Fig. 4.5, where we show our results for B_2/m^2 (on a logarithmic scale) versus chain length. For the lowest temperature examined ($T^* = 0.75$), our results suggest exponential scaling $B_2/m^2 \sim -\exp(m)$. The scaling gradually becomes weaker when the temperature increases towards the Boyle temperature, where B_2/m^2 is independent of m.

The observed exponential scaling is consistent with previous predictions obtained from approximate field-theoretic simulations, by Raos and Allegra.[64] Their Gaussian Cluster Theory suggests that exponential scaling is only obtained if the two chain molecules are allowed to relax their conformation as they approach each other and interact. Exponential scaling thus

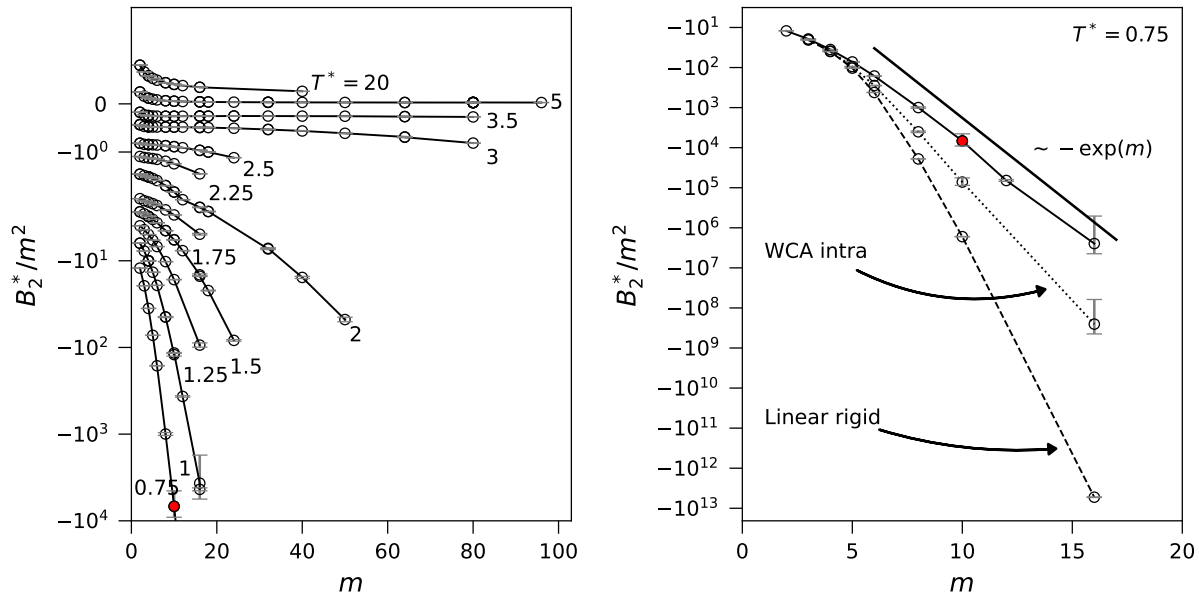


Figure 4.5: Left: Scaling of the second virial coefficient per segment segment interaction with the number of LJ segments per chain m, visualized using a symlog scale for the y-axis. Right: Magnification of the results in the poor solvent regime (for a temperature $T^* = 0.75$) and a comparison to results for freely-jointed chains with the intramolecular segment-segment potential replaced by a repulsive WCA[47–50] potential (dotted lines) and results for linear, rigid Lennard-Jones chains (dashed lines). Error bars denote one estimated standard error; lines are a guide for the eye. The red-coloured symbol in both graphs denotes the same data-point.

corresponds to the true, equilibrium behaviour of the second virial coefficient. Insufficient relaxation was shown to result in (apparent) values for B_2/m^2 approaching zero in the poor solvent regime, providing a possible explanation of the disputed[64–66] experimental results of Nierlich *et al.*[67] which suggest exactly that behaviour.

Related to this, our simulation results in the poor solvent regime solve a discrepancy between the Monte Carlo simulation data of Harismiadis and Szleifer[20] and the Monte Carlo simulation data of Withers *et al.*[24] for the second virial coefficient of chains with a truncated (at 2.5σ) and shifted Lennard-Jones potential. While the data of Harismiadis shows similar trends as our data, the data of Withers *et al.* suggests a qualitatively different poor-solvent-regime scaling $B_2/m^2 \sim m^{-1}$. Noting the equivalence of this scaling with the experimental results of Nierlich *et al.*[67], we speculate that the simulation method of Withers *et al.* did not adequately sample relevant chain conformations.

4.6.2.2 Effect of Chain Folding

It is tempting to suggest that the stronger scaling of B_2 with chain length m in the poor solvent regime is caused by the folding of the chains.[18,20] After all, this is what differentiates this regime from the theta and good solvent regimes. However, we show in Figs. 4.4 and 4.5 that such reasoning is flawed.

Our argument is twofold. First, linear-rigid chains, which do not fold by definition, display a more negative second virial coefficient than freely-jointed chains. Moreover, in the poor solvent regime, the scaling of the second virial coefficient of linear-rigid chains with chain length (dashed line in the right diagram shown in Fig. 4.5) is stronger than for freely-jointed chains. The second part of our argumentation follows from a comparison between the second virial coefficient of freely-jointed Lennard-Jones chains that only exhibit repulsive *intramolecular* interactions (crosses in Fig. 4.4) to our previous results, where the segments within a chain (separated by two or more bonds) interact by the full, attractive Lennard-Jones potential. As shown by the results for the radius of gyration in the lower diagram of Fig. 4.4, the chains with repulsive intramolecular interactions do not fold when temperature decreases. However, the second virial coefficient of these chains is not less negative in the poor solvent regime; in fact, it is slightly more negative, and its scaling with chain length is slightly stronger than for Lennard-Jones chains (see the dotted line in the right diagram of Fig. 4.5 for the scaling at $T^* = 0.75$).

Folding is thus not causing the stronger scaling of B_2 with chain length in the poor solvent regime. The actual source of this behaviour must simply lie in the exponential dependence of the Mayer-f function $f_{\mathrm{M}}(\mathbf{12}) = \exp(-\beta U^{\mathrm{inter}}(\mathbf{12}))-1$ on inverse temperature β, combined with the fact that the interaction energy of two chains $U^{\mathrm{inter}}(\mathbf{12})$ may comprise several (maximal m^2) simultaneous segment-segment interactions. This is easily illustrated by expanding the Mayer function about $\beta = 0$. At high temperatures, this leads to $f_{\mathrm{M}}(\mathbf{12}) \propto -\beta U^{\mathrm{inter}}(\mathbf{12}) \sim m^2$, where we assumed all m^2 segment-segment interactions contribute to $U^{\mathrm{inter}}(\mathbf{12})$. When temperature is lowered, higher-order expansion terms $\frac{1}{2}(\beta U^{\mathrm{inter}}(\mathbf{12}))^2$, $-\frac{1}{6}(\beta U^{\mathrm{inter}}(\mathbf{12}))^3$, etc. become relevant, gradually increasing the scaling with chain length to m^4, m^6 and so on, until the scaling of the full exponential, $f_{\mathrm{M}}(\mathbf{12}) = \exp(-\beta U^{\mathrm{inter}}(\mathbf{12}))-1 \sim \exp(m^2)$ is obtained. Of course, the actual number of segment-segment interactions for a given configuration of two chains will typically be smaller then m^2; the scaling of the second virial coefficient will therefore be weaker than $B_2/m^2 \sim -\exp(m^2)$.

Our results for the perturbation contributions to the second virial coefficient shown in Fig. 4.6 display exactly this behaviour. At the level of a first-order perturbation theory, B_2/m^2 goes to zero as chain length increases. Increasing the order of the perturbation theory to second-order

and then third-order qualitatively changes the dependence of B_2 on chain length to that observed for the full B_2, i.e. at the level of a third-order perturbation theory, B_2/m^2 becomes more negative for longer chain lengths. Having that said, the results shown in figure 4.7 clearly demonstrate that a third-order perturbation theory is not a sufficient model to quantify the full behaviour of B_2. For further details on the modelling of B_2, we refer to Sec. 4.7.

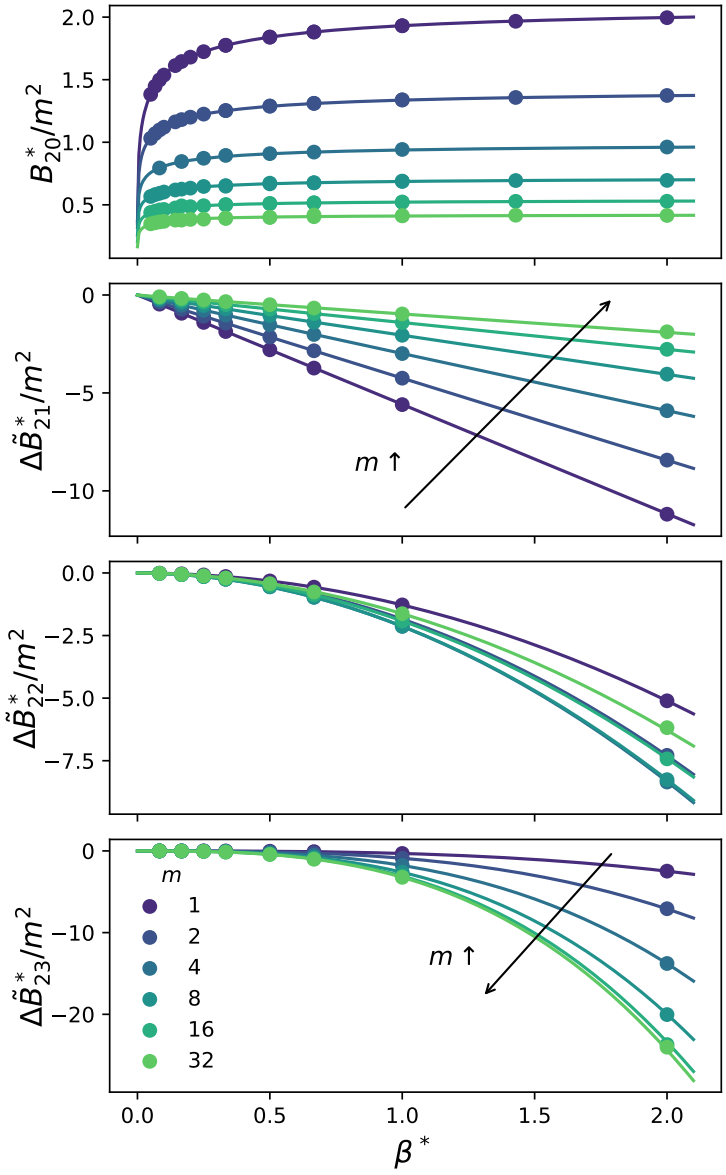


Figure 4.6: The BH reference contribution (B_{20}) and first three BH perturbation contributions (ΔB_{2i}, $i = 1, 2, 3$) to the second virial coefficient, as obtained from Monte Carlo simulations (circles) and the model developed in Sec. 4.7 (lines).

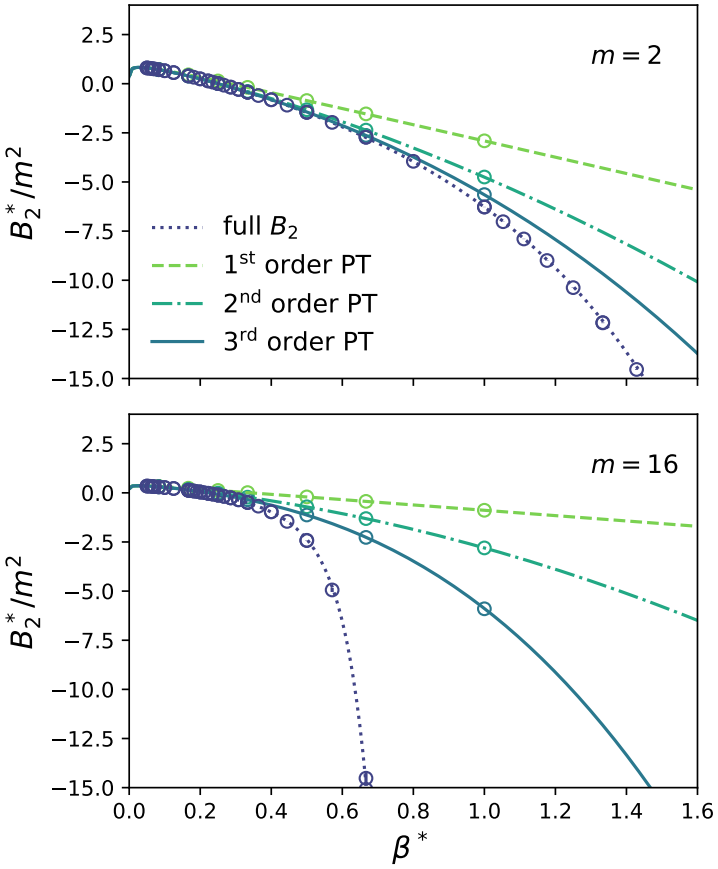


Figure 4.7: Comparison of the full second virial coefficient to predictions obtained from first-order, second-order and third-order BH perturbation theory (PT). Symbols are Monte Carlo simulation results, whereas lines are obtained from the model of Sec. 4.7.

4.6.2.3 Boyle Temperature, Single-chain Collapse and Polymer-solvent Phase Separation

The Boyle temperature T_B^* defines the temperature at which the second virial coefficient equals zero,

$$B_2(T_B^*) = 0 \tag{4.23}$$

At this temperature, the repulsive and attractive contributions to B_2 cancel. Several mean-field theories (e.g. Flory-Huggins theory[6–9] and the theory of Sanchez[11]) predict that the Boyle temperature T_B^*, the temperature of single-chain collapse T_θ^*, and the critical temperature for polymer-solvent phase separation T_c^* converge to the same value θ^* for very long polymers. Convincing evidence from either experiments or molecular simulations of off-lattice, 3-D

chain molecules is scarce however. Therefore we estimate the Boyle temperature based on our MC simulation data, extrapolate to infinite chain length, and compare the result to our estimate of the single-chain collapse temperature of infinitely long chains $T_\theta^{*,\infty} \approx 4.5$, and an estimate for the critical temperature of infinitely long LJ chains $T_c^{*,\infty} \approx 4.6$ obtained from the literature.[21]

To calculate the Boyle temperature, we fitted a fourth-order polynomial in β^* to the MC data for B_2 of freely-jointed LJ chains; the inverse of the Boyle temperature β_B^* was then obtained as one of the roots of the polynomial. Only data within the range $T^* = [2.5, 6]$ was used for this analysis, typically corresponding to a total of about 12 data-points per molecule type. The standard deviation of the inverse Boyle temperature $\Delta\beta_B^*$ was estimated by repeating this procedure 1500 times with each data-point for B_2 given a random perturbation according to a Gaussian distribution centred around that data-point and with width equal to the estimated standard error of that data-point. The uncertainty in the Boyle temperature T_B^* was then estimated by linear error propagation, using $\Delta T_B^* = \left|\partial T_B^*/\partial \beta_B^*\right| \Delta\beta_B^* = T_B^{*2}\Delta\beta_B^*$.

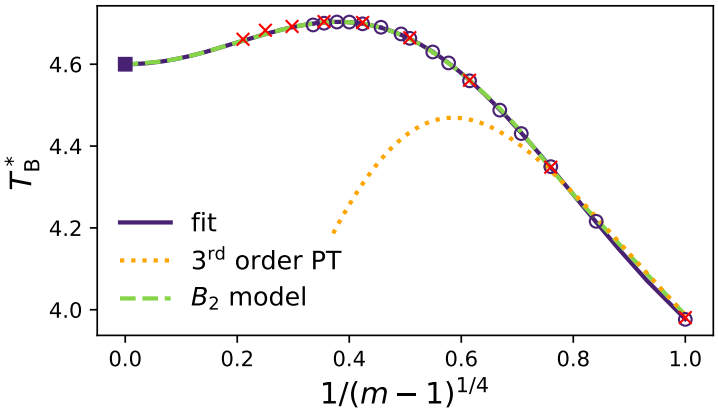


Figure 4.8: Boyle temperature T_B^* as obtained from Monte Carlo simulations of chains with up to 80 segments (circles), MSMC simulations of Schultz and Kofke[26] (crosses), the fit of eq. (4.24) (solid), the third-order BH perturbation theory of Sec. 4.7 (dotted) and the model for B_2 of Sec. 4.7 (dashed).

As shown in Fig. 4.8, our results for the Boyle temperature suggest an interesting trend. The Boyle temperature first increases with chain length, but for chains longer than about 50 segments it starts to decrease towards its value at infinite chain length $T_B^{*,\infty}$. If one would estimate the value of the Boyle temperature at infinite chain length from the data corresponding to 50 segments or less, this would lead to $T_B^{*,\infty} \approx 5.5$, which is much larger than the theta temperature $\theta^* \approx 4.5$ estimated in Sec. 4.6.1 and the VLE critical temperature

of infinitely long LJ chains $T_c^{*,\infty} \approx 4.6$ estimated by Sheng *et al.*[21]*. Only if the data for longer chains is included in the extrapolation, our prediction for $T_B^{*,\infty}$ might actually approach those values. Although the precise value of $T_B^{*,\infty}$ is difficult to estimate from our data (longer chain lengths need to be examined for this), a value close to 4.6 seems plausible to suggest.

We fitted our data for the Boyle temperature using the following equation

$$T_B^*(m) = \theta^* + \frac{a}{\sqrt{m}} + \frac{b}{m} + \frac{c}{m^{\frac{3}{2}}} + \frac{d}{m^2} + \frac{e}{m^{\frac{5}{2}}} \tag{4.24}$$

where we assumed $\theta^* = 4.6$. The above equation provides the Boyle temperature of freely-jointed LJ chains of any length, which will turn out useful for modelling the second virial coefficient in Sec. 4.7. The parameters are given by $a = 1.607$, $b = -7.444$, $c = 10.204$, $d = -9.781$, $e = 4.528$.

The non-monotonous behaviour of the Boyle temperature shown in Fig. 4.8 has been observed experimentally for e.g. polystyrene in decalin.[70] Prior molecular simulation studies of 3-D off-lattice Lennard-Jones polymers by Schultz and Kofke[26] and Schnabel and Janke[25] also predicted this behaviour. Both, approximate theoretical approaches (such as polymer perturbation theory and self-consistent field theory)[71] and molecular simulations of lattice models for polymers[72–74] show either a monotonous approach to T_B^∞ from above or below, while earlier molecular simulations based on 3-D off-lattice polymer models[18,19,21,23,24] either have been performed for too short chains and too high temperatures or are of too low statistical accuracy for the maximum to be observed. It seems that the qualitative features of how the Boyle temperature approaches its value at infinite chain length are sensitive to assumptions made in the polymer model, e.g. 2-D or 3-D, rigid or flexible bond-lengths/bond-angles,[24,26] linear or branched chains, and short or long-ranged interactions.[18]

4.7 Model for the Second Virial Coefficient of Freely-Jointed Lennard-Jones Chains

As shown in Figs. 4.7 and 4.8, the usefulness of liquid-state perturbation theory in describing the second virial coefficient of LJ chain molecules is limited. Although the perturbation

*We note that another estimate $T_c^{*,\infty} = 4.3$ based on more recent molecular simulation data for the VLE by Silmore and Panagiotopoulos[68] is available but that data has recently been questioned[69] due to not using any tail corrections to handle the long-ranged part of the LJ interactions.

Table 4.1: Model constants for the dimensionless correlation integrals I_n defined in eqs. (4.37) and (4.38), and analytic solutions of the mean-field constant $\alpha_{12-6,n}$ of eq. (4.34).

n	a_{n0}	a_{n1}	a_{n2}	a_{n3}
1	0.21144	-0.44187	-2.1353	5.2355
2	-2.7389	3.7530	26.715	-32.603
3	2.9748	-4.9832	670.40	-771.82
	b_{n0}	b_{n1}	b_{n2}	b_{n3}
1	-0.10426	0.06736	5.0746	-3.8369
2	2.6435	-2.1370	-37.511	33.732
3	-11.954	11.590	-338.51	240.06
	$n=1$	$n=2$	$n=3$	
$\alpha_{12-6,n}$	8/9	-128/315	1024/3465	

Table 4.2: Model constants for the empirical correction $\Delta B_2^{\mathrm{B}^*}$ defined in eqs. (4.39) and (4.40).

n	c_{n0}	c_{n1}
0	21.237	-27.918
1	-38.767	-14.730
2	85.121	363.80
3	-441.20	42.768

Table 4.3: Model constants for the empirical low temperature correction $\Delta B_2^{\mathrm{LT}^*}$ defined in Eqs. (4.41)-(4.42).

n	d_{n0}	d_{n1}	d_{n2}
0	0.4963	-2.1319	2.0046
1	7.8158	-30.068	22.809
2	0.12878	7.6629	-6.578
3	1779.7	-5999.0	4224.6
4	0.33082	2.4526	-1.9026

expansion of Eq. (4.16) converges rapidly at high temperatures (good solvent regime), at low temperatures (poor solvent regime) even a third-order perturbation theory can not accurately quantify B_2. Also, for chains of more than three segments, the Boyle temperature (theta regime) is significantly underestimated. Given the vastly different scaling of B_2 with chain length m in each solvent regime, we add two separate empirical corrections to address these deficiencies: one designed to improve the description of the Boyle temperature of chains (ΔB_2^{B}) and another one to improve the description at low temperatures (ΔB_2^{LT}),

$$B_2 = B_{20} + \Delta\tilde{B}_{21} + \Delta\tilde{B}_{22} + \Delta\tilde{B}_{23} + \Delta B_2^{\mathrm{B}} + \Delta B_2^{\mathrm{LT}} \tag{4.25}$$

Here, B_{20} is the second virial coefficient of the repulsive reference fluid and $\Delta\tilde{B}_{2n}$ are the perturbation contributions of order n (see section 4.4.2 for a definition). In the following paragraphs, we will develop an algebraic model for each term of Eq. (4.25).

4.7.1 Reference Contribution

At the level of the reference fluid, the segments of the chains interact solely by the positive (BH) part of the LJ potential (see Eq. (4.7)). Following previous work,[75,76] the second virial coefficient of two such BH chain molecules is mapped onto the second virial coefficient of two chains of tangentially bonded hard spheres (denoted as THSC) of effective diameter $d(T^*, m)$,

$$B_{20}^* = B_2^{\mathrm{THSC}*}(m) d^*(T^*, m)^3 \tag{4.26}$$

The second virial coefficient of THSC molecules is well studied.[75,77,78] As shown in a previous work,[76] the available molecular simulation data for B_2 of THSCs is described accurately by a simple modification to the second virial coefficient of two THSCs with all segments fixed in a linear, rigid conformation,[61,62] according to

$$B_2^{\mathrm{THSC}*} = \frac{\pi}{6} m \left(\frac{11m-3}{2m} + 1.3880 \frac{(m-1)^{2-\gamma_0}}{m} \right) \tag{4.27}$$

where

$$\gamma_0 = \gamma_0^\infty + \frac{0.6388}{m^3} \tag{4.28}$$

The parameter $\gamma_0^\infty = 0.2079$ defines the asymptotic scaling of the model for large values of the chain length, as $B_{20}/m^2 \sim m^{-\gamma_0^\infty}$. Its value is reasonably close to estimates from theory and simulations $\gamma_0^\infty = [0.2, 0.25]$ (see Yethiraj and Hall[60] and references therein). Note that

Eq. (4.27) in Ref.[76] (Eq. A1 in that paper) contains a misprint in the value of the constant 1.3880.

For BH monomers, the hard-sphere diameter $d(T^*, m = 1)$ can be calculated accurately using the model of van Westen[79]. We extend this model to BH chain molecules by introducing four additional parameters, which were adjusted to our molecular simulation data for B_{20}, comprising chain lengths $2 \leq m \leq 32$ and temperatures $1 \leq T^* \leq 20$. The resulting expression for the dimensionless hard-sphere diameter $d^* = d/\sigma$ is

$$d^*(T^*, m) = \left(1 + C_0 T^* + D \ln(1 + T^*) + C_4 T^{*2}\right)^{-\frac{C_5}{24}} \tag{4.29}$$

where only the coefficients C_0 and C_5 are modified with respect to the monomer model (for which $C_0 = C_5 = 1$), according to

$$\begin{aligned} C_0 &= 1 + c_0\left(\frac{m-1}{m}\right) + c_1\left(\frac{m-1}{m}\right)\left(\frac{m-2}{m}\right) + c_2\left(\frac{m-1}{m}\right)\left(\frac{m-2}{m}\right)\left(\frac{m-3}{m}\right) \\ C_5 &= 1 + c_3\left(\frac{m-1}{m}\right) \end{aligned} \tag{4.30}$$

with $\mathbf{c} = [0.12055, 0.32666, -0.90893, -0.48780]$.

The remaining coefficients are equivalent to those of the monomer model[79] and are here listed for completeness:

$$D = C_1 T^{*\frac{1}{4}} + C_2 T^{*\frac{3}{4}} + C_3 T^{*\frac{5}{4}} \tag{4.31}$$

with

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{pmatrix} = \begin{pmatrix} 0 & 0.010936 & 0 \\ -0.20090 & -0.012707 & 0 \\ 0.014042 & 0.073595 & 0.012846 \\ 0.0037153 & 0.0050538 & 0.049100 \end{pmatrix} \begin{pmatrix} 1 \\ \alpha' \\ \alpha'^2 \end{pmatrix} \tag{4.32}$$

$$\alpha' = \frac{1}{\alpha_{12-6,1}} - \frac{1}{\alpha_{7-6,1}} \tag{4.33}$$

and the dimensionless (Van-der-Waals) mean-field constant

$$\alpha_{\nu-6,n} = -\int_1^\infty \left(u^*_{\mathrm{Mie},\nu-6}(r^*)\right)^n r^{*2} \mathrm{d}r^* \tag{4.34}$$

where $u^*_{\mathrm{Mie},\nu-6}$ is the dimensionless Mie ν-6 pair potential, which for $\nu = 12$ is the LJ potential. For $n = 1$, this results in

$$\alpha_{\nu-6,1} = \frac{\nu}{\nu-6}\left(\frac{\nu}{6}\right)^{\frac{6}{\nu-6}}\left(\frac{1}{3} - \frac{1}{\nu-3}\right) \tag{4.35}$$

Results for $n = \{2, 3\}$ are given in table 4.1.

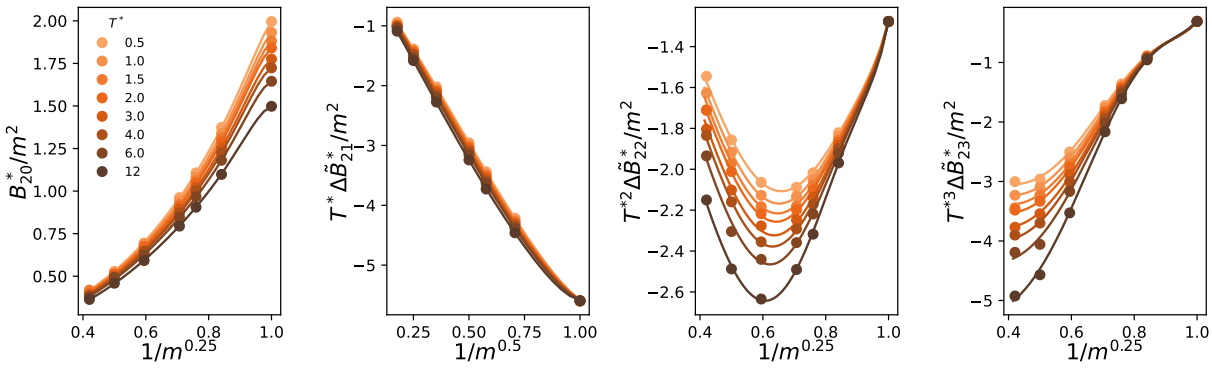


Figure 4.9: Scaling of the reference and perturbation contributions to the second virial coefficient with chain length m. Comparison of the models of eqs. (4.26)–(4.30) (lines) and MC simulation results (symbols) for chain lengths $m = \{1, 2, 3, 4, 8, 16, 32\}$.

4.7.2 Perturbation Contributions

The BH perturbation contributions to the second virial coefficient defined in eq. (4.17) can be written as

$$\Delta\tilde{B}^*_{2n} = \frac{(-1)^{n+1}}{2n!}\beta^{*n} I_n \tag{4.36}$$

where I_n define the dimensionless correlation integrals of order n. The latter are modelled by empirically extending the exactly known results for LJ monomers ($m = 1$) to chains,

$$I_n = 4\pi m^{2-\gamma_n^\infty}\left(-\alpha_{12-6,n} + A_n\frac{m-1}{m} + B_n\frac{m-1}{m}\frac{m-2}{m}\right) \tag{4.37}$$

using the following empirical coefficients $X_n \in \{A_n, B_n\}$

$$X_n = x_{n0} + x_{n1}m^{-\gamma_n^\infty} + (x_{n2} + x_{n3}m^{-\gamma_n^\infty})(d^* - 1) \tag{4.38}$$

where $x_{ni} \in \{a_{ni}, b_{ni}\}$ are adjustable model constants. The effective hard-sphere diameter $d^*(T^*, m)$ (defined in eq. (4.29)) appearing in this equation serves to model the temperature dependence of the perturbation terms. The parameter γ_n^∞ defines the asymptotic scaling of the perturbation term ΔB^*_{2n} with chain length for $m \to \infty$.

To determine the values of the coefficients γ_n^∞, we analyse the scaling of our Monte Carlo simulation results for $\Delta\tilde{B}_{2n}/m^2$ with chain length m in figure 4.9. For the first order perturbation contribution, we observe $\Delta\tilde{B}_{21}/m^2 \sim m^{-0.5}$ and thus apply $\gamma_1^\infty = 0.5$. The second-order perturbation term shows a reversal in scaling with m as m increases: $\Delta\tilde{B}_{22}/m^2$ increases with m for $m \lesssim 8$ and decreases with m for $m > 8$. At least for temperatures close to the Boyle temperature, we expect this reversal in the dependence on chain length to take place for all perturbation terms; otherwise, no finite Boyle temperature of long chains can be predicted from an infinite-order perturbation theory (if any of the perturbation terms $\Delta B^*_{2n}/m^2$ is non-zero for $m \to \infty$, the Boyle temperature would diverge to positive infinity). Although our MC simulation results for the third-order perturbation term are inconclusive about this point, we expect the chains considered in our simulations ($m \leq 32$) are simply too short to fully reveal the same trend as observed for the second-order perturbation term. For practical reasons, we assume the same asymptotic scaling of the second- and third-order perturbation contribu-

tions with chain length as for the reference contribution B_{20}/m^2, specifically $\gamma_2^\infty = \gamma_3^\infty = \gamma_0^\infty$. This ensures a finite, non-zero Boyle temperature already at the level of a second-order, or third-order perturbation theory.

The model constants x_{ni} of eq. (4.38) are adjusted to our Monte Carlo simulation data for ΔB_{2n}^*, which comprises chain lengths $2 \leq m \leq 32$ and temperatures $0.5 \leq T^* \leq 12$. The values of the fitted constants are provided in table 4.1.

4.7.3 Empirical Correction

The first empirical correction is designed to improve the prediction of the Boyle temperature for chains with $m > 2$. We apply the Ansatz function

$$\frac{\Delta B_2^{\mathrm{B}^*}}{m^2} = \beta^{*3} m^{-\gamma_0^\infty} \left(\frac{m-1}{m}\right)\left(\frac{m-2}{m}\right)\left(C_1 + \frac{C_2}{m^{\frac{1}{4}}} + \frac{C_3}{\sqrt{m}}\right)\tanh(C_0 m\beta^*) \tag{4.39}$$

with

$$C_n = c_{n0} + c_{n1} m^{-\gamma_0^\infty} \tag{4.40}$$

The scaling $\Delta B_2^{\mathrm{B}^*}/m^2 \sim m^{-\gamma_0^\infty}$ imposed by this equation ensures that the model predicts a finite Boyle temperature for $m \to \infty$. The 8 model constants c_{ij} were obtained by minimizing the absolute relative deviations between the Boyle temperature predicted by the virial coefficient model of Eq. (4.25) (with $\Delta B_2^{\mathrm{LT}} = 0$) and the fit to MC simulation results of eq. (4.24). Values of the correlated constants are provided in table 4.2.

The second empirical correction addresses the model's performance at temperatures below the Boyle temperature. This correction must be zero at the Boyle temperature and very close to zero for temperatures $T^* > T_{\mathrm{B}}^*$. At low temperatures, the correction should capture the exponential scaling of B_2 with m (see section 4.6.2.1). We choose the following Ansatz function, which is loosely inspired by the functional form of a cluster expansion of the Mayer-f function of two chains,[2,18]

$$\frac{\Delta B_2^{\mathrm{LT}*}}{m^2} = \tanh^2\left(\frac{D_0}{m}\beta^*\right)\left\{-D_1\left(\exp\left[mD_2\left(\beta^* - \beta^*_{\mathrm{Boyle}}(m)\right)\right]-1\right) - \frac{m-1}{m}D_3\left(\exp\left[mD_4\left(\beta^* - \beta^*_{\mathrm{Boyle}}(m)\right)\right]-1\right)\right\} \tag{4.41}$$

where the inverse Boyle temperature $\beta^*_{\mathrm{Boyle}}(m)$ is obtained from the model of eq. (4.24). The empirical coefficients are described by the following equation

$$D_i = d_{i0} + \frac{d_{i1}}{\sqrt{m}} + \frac{d_{i2}}{m^{3/4}} \tag{4.42}$$

The 15 model constants d_{ij} were used to correlate our Monte Carlo simulation results for B_2, comprising $0.25 \leq T^* \leq 1000$ for $m = 1$ and $0.2 \leq T^* \leq 20$ for $2 \leq m \leq 128$. See table 4.3 for the correlated constants.

4.7.4 Model Evaluation

In figure 4.10, results obtained from our model for B_2^* are compared to simulation data. The left plot shows results for temperatures near and above the Boyle temperature while the right diagram additionally includes results for lower temperatures within the poor solvent regime. The proposed model accurately captures the different scaling regimes of B_2^*. For the high temperature data (left), the absolute deviations of the model results for B_2^*/m^2 to the reference data are all below 0.07. The model captures the Boyle temperature accurately for the chain lengths shown. This is also illustrated in figure 4.8 where the Boyle temperature predicted by the model (dashed lines) is compared to the Boyle temperature from simulations (symbols) and the fit of eq. (4.24) (solid line). The two lines are almost indistinguishable and converge to the same value $\theta^* = 4.6$ for $m \to \infty$. Slight deviations occur for $2 \leq m \leq 3$. In this range, the Boyle temperature is mainly determined by the perturbation terms, as the magnitude of the empirical correction $\Delta B_2^{\mathrm{B}*}$ is small compared to that of the perturbation terms.

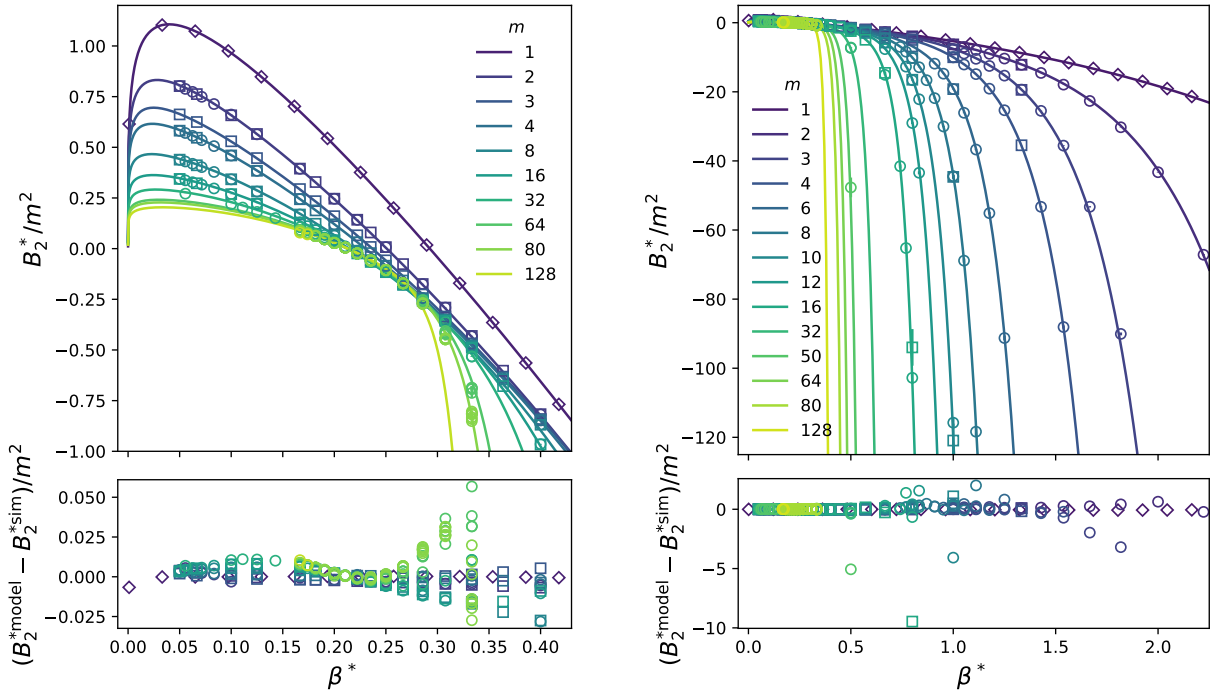


Figure 4.10: Comparison between model predictions (eq. (4.25), solid lines) and simulation results (symbols) for the second virial coefficient of chains comprising m freely-jointed LJ segments. Simulations comprise results from numerical integration ($m = 1$, diamonds), MC simulations with Zeno (circles), and NVT-MC simulations (squares). Left: B_2^*/m^2 evaluated for temperatures close to or above the Boyle temperature. Right: Results and absolute deviations for $B_2^*/m^2 \geq -120$.

4.8 Conclusion

The molecular simulation data for the second virial coefficient and single-chain properties of Lennard-Jones chains generated in this work provide evidence that the Boyle temperature T_B, the temperature for single-chain collapse T_θ, and the upper critical solution temperature for polymer-solvent phase separation T_c of 3-D off-lattice chains converge to the same value if the number of segments in the chain molecules becomes very large. This confirms the intimate link between the conformational properties of single chains and polymer-solvent phase behaviour predicted by mean-field theories of polymers on a lattice, such as Flory-Huggins theory.[6–9] Somewhat counterintuitively, our simulation data (especially those of Lennard-Jones chains with repulsive intramolecular interactions) also reveal that the actual effect of changes in chain conformation (due to single-chain collapse) on the second virial coefficient is small, even for temperatures quite far below the Boyle temperature. Lennard-Jones chains with repulsive intramolecular interactions do not exhibit a coil-globule transition; nevertheless, the second virial coefficient of such chains displays all three scaling regimes: good solvent, theta

solvent and poor solvent. This does not mean the single-chain collapse temperature of real polymers can not be used as a predictor for polymer-solvent phase equilibria (see e.g. Zeng *et al.* [80]). However, it suggests that the modelling of the free energy of polymers (in solution), and, as a result, polymer(-solvent) phase equilibria, might be done based on molecular models that do not undergo single-chain collapse. Specifically, a molecular model in which only the segments of different chains attract each other, while the segments within the same chain are repulsive, might be sufficient, thereby considerably simplifying the development of models for the free energy of polymers (in solution). The analytic model for the second virial coefficient of LJ chains proposed in this work will further benefit such developments.

References

[1] P. J. Flory. *Principles of polymer chemistry*. Cornell University Press, Ithaca, New York, 1953.

[2] H. Yamakaw. *Modern theory of polymer solutions*. Harper & Row, New York, 1971.

[3] J. des Cloizeaux and G. Jannink. *Polymers in solution*. Oxford University Press, Oxford, 1953.

[4] M. Doi. *Introduction to Polymer Physics*. Clarendon Press, Oxford, 1996.

[5] M. Rubinstein and R. H. Colby. *Polymer physics*. Oxfor University Press, Oxford, 2003.

[6] P. J. Flory. Thermodynamics of High Polymer Solutions. *J. Chem. Phys.*, **9**(8):660–660, 1941. doi:10.1063/1.1750971.

[7] P. J. Flory. Thermodynamics of High Polymer Solutions. *J. Chem. Phys.*, **10**(1):51–61, 1942. doi:10.1063/1.1723621.

[8] M. L. Huggins. Solutions of Long Chain Compounds. *J. Chem. Phys.*, **9**(5):440–440, 1941. doi:10.1063/1.1750930.

[9] M. L. Huggins. THERMODYNAMIC PROPERTIES OF SOLUTIONS OF LONG-CHAIN COMPOUNDS. *Annals of the New York Academy of Sciences*, **43**(1):1–32, 1942. doi:https://doi.org/10.1111/j.1749-6632.1942.tb47940.x.

[10] C. Williams, F. Brochard, and H. L. Frisch. Polymer collapse. *Ann. Rev. Phys. Chem.*, **32**:433, 1981.

[11] I. C. Sanchez. Phase transition behavior of the isolated polymer chain. *Macromolecules*, **12**:980–987, 1979.

[12] G. Luna-Bárcenas, J. C. Meredith, I. C. Sanchez, K. P. Johnston, D. G. Gromov, and J. J. de Pablo. Relationship between polymer chain conformation and phase boundaries in a supercritical fluid. *J. Chem. Phys.*, **107**(24):10782–10792, 1997. doi:10.1063/1.474194.

[13] J. McMillan, William G. and J. E. Mayer. The Statistical Thermodynamics of Multicomponent Systems. *J. Chem. Phys.*, **13**(7):276–305, 1945. doi:10.1063/1.1724036.

[14] M. P. Taylor, Y. Ye, and S. R. Adhikari. Conformation of a flexible polymer in explicit solvent: Accurate solvation potentials for Lennard-Jones chains. *J. Chem. Phys.*, **143**(20):204901, 2015. doi:10.1063/1.4935952.

[15] T. L. Hill. *Statistical Mechanics*. McGraw-Hill, New York, third edition, 1956.

[16] S. Vafaei, B. Tomberli, and C. G. Gray. McMillan-Mayer theory of solutions revisited: Simplifications and extensions. *J. Chem. Phys.*, **141**(15):154501, 2014. doi:10.1063/1.4897980.

[17] A. Baumgartner. Statics and dynamics of the freely jointed polymer chain with Lennard-Jones interaction. *J. Chem. Phys.*, **72**:871, 1980.

[18] J. M. Wichert and C. K. Hall. Second virial coefficient calculations for square-well chain molecules. *Macromolecules*, **27**:2744–2756, 1994.

[19] J. Dautenhahn and C. K. Hall. Monte Carlo Simulation of Off-Lattice Polymer Chains: Effective Pair Potentials in Dilute Solution. *Macromolecules*, **27**:5399–5412, 1994.

[20] V. I. Harismiadis and I. Szleifer. Second virial coefficients of chain molecules: a Monte Carlo study. *Mol. Phys.*, **81**:851–856, 1994.

[21] Y. Sheng, A. Z. Panagiotopoulos, S. K. Kumar, and I. Szleifer. Monte Carlo calculation of phase equilibria for a bead-spring polymeric model. *Macromolecules*, **27**:400–406, 1994.

[22] C. Vega, S. Lago, and B. Garzón. Linear hard sphere models. Virial coefficients and equation of state. *Mol. Phys.*, **82**:1233–1247, 1994.

[23] Y. C. Chiew and V. Sabesan. Second virial coefficients of Lennard-Jones chains. *Fluid Phase Equilib.*, **155**:75–83, 1999.

[24] I. M. Withers, A. V. Dobrynin, M. L. Berkowitz, and M. Rubinstein. Monte Carlo simulation of homopolymer chains. I. Second virial coefficient. *J. Chem. Phys.*, **118**:4721, 2003.

[25] S. Schnabel and W. Janke. Collapse transition of a Lennard-Jones polymer. *Phys. Rev. E*, **108**:064502, 2023. doi:10.1103/PhysRevE.108.064502.

[26] A. J. Schultz and D. A. Kofke. Methodical evaluation of Boyle temperatures using Mayer sampling Monte Carlo with application to polymers in implicit solvent. *J. Chem. Phys.*, **161**(15):154108, 2024. doi:10.1063/5.0227411.

[27] M. P. Taylor. Square-well diatomics Exact low density results. *Mol. Phys.*, **82**(6):1151–1164, 1994. doi:10.1080/00268979400100814.

[28] M. P. Taylor, W. Paul, and K. Binder. Phase transitions of a single polymer chain: A Wang–Landau simulation study. *J. Chem. Phys.*, **131**(11):114907, 2009. doi:10.1063/1.3227751.

[29] Y. Zhou, M. Karplus, J. M. Wichert, and C. K. Hall. Equilibrium thermodynamics of homopolymers and clusters: Molecular dynamics and Monte Carlo simulations of systems with square-well interactions. *J. Chem. Phys.*, **107**(24):10691–10708, 1997. doi:10.1063/1.474186.

[30] M. P. Taylor. Collapse transition of isolated Lennard-Jones chain molecules: Exact results for short chains. *J. Chem. Phys.*, **114**(14):6472–6484, 2001. doi:10.1063/1.1350578.

[31] J. Gross and G. Sadowski. Perturbed-Chain SAFT: an equation of state based on a pertubation theory for chain molecules. *Ind. Eng. Chem. Res.*, **40**:1244–1260, 2001.

[32] T. van Westen and J. Gross. Accurate thermodynamics of simple fluids and chain fluids based on first-order perturbation theory and second virial coefficients: uv-Theory. *J. Chem. Phys.*, **155**:244501, 2021.

[33] H. G. Honnell and C. K. Hall. A new equation of state for athermal chains. *J. Chem. Phys.*, **90**:1841, 1989.

[34] M. S. Wertheim. Fluids with highly directional attractive forces. IV. Equilibrium polymerization. *J. Stat. Phys.*, **42**:477–492, 1986.

[35] M. S. Wertheim. Thermodynamic perturbation theory of polymerization. *J. Chem. Phys.*, **87**:7323, 1987.

[36] G. Jackson, W. G. Chapman, and K. E. Gubbins. Phase equilibria of associating fluids: spherical molecules with multiple bonding sites. *Mol. Phys.*, **65**:1–31, 1988.

[37] W. G. Chapman, G. Jackson, and K. E. Gubbins. Phase equilibria of associating fluids: chain molecules with multiple bonding sites. *Mol. Phys.*, **65**:1057–1079, 1988.

[38] W. G. Chapman, K. E. Gubbins, G. Jackson, and M. Radosz. New reference equations of state for associating liquids. *Ind. Eng. Chem. Res.*, **29**:1709–1721, 1990.

[39] C. McCabe and A. Galindo. SAFT associating Fluids and Fluid Mixtures. In A. R. H. Goodwin, J. V. Sengers, and C. J. Peters (editors), *Applied Thermodynamics of Fluids*, chapter 8, pages 215–279. The Royal Society of Chemistry, Cambride, 2010.

[40] D. A. McQuarrie. *Statistical Mechanics*. Harper & Row, New York, 1976.

[41] T. L. Hill. *An introduction to statistical thermodynamics*. Dover Publications, New York, third edition, 1986.

[42] S. Caracciolo, B. M. Mognetti, and A. Pelissetto. Virial coefficients and osmotic pressure in polymer solutions in good-solvent conditions. *J. Chem. Phys.*, **125**(9):094903, 2006. doi:10.1063/1.2338913.

[43] T. van Westen and J. Gross. Accurate first-order perturbation theory for fluids: uf-Theory. *J. Chem. Phys.*, **154**:041102, 2021.

[44] T. van Westen, M. Hammer, B. Hafskjold, A. Aasen, J. Gross, and O. Wilhelmsen. Perturbation theories for fluids with short-ranged attractive forces: A case study of the Lennard-Jones spline fluid. *J. Chem. Phys.*, **156**:104504, 2022.

[45] A. Reimer, T. van Westen, and J. Gross. Physically based equatin of state for Mie ν-6 fluids. *J. Chem. Phys.*, **158**:164506, 2022.

[46] J. A. Barker and D. Henderson. Perturbation theory and equation of state for fluids. II. A succesful

theory of liquids. *J. Chem. Phys.*, **47**:4714–4721, 1967.

[47] D. Chandler and J. D. Weeks. Equilibrium structure of simple liquids. *Phys. Rev. Lett.*, **25**:149–151, 1970.

[48] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.*, **54**:5237, 1971.

[49] H. C. Andersen, J. D. Weeks, and D. Chandler. Relationship between the hard-sphere fluid and fluids with realistic repulsive forces. *Phys. Rev. A*, **4**:1597–1606, 1971.

[50] H. C. Andersen, D. Chandler, and J. D. Weeks. Roles of repulsive and attractive forces in liquids: the equilibrium theory of classical fluids. *Advances in Chemical Physics*, **34**:105, 1976.

[51] R. S. Berry, T. L. Beck, H. L. Davis, and J. Jellinek. Solid-liquid phase behavior in microclusters. In I. Prigogine and S. A. Rice (editors), *Evolution of size effects in chemical dynamics Part 2*, chapter 3, pages 75–138. Wiley, New York, 1988.

[52] D. D. Frantz. Magic numbers for classical Lennard-Jones cluster heat capacities. *J. Chem. Phys.*, **102**(9):3747–3768, 1995. doi:10.1063/1.468557.

[53] M. N. Rosenbluth and A. W. Rosenbluth. Monte-Carlo calculation of the average extension of molecular chains. *J. Chem. Phys.*, **23**(2):356–359, 1955.

[54] D. Frenkel and B. Smit. Unexpected length dependence of the solubility of chain molecules. *Mol. Phys.*, **75**(5):983–988, 1992.

[55] D. Frenkel, G. C. A. M. Mooij, and B. Smit. Novel scheme to study structural and thermal properties of continuously deformable molecules. *J. Phys. Condens. Matter*, **4**(12):3053–3076, 1992.

[56] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Academic Press, San Diego, 2nd edition, 2002.

[57] D. Juba, D. J. Audus, M. Mascagni, J. F. Douglas, and W. Keyrouz. ZENO: Software for calculating hydrodynamic, electrical, and shape properties of polymer and particle suspensions. *J. Res. Natl. Inst. Stand. Technol*, **122**(20):10–6028, 2017.

[58] A. Bansal, A. J. Schultz, J. F. Douglas, and D. A. Kofke. Probabilistic computations of virial coefficients of polymeric structures described by rigid configurations of spherical particles: A fundamental extension of the ZENO program. *J. Chem. Phys.*, **157**(22):224801, 2022. doi:10.1063/5.0127465.

[59] M. P. Taylor, W. Paul, and K. Binder. Applications of the Wang-Landau algorithm to phase transitions of a single polymer chain. *Polym. Sci. Ser. C*, **55**:23–38, 2013. doi:10.1134/S1811238213060040.

[60] A. Yethiraj and C. K. Hall. Generalized Flory equations of state for square-well chains. *J. Chem. Phys.*, **95**:8494, 1991.

[61] D. C. Williamson and G. Jackson. Excluded volume for a pair of linear chains of tangent hard spheres with an arbitrary relative orientation. *Mol. Phys.*, **86**:819–836, 1995.

[62] K. M. Jaffer, S. B. Opps, and D. E. Sullivan. The nematic-isotropic phase transition in linear fused hard-sphere chain fluids. *J. Chem. Phys.*, **110**:11630, 1999.

[63] T. van Westen, T. J. H. Vlugt, and J. Gross. An analytical approximation for the orientation-dependent excluded volume of tangent hard sphere chains of arbitrary chain length and flexibility. *J. Chem. Phys.*, **37**:044906, 2012.

[64] G. Raos and G. Allegra. Chain Interactions in Poor-Solvent Polymer Solutions: Equilibrium and Nonequilibrium Aspects. *Macromolecules*, **29**(20):6663–6670, 1996. doi:10.1021/ma960503i.

[65] S. T. Sun, I. Nishio, G. Swislow, and T. Tanaka. The coil–globule transition: Radius of gyration of polystyrene in cyclohexane. *J. Chem. Phys.*, **73**:5971, 1980.

[66] I. H. Park, Q.-W. Wang, and B. Chu. Transition of Linear Polymer Dimensions from θ to Collapsed Regime. 1. Polystyrene/Cyclohexane System. *Macromolecules*, **20**:1965, 1987.

[67] M. Nierlich, J. P. Cotton, and B. Farnoux. Observation of the collapse of a polymer chain in poor solvent by small angle neutron scattering. *J. Chem. Phys.*, **69**:1379, 1980.

[68] M. P. H. Kevin S. Silmore and A. Z. Panagiotopoulos. Vapour–liquid phase equilibrium and surface tension of fully flexible Lennard–Jones chains. *Mol. Phys.*, **115**(3):320–327, 2017. doi:10.1080/00268976.2016.1262075.

[69] E. L. Granados-Bazán, S. E. Quiñones-Cisneros, and U. K. Deiters. Interfacial properties of binary mixtures of Lennard-Jones chains in planar interfaces by molecular dynamics simulation. *J. Chem. Phys.*, **154**(8):084704, 2021. doi:10.1063/5.0042340.

[70] Y. Nakamura, N. Inoue, T. Norisuye, and A. Teramoto. Ternary Cluster Integral for Polystyrene in trans-Decalin near the θ Point. *Macromolecules*, **30**(3):631–636, 1997. doi:10.1021/ma961222d.

[71] P. Zhang, N. M. Alsaifi, and Z.-G. Wang. Revisiting the θ Point. *Macromolecules*, **53**(23):10409–10420, 2020. doi:10.1021/acs.macromol.0c01317.

[72] M. Janssens and A. Bellemans. On the Osmotic Second Virial Coefficient of Polymer Solutions. *Macromolecules*, **9**(2):303–307, 1976. doi:10.1021/ma60050a024.

[73] W. Bruns. The Ideal and the Pseudoideal State of Macromolecules: A Comparison. *Macromolecules*, **17**:2826–2830, 1984.

[74] A. Z. Panagiotopoulos, V. Wong, and M. A. Floriano. Phase Equilibria of Lattice Polymers

from Histogram Reweighting Monte Carlo Simulations. *Macromolecules*, **31**(3):912–918, 1998. doi:10.1021/ma971108a.

[75] T. van Westen, T. J. H. Vlugt, and J. Gross. On the vapor-liquid equilibrium of attractive chain fluids with variable degree of molecular flexibility. *J. Chem. Phys.*, **142**:224504, 2015.

[76] T. van Westen, P. Rehner, T. J. H. Vlugt, and J. Gross. Generic low-density corrections to the equation of state of chain molecules with repulsive intermolecular forces. *J. Chem. Phys.*, **160**(17):174105, 2024. doi:10.1063/5.0197910.

[77] C. Vega, J. M. Labaig, L. G. MacDowell, and E. Sanz. The virial coefficients of the pearl-necklace model. *J. Chem. Phys.*, **113**(22):10398–10409, 2000. doi:10.1063/1.1322637.

[78] A. Yethiraj, K. G. Honnell, and C. K. Hall. Monte Carlo calculation of the osmotic second virial coefficieint of off-lattice athermal polymers. *Macromolecules*, **25**:3979–3983, 1992.

[79] T. van Westen. Algebraic second virial coefficient of the Mie m-6 intermolecular potential based on perturbation theory. *J. Chem. Phys.*, **154**:234502, 2021.

[80] X. Zeng, A. S. Holehouse, A. Chilkoti, T. Mittag, and R. V. Pappu. Connecting Coil-to-Globule Transitions to Full Phase Diagrams for Intrinsically Disordered Proteins. *Biophysical journal*, **119**:402–418, 2000.

5 Beyond One-Fluid Approximations for the Thermodynamics of Fluid Mixtures

The content of this chapter is a literal quote of the publication:

A. Reimer, T. Winkler-Markert and J. Gross. Beyond One-Fluid Approximations for the Thermodynamics of Fluid Mixtures. *J. Chem. Phys.*, **164**, 014117 (2026), doi:10.1063/5.0307122

5.1 Abstract

Developing analytic equations of state for fluid mixtures based on perturbation theories requires simplifying approximations, in which mixture properties are determined from effective pure component properties. While these one-fluid approximations reduce model complexity, they also introduce inaccuracies. In this work, a simple algebraic correction factor for perturbation theories applying a one-fluid approximation is proposed. The correction factor is defined as the ratio of the rigorous first-order perturbation term in Helmholtz energy to the respective first-order perturbation term in the one-fluid approximation. An approximate closed-form expression for the correction factor is derived in terms of the fundamental measures of the particles using symbolic regression. The new approach is thoroughly evaluated by applying it to a range of equations of state, including the uv-theory and the PCP-SAFT equation of state, to predict the thermodynamic properties of square-well, Lennard–Jones, and real-substance mixtures. New molecular simulation data for strongly size-asymmetric binary, ternary, and quinary Lennard–Jones mixtures is generated to test the proposed approach. Compared to the predictions obtained from a one-fluid approach, the correction factor significantly improves the accuracy of predicted phase equilibria and thermodynamic properties for model fluids. Its impact on real-fluid predictions with the PCP-SAFT equation of state is rather small, because PCP-SAFT segment size parameters of most substances are rather similar whereas the correction factor primarily accounts for segment size asymmetry.

5.2 Introduction

Equations of state (EoS) are powerful tools for the development and optimization of processes in the chemical and pharmaceutical industries, as they enable accurate predictions of thermodynamic properties at low computational cost. A particular strength of EoS is their ability to predict the phase behaviour of multicomponent systems, which is essential for various applications, such as thermodynamic separation processes. EoS can be roughly divided into two categories: physically based models and fully empirical models.[1]

Fully empirical models usually comprise a large number of adjustable parameters that are fitted to experimental data. These models are often intended to describe a specific substance with high accuracy. Examples include the water model of Wagner and Pruß[2] or the models for nitrogen[3,4] and carbon dioxide[5] developed by Roland Span and collaborators. Extending empirical EoS to other substances or mixtures is not straightforward and often requires the introduction of a significant amount of additional parameters. One example of such an extended model is the GERG-2008 equation of state[6], which is parametrized to describe 21 different substances and their mixtures.

Physically based EoS are often less accurate than fully empirical models, but offer a more rigorous treatment of mixtures and a wider range of applications, for example the ability to describe molecules with different sizes, shapes, and interaction strengths within a single framework.[7] Physically based models usually depend on a small number of empirical parameters and can provide reasonable predictions, even for substances with limited experimental data.

A popular class of physically based EoS are those based on perturbation theories[8–12], including the statistical associating fluid theory (SAFT)[13] and its variants (e.g., SAFT-VR-Mie[14] and PC-SAFT[15]). More recent advances include the uf-theory[16] and the uv-theory[7,17]. In perturbation theories, the intermolecular potential is divided into a reference and a perturbation part, and the Helmholtz energy is formulated as a series expansion about the reference fluid. Each perturbation term in this series is defined in terms of the structural properties of the reference fluid only. For fluid mixtures, a rigorous description of the perturbation terms requires the structural properties for all combinations of species in the reference-fluid mixture. Models for structural properties in mixtures, in particular multi-particle correlation functions, are scarce. Moreover, even when such models are available, evaluating all integrals occurring in the series expansion is computationally expensive.
To avoid these difficulties, the standard approach for perturbation terms is to develop models for pure fluids first and extend them to mixtures using one-fluid approximations.[11,18,19] In

these approaches, the mixture is treated as an effective pure fluid, characterized by interaction parameters obtained from the composition of the mixture and the interaction parameters of the respective pure components, through mixing rules.[20] For models with a weak theoretical foundation, these mixing rules are empirical, whereas for models with well established basis in statistical mechanics, the mixing rules are often obvious. The impact of applying a one-fluid approximation (OFA) on the accuracy of equation-of-state predictions is shown by van Westen and Gross[16,17]. The authors compute vapour–liquid equilibria of binary Lennard-Jones mixtures using the uv-theory[17] and the uf-theory[16], and compare results obtained with a more rigorous mixture treatment based on the pair-correlation function of a hard sphere mixture with those obtained using the OFA. They show that the results obtained with the rigorous mixture treatment agree significantly better with data from molecular simulation than the results obtained using the OFA. Similarly, Galliéro *et al.*[21] analysed the limitations of the van der Waals one-fluid model for predicting the viscosity of Lennard–Jones mixtures and found that, while the approximation performs well for energy asymmetry, it fails for size-asymmetric systems in dense or low-temperature states.

In this work, an algebraic correction factor for describing mixtures with perturbation theories is developed. This correction factor partially compensates for the loss of accuracy associated with the OFA while retaining its low computational cost. The correction factor is defined as the ratio of Helmholtz energies from two first-order perturbation terms: one with a full mixture description and the other one based on the OFA. These terms are evaluated numerically to generate reference data. Symbolic regression is employed to identify an approximate algebraic function. The resulting function can be used as a correction factor to the first-order perturbation terms of OFA-based EoS to improve their accuracy.

The correction factor is developed for mixtures of particles interacting with a square-well potential, using a hard-sphere mixture as reference system. For this system, the correction factor is independent of temperature, making it particularly convenient for an initial application. The resulting simple analytic function significantly improves property predictions across a broad range of binary square-well mixtures, including strongly asymmetric cases.

We assess the generality and transferability of the correction factor derived for square-well mixtures by incorporating it into various Lennard-Jones EoS, including the PC-SAFT equation of state[15], and the uv-theory[17]. Pressures, internal energies, and vapour-liquid equilibria calculated with the (corrected) EoS are compared to results from molecular simulation for a broad range of binary, ternary and quinary Lennard-Jones mixtures. In most cases, the correction factor improves the equation of state predictions. In addition, the influence of the correction on phase equilibrium predictions for binary mixtures of real substances calculated with the PCP-SAFT equation of state[15,22,23] is examined.

5.3 Molecular Model

The square-well potential between two particles of species i and j at a relative distance r is defined as

$$u_{ij}(r) = \begin{cases} \infty & \text{for } r < \sigma_{ij} \\ -\varepsilon_{ij} & \text{for } \sigma_{ij} \leq r < \lambda^{\mathrm{sw}}\sigma_{ij} \\ 0 & \text{for } r \geq \lambda^{\mathrm{sw}}\sigma_{ij} \end{cases} \tag{5.1}$$

with the size parameter σ_{ij}, the energy parameter ε_{ij} and the dimensionless potential width λ^{sw}. We consider only $\lambda^{\mathrm{sw}} = 1.5$ in this study, noting that the properties of square-well fluids for this value of λ^{sw} roughly resemble the properties of Lennard-Jones fluids of analogous parameters (ε_{ij}, σ_{ij}). The interaction parameters of two particles of different species are defined through Lorentz-Berthelot combining rules[24,25], as

$$\begin{aligned} \sigma_{ij} &= \frac{\sigma_{ii} + \sigma_{jj}}{2} \\ \varepsilon_{ij} &= \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \end{aligned} \tag{5.2}$$

The square-well pair potential is divided into a hard-sphere (hs) potential,

$$u_{0,ij}(r) = u_{ij}^{\mathrm{hs}}(r) = \begin{cases} \infty & \text{for } r < \sigma_{ij} \\ 0 & \text{for } r \geq \sigma_{ij} \end{cases} \tag{5.3}$$

and the corresponding perturbation potential,

$$w_{ij}(r) = u_{ij}(r) - u_{0,ij}(r) \tag{5.4}$$

In sections 5.4.2, 5.5, and 5.7, mixtures of particles interacting with the Lennard-Jones (LJ) potential

$$u_{ij}^{\mathrm{LJ}}(r) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{5.5}$$

are investigated. Two different divisions of the LJ potential into a reference contribution and a perturbation contribution are considered. The Weeks-Chandler-Andersen[26–28] (WCA)

division is defined by the reference potential

$$u_{0,ij}^{\mathrm{WCA}}(r) = \begin{cases} u_{ij}(r) + \varepsilon_{ij} & \text{for } r < 2^{1/6}\sigma_{ij} \\ 0 & \text{for } r \geq 2^{1/6}\sigma_{ij} \end{cases} \tag{5.6}$$

and the Barker-Henderson[29] (BH) division is defined by

$$u_{0,ij}^{\mathrm{BH}}(r) = \begin{cases} u_{ij}(r) & \text{for } r < \sigma_{ij} \\ 0 & \text{for } r \geq \sigma_{ij} \end{cases} \tag{5.7}$$

For both divisions, the corresponding perturbation potentials are given by

$$w_{ij}(r) = u_{ij}^{\mathrm{LJ}}(r) - u_{0,ij}(r) \tag{5.8}$$

5.4 Derivation of the Correction Factor

A system with volume V, temperature T and a total number of particles N, comprising N^{c} distinct species is considered. The composition of the mixture is defined by the mole fractions $x_i = N_i/N$, where N_i denotes the number of particles of species i.

Assuming a u-expansion[9,30], the reduced Helmholtz energy can be described based on the following series expansion,

$$a = \frac{\beta A}{N} = a_0 + \Delta a_1 + \Delta a_2 + \Delta a_3 + \cdots \tag{5.9}$$

where a_0 is the Helmholtz energy of the reference fluid of intermolecular potential $u_{0,ij}(r)$ (see eq. (5.3)), and Δa_1, Δa_2, and Δa_3 are perturbation terms describing the effect of the perturbation potential $w_{ij}(r)$ (eq. (5.4)) on the Helmholtz energy.

For fluid mixtures, the first-order perturbation term is defined rigorously as

$$\Delta a_1 = 2\pi\rho\beta \sum_{i=1}^{N^{\mathrm{c}}} \sum_{j=1}^{N^{\mathrm{c}}} x_i x_j \int_0^{\infty} w_{ij}(r) g_{0,ij}(r) r^2 \mathrm{d}r \tag{5.10}$$

where $g_{0,ij}(r)$ is the radial distribution function (RDF) of a particle of species i and a particle of species j in the reference fluid. Note that the RDF depends not only on the distance between the particles but also on temperature, density, composition, and the pair potentials of all

components. For brevity these dependencies are not written explicitly here. For particles interacting with a square-well or LJ potential eq. (5.10) can be expressed as

$$\Delta a_1 = 2\pi\rho\beta \sum_{i=1}^{N^c}\sum_{j=1}^{N^c} x_i x_j \varepsilon_{ij}\sigma_{ij}^3 I_{ij} \tag{5.11}$$

with the dimensionless correlation integrals

$$I_{ij} = \int_0^\infty g_{0,ij}(r^*) w^*(r^*) r^{*2} \mathrm{d}r^* \tag{5.12}$$

where $r^* = r/\sigma_{ij}$ is the reduced radial distance and $w^*(r^*) = w_{ij}(r^*)/\varepsilon_{ij}$ is the reduced perturbation potential.

In this dimensionless form, the RDFs $g_{0,ij}(r^*)$ show only minor variations across different component pairs, so that $g_{0,x}(r^*) \approx g_{0,ij}(r^*)$ provides a good approximation. This is the central assumption of the one-fluid approximation (OFA),[18,19] in which $g_{0,x}(r^*)$ corresponds to the RDF of an effective pure fluid constructed to reproduce the mixture's packing fraction. Consequently, all pair-specific correlation integrals I_{ij} are replaced by a single correlation integral I_x of the effective pure fluid,

$$I_{ij}(\rho, T, \mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varepsilon}) \approx I_x(\rho, T, \sigma_x, \varepsilon_x), \quad \forall i,j \in \{1,\ldots,N^c\}$$

with interaction parameters σ_x and ε_x defined through mixing rules. For square-well fluids, the correlation integrals are independent of $\boldsymbol{\varepsilon}$ and ε_x; this dependence becomes relevant only for fluids with soft repulsive interactions. Applying the OFA to eq. (5.11) yields

$$\Delta a_1 \approx \Delta a_1^{\mathrm{OFA}} = 2\pi\rho\beta \left(\sum_{i=1}^{N^c}\sum_{j=1}^{N^c} x_i x_j \varepsilon_{ij}\sigma_{ij}^3 \right) I_x \tag{5.13}$$

To correct for the inaccuracies introduced by the OFA in the first order perturbation term, we introduce a correction factor f_x,

$$\Delta a_1^{\mathrm{OFA},\, f_x} = 2\pi\rho\beta \left(\sum_{i=1}^{N^c}\sum_{j=1}^{N^c} x_i x_j \varepsilon_{ij}\sigma_{ij}^3 \right) f_x I_x \tag{5.14}$$

which is rigorously defined as

$$f_x^{\text{true}} = \frac{\Delta a_1}{\Delta a_1^{\text{OFA}}} = \frac{\sum_{i=1}^{N^c}\sum_{j=1}^{N^c} x_i x_j \varepsilon_{ij}\sigma_{ij}^3 I_{ij}}{\left(\sum_{i=1}^{N^c}\sum_{j=1}^{N^c} x_i x_j \varepsilon_{ij}\sigma_{ij}^3\right) I_x} \tag{5.15}$$

We propose a simple analytical function $f(\mathbf{z})$ that approximates f_x^{true}. The input variables $\mathbf{z}$ of this function must be chosen such that they can capture the structural differences between the mixture and the effective pure fluid.

This is most intuitively analysed for square-well fluids, for which the correction factor simplifies as

$$f_x^{\text{true}} = \frac{\sum_{i=1}^{N^c}\sum_{j=1}^{N^c} x_i x_j \varepsilon_{ij}\sigma_{ij}^3 \int_1^{1.5} g_{ij}^{\text{hs}}(r_{ij}^*) r_{ij}^{*2} \mathrm{d}r_{ij}^*}{\left(\sum_{i=1}^{N^c}\sum_{j=1}^{N^c} x_i x_j \varepsilon_{ij}\sigma_{ij}^3\right) \int_1^{1.5} g^{\text{hs}}(r_x^*) r_x^{*2} \mathrm{d}r_x^*} \tag{5.16}$$

where $g_{ij}^{\text{hs}}(r_{ij}^*)$ is the RDF of hard-spheres of species i and j within a hard-sphere mixture and $g^{\text{hs}}(r_x^*)$ is the RDF in a pure hard-sphere fluid at dimensionless distance $r_x^* = r/\sigma_x$. The size parameter σ_x of the effective pure fluid can be defined through different mixing rules; in this work, we apply

$$\sigma_x^3 = \sum_{i=1}^{N^c} x_i \sigma_{ii}^3 \tag{5.17}$$

which is the same mixing rule used in the uv-theory[17] and the PC-SAFT equation of state[15]. For square-well mixtures, a definition of the energy parameter ε_x is not required to evaluate I_x, as the hard-sphere RDF is independent of temperature.

Equation (5.16) suggests that, for square-well fluids, the variables $\mathbf{z}$ are those variables defining the structural properties of hard-sphere fluid mixtures. These variables are well known, being the fundamental measures of the hard-spheres, and density,[31–33]

$$\zeta_n = \frac{\pi}{6}\rho \sum_{i=1}^{N^c} x_i \sigma_{ii}^n \tag{5.18}$$

with $0 \leq n \leq 3$. Boublík[34] developed an algebraic model for the RDF of two hard-spheres of types i and j in a hard-sphere mixture. In this model, the RDF is expressed in terms of the variables ζ_n, i.e., $g_{ij}^{\text{hs}}(r_{ij}^*, \zeta_0, \zeta_1, \zeta_2, \zeta_3)$. For a pure hard-sphere fluid, the model reduces to a function of the packing fraction $\eta = \zeta_3$, i.e., $g^{\text{hs}}(r^*, \eta)$. The full expressions are provided in appendix C.1.2.

Given that the main structural differences between the mixture and the effective pure fluid are captured in their respective RDFs, which are functions of the variables ζ_n in the model of Boublík[34], we expect these variables to be suitable inputs for the algebraic correction factor and approximate

$$f_x(\boldsymbol{\zeta}) \approx f_x^{\text{true}} \tag{5.19}$$

with $\boldsymbol{\zeta} = [\zeta_0, \zeta_1, \zeta_2, \zeta_3]$.

5.4.1 Symbolic Regression

The genetic symbolic regression algorithm implemented in the open-source library PySR[35] is employed to identify algebraic functions $f_x(\boldsymbol{\zeta})$ approximating the exact correction factor f_x^{true} defined in eq. (5.16). Symbolic regression is a machine-learning approach that aims to find human-interpretable expressions, as functions of predefined variables (features), capable of accurately representing a given dataset[35].

A reference dataset for the regression is generated by numerically evaluating the correlation integrals I_{ij} and I_x, using the hard-sphere RDF model proposed by Boublík[34]. In total, 9818 data points are considered, including data for binary square-well mixtures and pure square-well fluids with 17 different size parameter ratios $1 \leq \sigma_{22}/\sigma_{11} \leq 5$ and 5 different energy parameter ratios $1 \leq \varepsilon_{22}/\varepsilon_{11} \leq 2$ at 10 different mole fractions $0 \leq x_1 \leq 1$, and 21 dimensionless mixture densities $0.01 \leq \rho^*_{\text{mix}} \leq 1.01$, with $\rho^*_{\text{mix}} = \rho \sum_{i=1}^{N^c} x_i \sigma_{ii}^3$. Only state points with packing fractions $\eta < 0.54$ are taken into account. Note that the RDF of hard spheres, and thus f_x^{true}, does not depend on temperature.

To ensure that the symbolic regression algorithm identifies dimensionless expressions for $f_x(\boldsymbol{\zeta})$, four dimensionless variables are introduced as features for the regression,

$$\Pi_0 = \zeta_3, \Pi_1 = \frac{\zeta_3\zeta_1}{\zeta_2^2}, \Pi_2 = \frac{\zeta_3^2\zeta_0}{\zeta_2^3}, \text{ and } \Pi_3 = \frac{\zeta_0\zeta_3}{\zeta_1\zeta_2}$$

The variables Π_0, Π_1 and Π_2 are linearly independent, whereas $\Pi_3 = \Pi_2/\Pi_1$ depends on the other variables. We chose to include Π_3 to guide the regression algorithm towards compact functional forms. The variables Π_1, Π_2 and Π_3 equal unity for pure substances and for mixtures composed of particles of identical size. In these cases, the correction factor must also reduce to unity as it should not alter the pure fluid model. The target function for the regression is

defined as

$$h_x(\Pi_0, \Pi_1, \Pi_2, \Pi_3) \approx f_x^{\text{true}} - 1 \tag{5.20}$$

such that the model is explicitly trained to vanish for pure fluids or size-symmetric mixtures. To guide the regression toward solutions that satisfy this condition, the custom operator $u(x) = 1 - x$ is included among the set of unary functions available to the symbolic regression algorithm. The full set of permitted unary operators includes $\{x^2, x^3, x^{-1}, \sqrt{x}, \log(x), u(x)\}$, and the allowed binary operators are $\{+, -, \times, /\}$. The final correction factor is chosen from a set of candidate expressions generated by multiple regression runs, prioritizing simplicity over accuracy. Details of the regression setup and the selection procedure are provided in appendix C.2.The selected expression for the target function is

$$h(\mathbf{\Pi}) = \Pi_0^3(\Pi_2 - 1) \tag{5.21}$$

which is equivalent to the correction factor

$$f_x(\zeta_0, \zeta_2, \zeta_3) = 1 + \left(\frac{\zeta_3^2 \zeta_0}{\zeta_2^3} - 1\right)\zeta_3^3 \tag{5.22}$$

Figure 5.1 shows a parity plot of symbolic regression predictions from eq. (5.22) against reference data for all data points used in the regression. The overall agreement is good, with most deviations occurring at high packing fractions. This trend is further emphasized, in figure 5.2, where the correction factor for square-well mixtures with $\sigma_{22}/\sigma_{11} = 3$ and $\varepsilon_{22}/\varepsilon_{11} = 1$ is shown across varying compositions. For these mixtures, the model predicts too low values for the correction factor at high packing fractions. At low packing fractions, the true correction factor takes values below unity, whereas the model consistently predicts values larger than one. Despite its simplicity, the functional form given in eq. (5.22) captures the trends of f_x^{true} with packing fraction and composition reasonably well.

5.4.2 Application to Lennard-Jones Mixtures and Mixtures of Real Substances

The proposed correction factor towards a one-fluid mixing rule can be expected to be transferable to Lennard-Jones mixtures. First, the square-well fluid ($\lambda^{\text{SW}} = 1.5$) has dimensionless properties that are somewhat comparable to those of a LJ fluid. Secondly, while the soft repulsion of a LJ fluid distinguishes it from a square-well fluid, one often represents softly repulsive fluids as hard-sphere fluids with an effective repulsive diameter. Such a diameter d_{ii} is used within eq. (5.18), instead of σ_{ii}, making the resulting expression for $f_x(\zeta_0, \zeta_2, \zeta_3)$

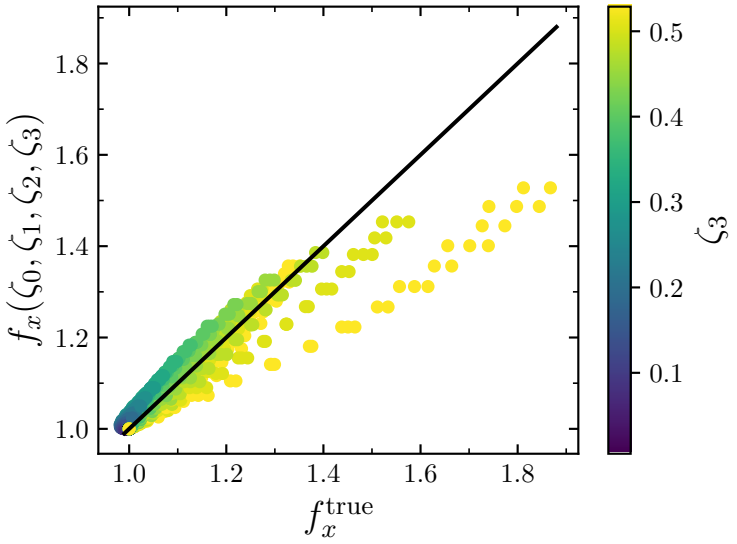


Figure 5.1: Parity plot of the algebraic correction factor $f_x(\zeta)$ from eq. (5.22) versus reference values f_x^{true} obtained by numerical integration of eq. (5.16). All points used in the symbolic regression are included. Colour indicates the packing fraction ζ_3.

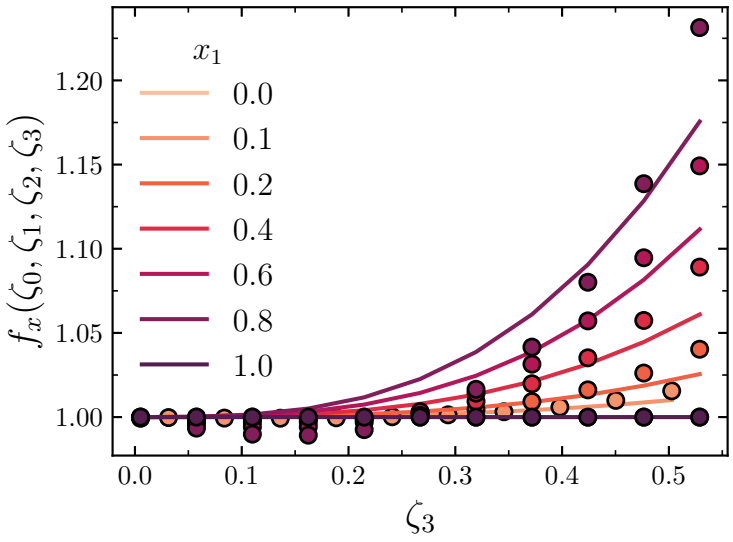


Figure 5.2: Correction factor $f_x(\zeta)$ from eq. (5.22) (lines) compared to reference data f_x^{true} (symbols) for binary square-well mixtures with size ratio $\sigma_{22}/\sigma_{11} = 3$ and energy parameter ratio $\varepsilon_{22}/\varepsilon_{11} = 1$ at various mole fractions x_1.

suitable for mixtures of Lennard-Jones fluids. To assess the transferability of the correction factor f_x of eq. (5.22) to Lennard-Jones mixtures, we apply f_x within two different equations of state: Both variants of the uv-theory[17], denoted as uv-BH-theory and uv-WCA-theory. Further, we test the transferability of the proposed model to mixtures of real substances by

applying f_x within the PC-SAFT equation of state[15].

To apply the correction factor within these EoS, one simply multiplies it with the respective first-order attractive perturbation contribution to the Helmholtz energy. For the *uv*-theory, the correction factor is also multiplied with the first-order perturbation contribution to the second virial coefficient (ΔB_{21}^{u} in the notation of ref. 17). Further details are provided in appendix C.3.

In both EoS, temperature-dependent effective diameters $\mathbf{d}(T)$ are used to map the properties of the BH or WCA reference fluids onto those of a hard-sphere (chain) fluid and the functions ζ_n that serve as input to f_x depend explicitly on temperature through $\mathbf{d}(T)$,

$$\zeta_n = \frac{\pi}{6}\rho \sum_{i=1}^{N^c} x_i m_i d_{ii}^n \tag{5.23}$$

where m_i is the number of segments in a chain molecule of component i. For the model fluids examined in this work, $m_i = 1$.

All evaluations of the PC-SAFT equation of state and the *uv*-theory in this work are performed using the open-source framework for equations of state FeOs[36]

5.5 Molecular Simulations

Molecular simulations of size- and energy-asymmetric binary, ternary and quinary LJ mixtures are performed. Single phase properties among temperature, pressure, density, and energy as well as radial distribution functions $g_{ij}(r)$ are obtained from molecular dynamics (MD) simulations. Phase equilibrium properties are determined using grand canonical Monte Carlo (GCMC) simulations with transition-matrix sampling and histogram reweighting.[37–41] An overview of the simulated systems comprising more than 5000 state points is provided in tables 5.2–5.4, and simulation details are given in appendix C.4.

5.6 Results for Square-Well Mixtures

Figure 5.3 shows the attractive contribution to the compressibility factor $Z^{\mathrm{att}} = Z - Z^{\mathrm{hs}}$ and the residual internal energy U^{res} for various binary square-well mixtures. Molecular simulation data from Akhouri and Solana[42] (symbols) is compared to predictions calculated with the first-order perturbation theories introduced in section 5.4: the rigorous mixture

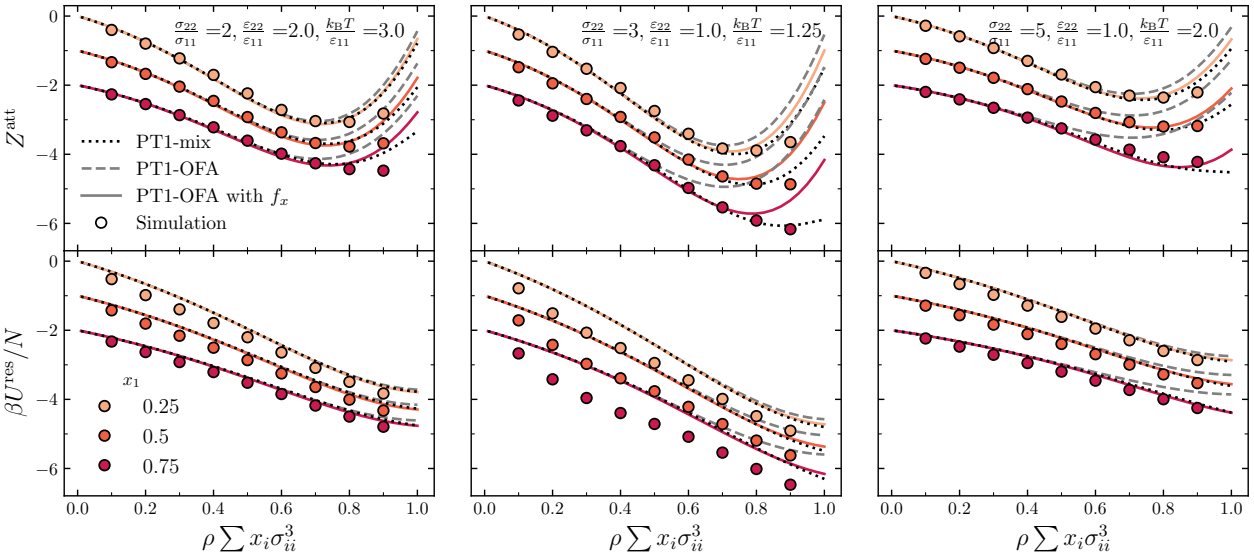


Figure 5.3: Comparison of molecular simulation data from Akhouri and Solana[42] for binary square-well mixtures with results from first order perturbation theories using the rigorous mixture description of eq. (5.11) (dotted), the OFA of eq. (5.13) (dashed), and the OFA with correction factor of eq. (5.14) (solid). For clarity, the lines and symbols representing different mole fractions x_1 are shifted downward by 1 relative to the curve above them. The shorthand notation $\sum$ on the x-axis indicates the sum over all components $\sum_{i=1}^{N^c}$.

formulation Δa_1 of eq. (5.11) (PT1-mix, dotted lines), the one-fluid approximation $\Delta a_1^{\mathrm{OFA}}$ of eq. (5.13) (PT1-OFA, dashed lines), and the mixture-corrected version of the one-fluid approximation $\Delta a_1^{\mathrm{OFA},f_x}$ of eq. (5.14) (PT1-OFA with f_x, solid lines). The simulation results for Z^{att} are obtained by subtracting the hard-sphere mixture contribution, Z^{hs}, from the total compressibility factor Z reported by Akhouri and Solana[42], where Z^{hs} is computed using the Boublík–Mansoori–Carnahan–Starling–Leland equation of state.[32,33]

The rigorous mixture predictions (PT1-mix) show significantly better agreement with the molecular simulation data than the predictions obtained with the one-fluid approximation (PT1-OFA), emphasizing the loss of accuracy introduced with applying the OFA. Deviations between the PT1-mix predictions and the simulation results are expected, since only a first-order perturbation term is considered and no cross-over approach is incorporated to account for density fluctuations near the critical point.[43–46] Higher-order contributions would be required for improved accuracy. The OFA results obtained using the correction factor (PT1-OFA with f_x) show significantly improved agreement with the molecular simulation data compared to the uncorrected OFA results.

Table 5.1: Mean absolute relative deviations (MARD) of equation of state predictions for the attractive compressibility factor Z^{att} and the residual internal energy U^{res} of binary square-well mixtures compared to results from molecular simulation of Akhouri and Solana[42].

EoS	MARD Z^{att}/%	MARD $U^{\text{res}}/(N\varepsilon_{11})$/%
PT1-mix	7.05	18.35
PT1-OFA	10.97	19.24
PT1-OFA with f_x	8.24	17.84
Gross and Sadowski[47]- OFA	8.58	10.21
Gross and Sadowski[47]- OFA with f_x	4.06	8.83

In figure 5.4, molecular simulation data from Akhouri and Solana[42] for an equimolar binary mixture with a size ratio of $\sigma_{22}/\sigma_{11} = 5$ and equal energy parameters are compared with predictions obtained with the second order perturbation theory for square-well fluids of Gross and Sadowski[47]. The dashed lines correspond to predictions calculated with the OFA. The solid lines are calculated with the correction factor of eq. (5.22) multiplied to the first-order perturbation term. Incorporating the correction factor substantially improves the model's accuracy.

Table 5.1 summarizes the mean absolute relative deviations

$$\text{MARD} = \frac{100\%}{N^{\text{points}}} \sum_{i=1}^{N^{\text{points}}} \left| \frac{X_i^{\text{EoS}} - X_i^{\text{sim}}}{X_i^{\text{sim}}} \right| \tag{5.24}$$

with $X \in \{Z^{\text{att}}, U^{\text{res}}\}$ for all EoS considered in this section, evaluated against $N^{\text{points}} = 648$ molecular simulation data points from Akhouri and Solana[42] for size-asymmetric binary square-well mixtures with a potential width of $\lambda^{\text{SW}} = 1.5$ at reduced mixture densities $\rho^*_{\text{mix}} < 1$. The data set includes binary mixtures with size ratios $1.5 \leq \sigma_{22}/\sigma_{11} \leq 5$, energy parameter ratios $1 \leq \varepsilon_{22}/\varepsilon_{11} \leq 2$ and dimensionless temperatures $1 \leq k_{\text{B}}T/\varepsilon_{11} \leq 4$.

Overall, incorporating the correction factor significantly improves the accuracy of the perturbation theories based on the one-fluid approximation, both for the equation of state of Gross and Sadowski[47] and for the numerically evaluated first-order perturbation theory.

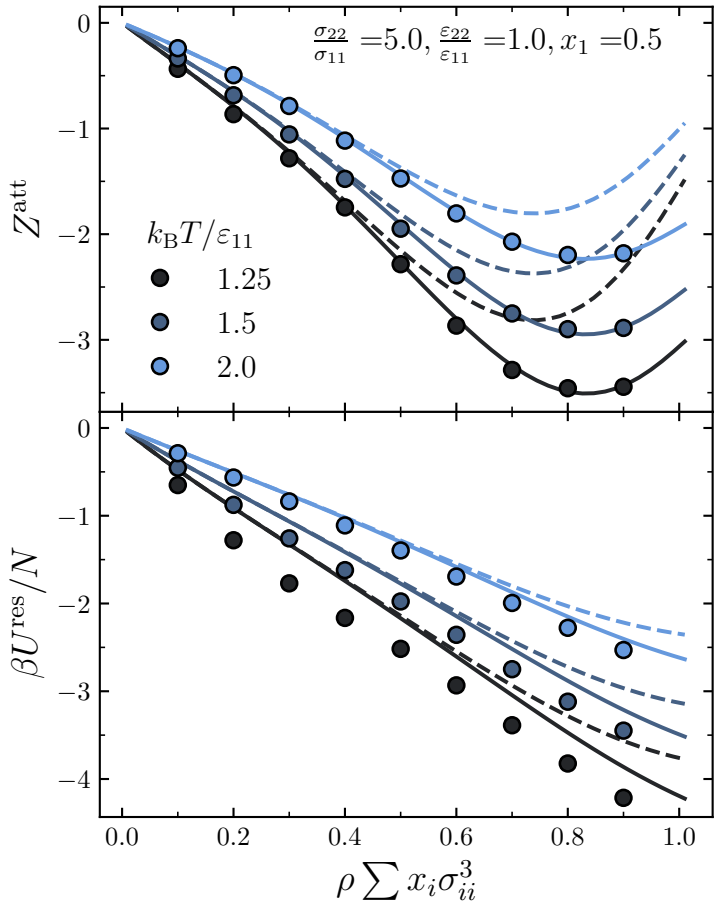


Figure 5.4: Attractive compressibility factor and residual energy of a binary square-well mixture at different temperatures obtained with the second-order perturbation theory of Gross and Sadowski[47] with OFA (dashed) and with OFA and correction factor (solid) compared to molecular simulation data[42] (symbols).

5.7 Results for Lennard-Jones Mixtures

5.7.1 Binary LJ mixtures

In figures 5.5 and 5.6, fluid property predictions calculated with the *uv*-BH-theory are compared to molecular simulation data for binary LJ mixtures. Figure 5.5 shows residual pressures of LJ mixtures with size ratios $\sigma_{22}/\sigma_{11} = 2$ (top) and $\sigma_{22}/\sigma_{11} = 3$ (bottom) at varying compositions. Although developed for square-well fluids, the correction factor improves the accuracy of the *uv*-theory predictions for both LJ mixtures. Figure 5.6 shows isotherms of the compressibility factor and residual internal energy for an equimolar mixture. The marked improvement in the description of the internal energy is especially rewarding,

Table 5.2: Comparison of **binary** LJ mixture property predictions obtained with the uv-BH-theory, the uv-WCA-theory and the PC-SAFT equation of state with and without correction factor, against results from molecular simulation. Column # indicates the number of molecular simulation results included in the comparison, and "Ref." lists the corresponding references, where entries marked with an asterisk (*) denote data from this work. Rows in bold indicate averages over all data points.

					MARD /%					
#	Ref.	$\frac{\sigma_{22}}{\sigma_{11}}$	$\frac{\varepsilon_{22}}{\varepsilon_{11}}$	$\frac{k_B T}{\varepsilon_{11}}$	uv-BH	uv-BH with f_x	uv-WCA	uv-WCA with f_x	PC-SAFT	PC-SAFT with f_x
Residual internal energy $\beta U^{\mathrm{res}}(T,\rho,x_1)/N$										
763	*	1.25	0.5, 1.0	1.0-3.0	1.2	1.2	1.3	1.3	4.4	4.3
1177	*	1.75	0.5, 1.0, 1.5	1.0-3.0	1.9	1.6	1.8	1.4	5.2	4.9
608	*	2	0.5, 1.0, 1.5	1.0-3.0	2.9	2.0	2.5	1.3	6.0	4.8
1244	*	3	0.5, 1.0, 1.5	1.0-3.0	4.4	2.7	3.9	2.6	8.1	5.8
4918					**2.4**	**1.8**	**2.3**	**1.6**	**5.8**	**4.9**
Compressibility factor $Z(T,\rho,x_1)$										
763	*	1.25	0.5, 1.0	1.0-3.0	1.0	0.9	0.7	0.6	2.6	2.5
1177	*	1.75	0.5, 1.0, 1.5	1.0-3.0	2.7	1.8	2.3	1.4	6.0	4.3
608	*	2	0.5, 1.0, 1.5	1.0-3.0	6.3	3.9	5.5	3.8	11.9	7.0
1244	*	3	0.5, 1.0, 1.5	1.0-3.0	8.5	3.4	7.2	7.6	12.9	5.2
4918					**4.1**	**2.2**	**3.5**	**3.1**	**7.6**	**4.4**

as this thermal property depends on temperature derivatives of the free energy, while the correction factor was developed for SW fluids, whose correlation integrals are temperature independent.

Figure 5.7 considers vapour-liquid equilibria (VLE) and compares model predictions obtained using the uv-WCA-theory and the PC-SAFT equation of state to results from molecular simulation. Additional figures showing VLEs of more systems and evaluations using the uv-BH-theory are provided in appendix C.3. Including the correction factor f_x significantly improves the accuracy of the model predictions for all three EoS.

Table 5.2 summarizes mean absolute relative deviations (MARDs) for a broad set of mixtures with varying size and energy parameter ratios. MARDs are shown for the compressibility factor and the residual internal energy for the uv-BH-theory, the uv-WCA-theory, and the PC-SAFT equation of state. In nearly all cases, the correction factor improves agreement with simulation data. The only exception is the compressibility factor prediction of the uv-WCA-theory at a size ratio of $\sigma_{22}/\sigma_{11} = 3$, where the correction leads to somewhat larger deviations at low temperatures ($k_B T/\varepsilon_{11} \lesssim 1.5$). Overall the results indicate good transferability of the correction factor f_x to binary mixtures of Lennard-Jones fluids.

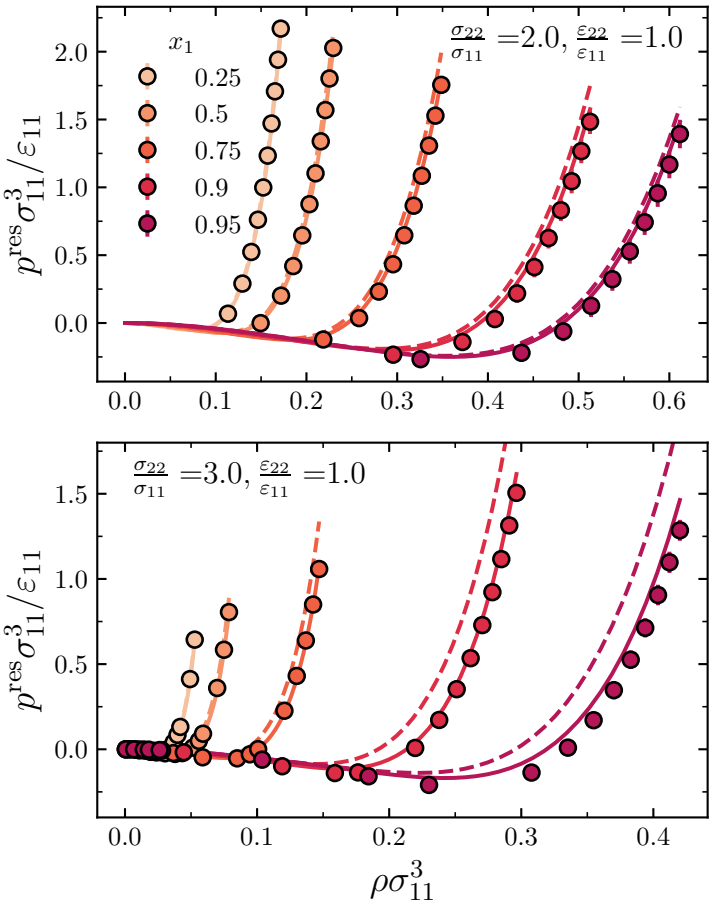


Figure 5.5: Residual pressures of binary LJ mixtures at different compositions at temperature $k_\mathrm{B}T/\varepsilon_{11} = 1.5$. Predictions from the *uv*-BH-theory without (dashed lines) and with the correction factor (solid lines) compared to results from molecular simulation from this work (circles).

5.7.2 Ternary and Quinary LJ Mixtures

In figure 5.8, *uv*-WCA-theory predictions of the residual energy for a ternary mixture (top) and a quinary mixture (bottom) at various compositions are compared with results from molecular simulation. Figure 5.9 shows compressibility factor and residual energy isotherms of an equimolar quinary LJ mixture calculated with the *uv*-BH-theory, compared with molecular simulation data. In both figures, predictions including the correction factor (solid lines) are in better agreement with the molecular simulation data (circles) than predictions calculated with the original theories (dashed lines). This is notable, as the correction factor f_x was determined solely from binary mixture data. Tables 5.3 and 5.4 summarize the MARDs for various ternary and quinary LJ systems. For the PC-SAFT EoS, the correction factor improves the prediction accuracy in all cases. For both versions of the *uv*-theory, introducing the correction factor

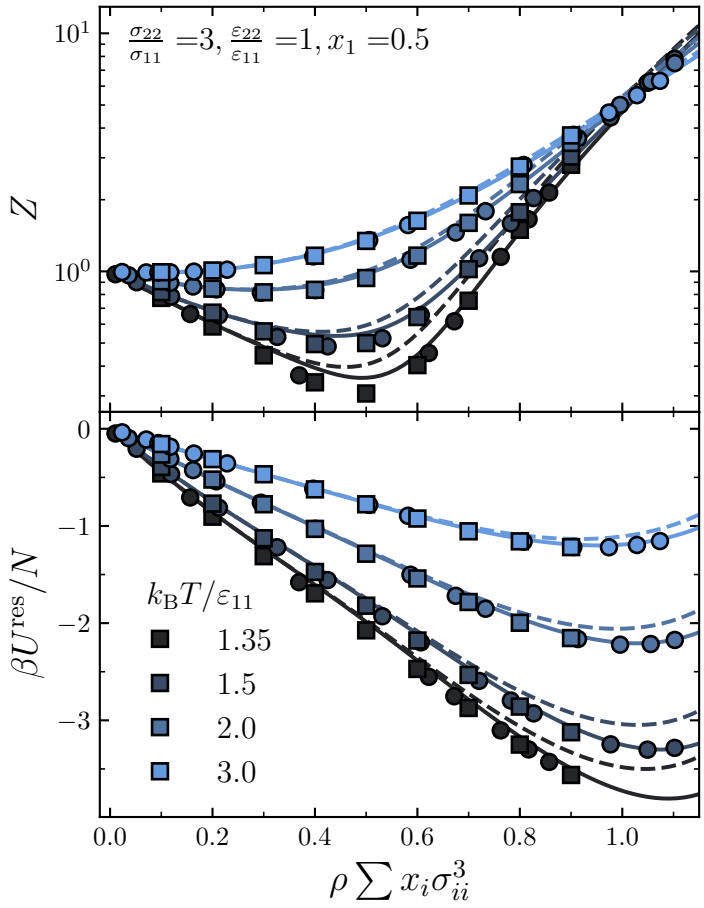
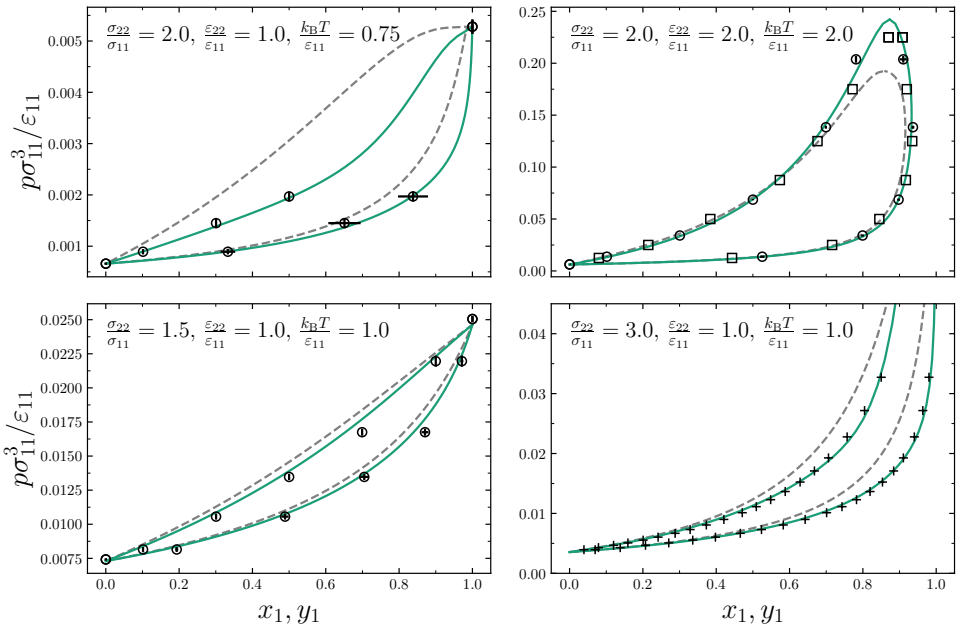


Figure 5.6: Isotherms of the compressibility factor and residual energy for an equimolar binary LJ mixture. Predictions from the *uv*-BH-theory without (dashed lines) and with correction factor (solid lines) compared to molecular simulation data from Akhouri and Solana[48] (squares) and this work (circles).

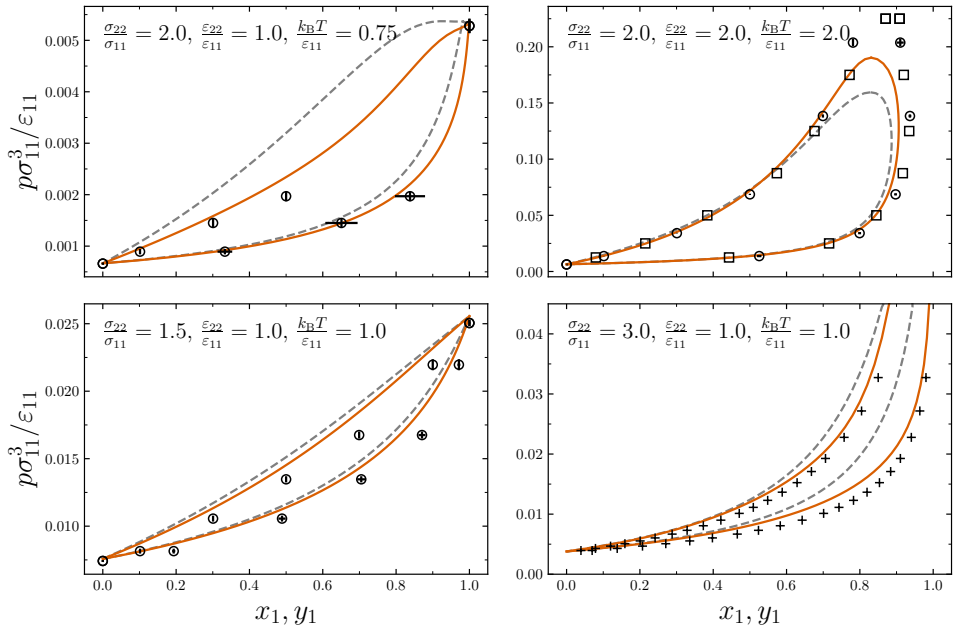
leads to a slight increase in MARD for a few ternary systems. However, the original *uv*-theory already provides very accurate results for these cases, and predictions with the correction factor remain satisfactory. The largest deviations between simulation and model predictions occur for data points close to the metastable phase region.

5.8 Results for Mixtures of Real Substances

The impact of the correction factor on predictions for real substances is assessed by comparing PCP-SAFT[15,22,23] phase equilibrium results with experimental data from the Dortmund Data Bank (DDB)[52]. PCP-SAFT extends PC-SAFT[15] by adding Helmholtz energy contributions accounting for polar interactions[22] and association[23]. Component-specific model parameters

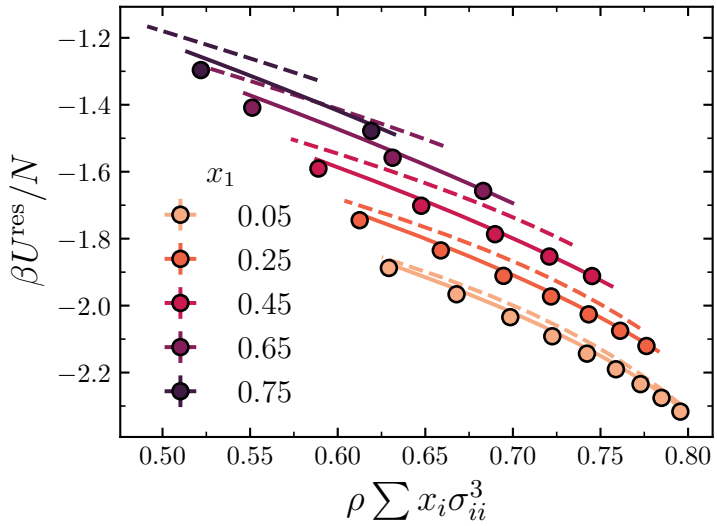


(a) VLEs predicted by the *uv*-WCA-theory.

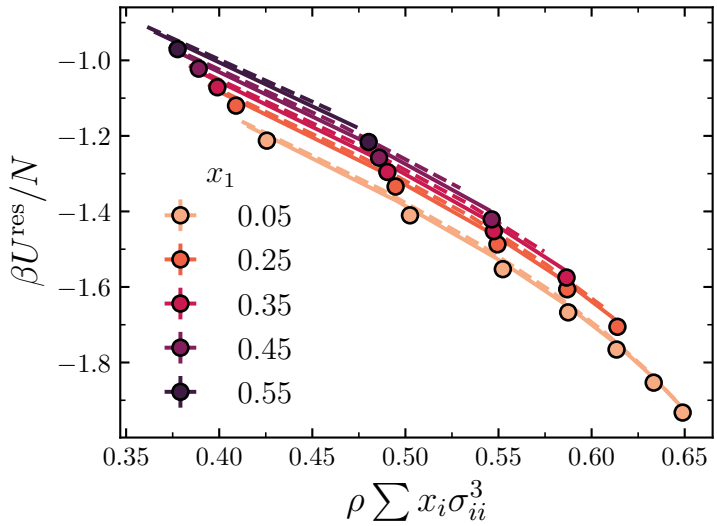


(b) VLEs predicted by the PC-SAFT equation of state.

Figure 5.7: vapour–liquid equilibria of binary Lennard-Jones mixtures. Predictions calculated with the *uv*-WCA-theory (top) and the PC-SAFT equation of state (bottom) without the correction factor (dashed lines) and with the correction factor (solid lines), compared to molecular simulation data from Vrabec *et al.* [49] (circles), Harismiadis *et al.* [50] (squares), and this work (plus symbols).



(a) Ternary mixture with $\sigma_{22}/\sigma_{11} = 2$, $\sigma_{33}/\sigma_{11} = 3$ at pressure $p\sigma_{11}^3/\varepsilon_{11} = 0.2$.



(b) Quinary mixture with $\sigma_{22}/\sigma_{11} = 1.5$, $\sigma_{33}/\sigma_{11} = 2$, $\sigma_{44}/\sigma_{11} = 2.5$, $\sigma_{55}/\sigma_{11} = 3$ at pressure $p\sigma_{11}^3/\varepsilon_{11} = 0.1$.

Figure 5.8: Residual energy of ternary (top) and quinary (bottom) LJ mixtures at $k_B T/\varepsilon_{11} = 2$. Predictions from the *uv*-WCA-theory without (dashed) and with correction factor (solid) are compared with results from molecular simulation from this work (circles). All components have identical energy parameters, $\varepsilon_{ii}/\varepsilon_{11} = 1$ ($i = 2$–5).

are taken from the parameter set of Esper *et al.* [53], which comprises 1842 substances.

In figure 5.10, VLEs of three binary mixtures are presented. The [top] diagram shows the VLE of methylcyclohexane and methane at ambient pressure. Both components are modelled as nonpolar and nonassociating, and the PCP-SAFT predictions are in good agreement with the experimental data. For this mixture, the size parameter ratio is close to unity ($\sigma_{22}/\sigma_{11} = 0.93$),

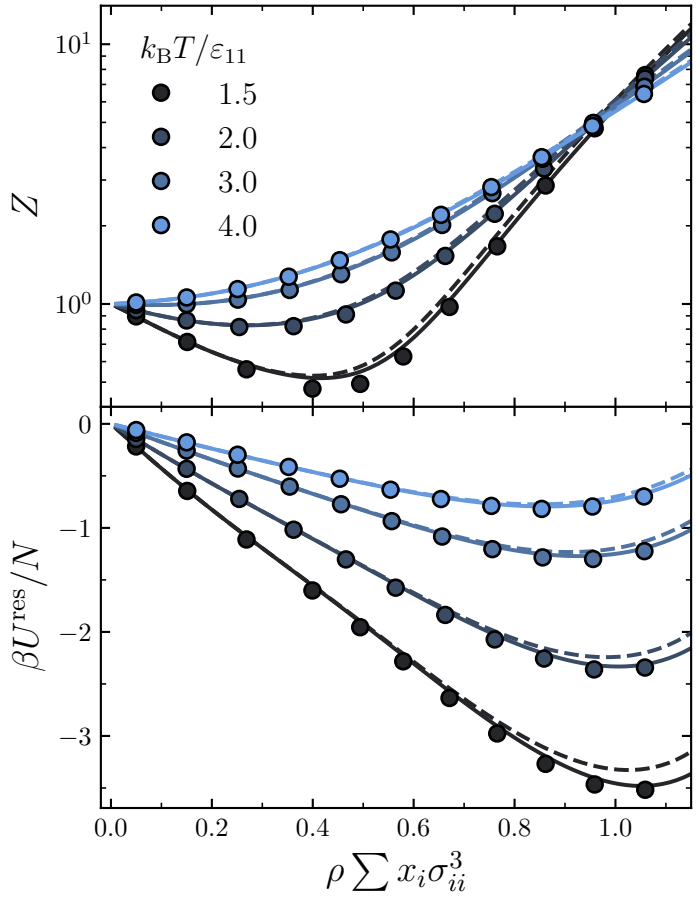


Figure 5.9: Compressibility and residual energy isotherms of an equimolar quinary LJ mixture calculated with the *uv*-BH-theory without (dashed) and with correction factor (solid) compared to results from molecular simulation of this work (circles). Size ratios $\sigma_{22}/\sigma_{11} = 1.5$, $\sigma_{33}/\sigma_{11} = 2$, $\sigma_{44}/\sigma_{11} = 2.5$, $\sigma_{55}/\sigma_{11} = 3$ and equal energy parameters.

and the correction factor is thus also close to unity, leading to results obtained with and without the correction factor that are practically indistinguishable. This mixture is representative of the majority of binary mixtures that can be described with the PCP-SAFT equation of state using the parameter set of Esper *et al.* [53]. Figure 5.11 shows the distribution of size parameters in this set, with most values clustering around the mean $\bar{\sigma} = 3.68$Å. Substantial deviations from this mean typically correspond to substances with pair interactions that are dominated by strong associating or polar effects, making them challenging to model within PCP-SAFT. Examples include water (2.15 Å), nitrogen dioxide (1.9 Å), ammonia (2.37 Å), carbon dioxide (2.58 Å) or benzyl alcohol (4.5 Å). Two representative mixtures with significant size asymmetry are shown in the middle and [bottom] diagrams of figure 5.10. For both, PCP-SAFT predictions deviate noticeably from the experimental data, with the correction factor improving agreement in the middle diagram but slightly worsening it in the

Table 5.3: Same as table 5.2, for **ternary** LJ mixtures. The column "Type" indicates whether equimolar (E) or non-equimolar (NE) state points are considered.

#	Type	Ref.	$\frac{\sigma_{22}}{\sigma_{11}}$	$\frac{\sigma_{33}}{\sigma_{11}}$	$\frac{\varepsilon_{22}}{\varepsilon_{11}}$	$\frac{\varepsilon_{33}}{\varepsilon_{11}}$	$\frac{k_B T}{\varepsilon_{11}}$	MARD /% uv-BH	uv-BH with f_x	uv-WCA	uv-WCA with f_x	PC-SAFT	PC-SAFT with f_x
Residual internal energy $\beta U^{\mathrm{res}}(T,p,x_1,x_2)/N$													
45	NE	*	1.5	2.0	1.0	1.0	2.0	3.1	2.6	1.8	1.2	4.1	3.5
45	NE	*	2.0	3.0	1.0	1.0	2.0	5.6	3.1	4.7	0.96	8.1	3.9
33	E	*	1.5	2.0	1.0	1.0	1.5–3.0	3.6	2.4	3.0	1.3	7.6	5.7
33	E	*	2.0	3.0	1.0	1.0	1.5–3.0	5.7	3.3	4.6	2.2	9.2	5.3
24	NE	51	1.25	1.5	1.5	2.0	2.0–3.0	0.19	0.58	0.46	0.95	1.7	0.92
20	E	51	1.05–1.75	1.125–2.0	1.0	1.0	1.0–3.0	1.6	0.8	1.5	0.32	4.8	3.6
20	E	51	1.05–1.75	1.125–2.0	1.75	2.0	1.0–3.0	0.79	0.47	0.89	1.3	2.4	1.1
Compressibility factor $Z(T,p,x_1,x_2)$													
45	NE	*	1.5	2.0	1.0	1.0	2.0	2.2	1.9	1.6	1.1	3.0	2.5
45	NE	*	2.0	3.0	1.0	1.0	2.0	2.1	0.9	1.7	0.8	1.8	0.99
33	E	*	1.5	2.0	1.0	1.0	1.5–3.0	1.5	1.1	1.2	0.84	2.4	1.9
33	E	*	2.0	3.0	1.0	1.0	1.5–3.0	2.3	1.4	1.9	1.3	2.9	2.0
24	NE	51	1.25	1.5	1.5	2.0	2.0–3.0	0.28	0.47	0.17	0.62	0.81	0.55
20	E	51	1.05–1.75	1.125–2.0	1.0	1.0	1.0–3.0	0.23	0.19	0.26	0.45	1.0	0.85
20	E	51	1.05–1.75	1.125–2.0	1.75	2.0	1.0–3.0	0.29	0.72	0.22	0.89	1.6	1.1

Table 5.4: Same as table 5.3, for **quinary** LJ mixtures with $\varepsilon_{ii}/\varepsilon_{11} = 1$ (i = 2–5) and size ratios $\sigma_{22}/\sigma_{11} = 1.5$, $\sigma_{33}/\sigma_{11} = 2$, $\sigma_{44}/\sigma_{11} = 2.5$, $\sigma_{55}/\sigma_{11} = 3$. All reference data are from molecular simulations performed in this work.

#	Type	$\frac{k_B T}{\varepsilon_{11}}$	$\frac{p\sigma_{11}^3}{\varepsilon_{11}}$	MARD /% uv-BH	uv-BH with f_x	uv-WCA	uv-WCA with f_x	PC-SAFT	PC-SAFT with f_x
Residual internal energy $\beta U^{\mathrm{res}}(T,p,\mathbf{x})/N$									
44	E	1.5-4.0	0.006-2.5	4.6	2.8	3.7	1.5	7.4	4.7
28	NE	2.0	0.1	6.1	4.6	4.7	2.3	7.5	5.1
Compressibility factor $Z(T,p,\mathbf{x})$									
44	E	1.5-4.0	006-2.5	1.7	1.1	1.3	0.88	2.2	1.6
28	NE	2.0	0.1	3.3	2.3	2.6	1.1	3.7	2.2

[bottom] diagram. This suggests that accurately capturing size asymmetry alone is insufficient to ensure predictive accuracy for these mixtures. The observed deviations are more likely due to unlike interactions that are not well captured by the combining rules applied in PCP-SAFT. A well-established approach to improve PCP-SAFT mixture predictions is the introduction of binary interaction parameters, k_{ij}, which modify the unlike energy parameters[54] as

$$\varepsilon_{ij} = (1 - k_{ij})\sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{5.25}$$

Binary interaction parameters, k_{ij}, are optimized for a representative system (benzene and 2-propanol) for both the original and the mixture-corrected PCP-SAFT EoS. The optimization

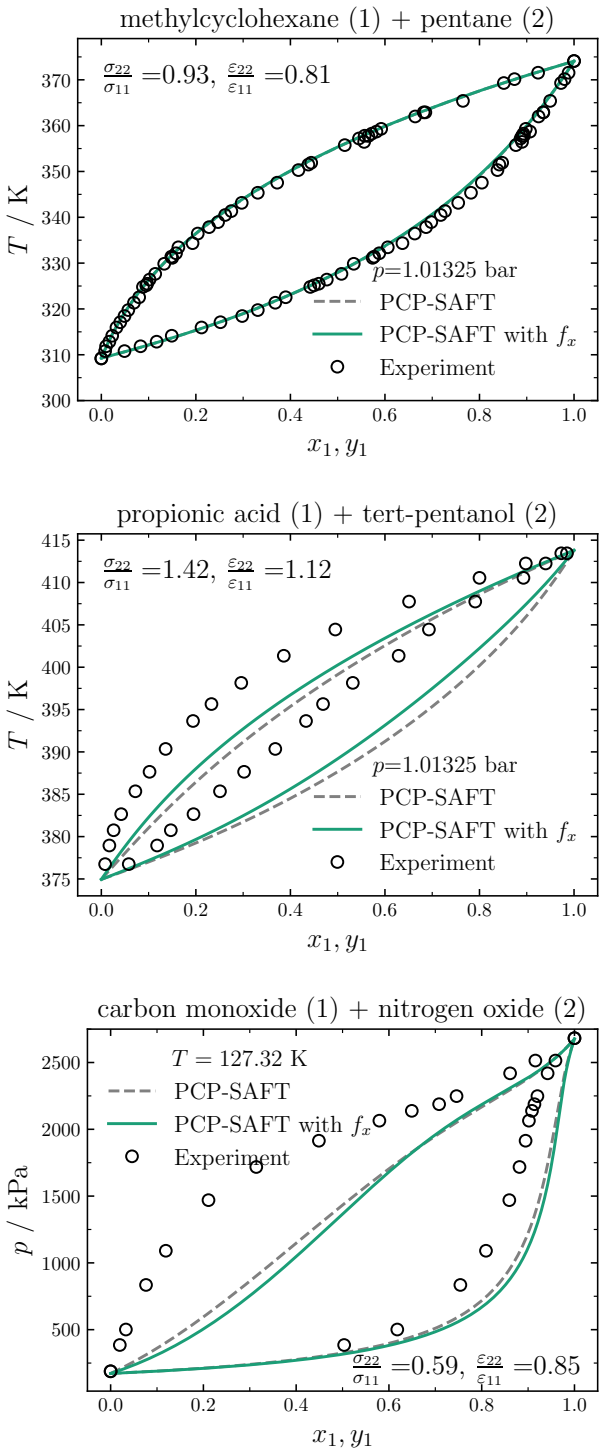


Figure 5.10: vapour–liquid equilibria of binary mixtures. Predictions calculated with the PCP-SAFT equation of state without (dashed lines) and with correction factor (solid lines) compared to experimental data from the DDB[52].

is performed using the FeOs estimator[36], to minimize the lengths of vectors orthogonal to the phase envelope that connect to experimental data. Figure 5.12 shows the optimized k_{ij}

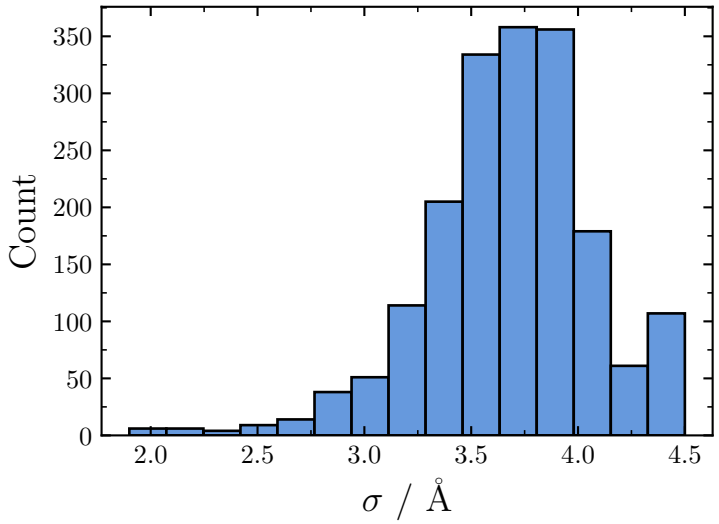


Figure 5.11: Distribution of size parameters σ within the PCP-SAFT parameter set of Esper *et al.* [53] comprising 1842 substances.

values and the corresponding VLEs. Although the optimized k_{ij} values differ slightly between the two equation of state variants, the resulting VLEs are essentially indistinguishable and in good agreement with the experimental data.

5.9 Conclusion and Outlook

In this work, we propose a method to develop correction factors for perturbation theories based on one-fluid approximations, aimed to improve the description of size-asymmetric mixtures. The method is applied to square-well fluids, using symbolic regression to obtain a simple algebraic expression, which is shown to significantly improve equation of state predictions for a broad range of binary mixtures. The same expression is applied within the *uv*-theory and the PC-SAFT equation of state, demonstrating its transferability to Lennard–Jones mixtures as it leads to improved predictions for both, homogeneous fluid properties and phase equilibria. To provide a comprehensive assessment, we perform molecular simulations of various binary, ternary, and quinary Lennard–Jones mixtures. The resulting data [...] can serve to benchmark for new and existing LJ mixture models in the future. The impact of the proposed correction factor for describing real mixtures with the PCP-SAFT equation of state and the parameter set of Esper *et al.* [53] is investigated. The key findings are: (1) for most mixtures, the impact of the correction factor is minor, as it primarily affects strongly size-asymmetric systems, while most substances are modelled with similar segment size

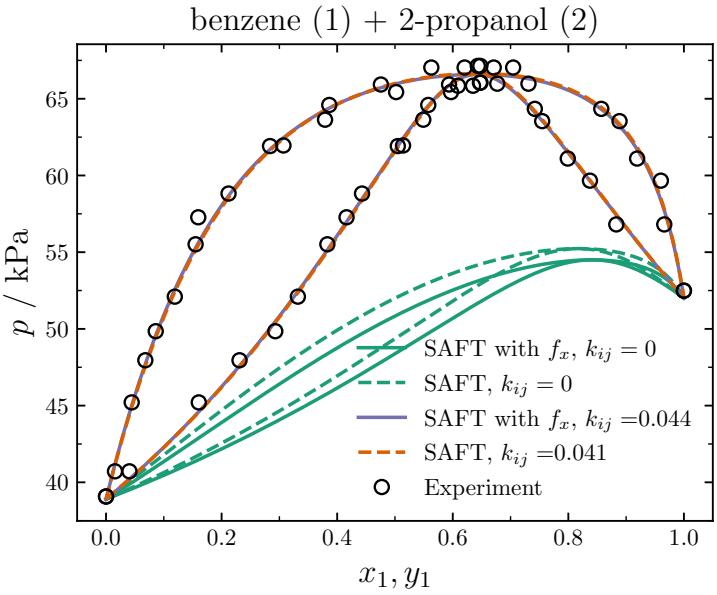


Figure 5.12: VLE of benzene (1) and 2-propanol (2) at temperature $T = 333.15$K. Predictions calculated with the PCP-SAFT equation of state without (dashed) and with correction factor (solid), using no ($k_{ij} = 0$) and optimized ($k_{ij} \neq 0$) binary interaction parameters, compared to experimental data[52]. Pure component parameters are taken from Esper *et al.*[53] ($\sigma_{22}/\sigma_{11} = 0.81$, $\varepsilon_{22}/\varepsilon_{11} = 0.70$), with 2-propanol modelled associating.

parameters, and (2) the main challenge in describing real mixtures accurately is not the accurate treatment of size asymmetry, but rather the presence of unlike interactions that deviate from the standard combining rules. Consequently, the correction factor cannot replace the need of binary interaction parameters for achieving accurate descriptions. Nevertheless, by systematically accounting for size asymmetry, the present approach provides a foundation for allowing a wider range of segment sizes in PCP-SAFT parametrizations and for more rigorously distinguishing the contributions of size asymmetry from deviations due to unlike interactions.

The correction factor presented in this work, eq. (5.22), was developed specifically for the mixing rule $\sigma_x^3 = \sum_{i=1}^{N^c} x_i \sigma_{ii}^3$. The same methodology could be applied in future studies to derive corrections for other mixing rules, such as the commonly used van der Waals mixing rule, $\sigma_x^3 = \sum_{j=1}^{N^c} \sum_{i=1}^{N^c} x_i x_j \sigma_{ij}^3$. In the present work, only square-well potentials with a potential width of $\lambda^{SW} = 1.5$ were considered. The impact of the well depth on the resulting correction factor remains to be examined in future studies.

References

[1] S. Stephan, J. Staubach, and H. Hasse. Review and comparison of equations of state for the Lennard-Jones fluid. *Fluid Phase Equilib.*, **523**:112772, 2020.

[2] W. Wagner and A. Pruß. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *J. Phys. Chem. Ref. Data*, **31**(2):387–535, 2002.

[3] R. Span, E. Lemmon, R. Jacobsen, and W. Wagner. A reference quality equation of state for nitrogen. *Int. J. Thermophys.*, **19**:1121–1132, 1998.

[4] R. Span, E. W. Lemmon, R. T. Jacobsen, W. Wagner, and A. Yokozeki. A reference equation of state for the thermodynamic properties of nitrogen for temperatures from 63.151 to 1000 K and pressures to 2200 MPa. *J. Phys. Chem. Ref. Data*, **29**(6):1361–1433, 2000.

[5] R. Span and W. Wagner. A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa. *J. Phys. Chem. Ref. Data*, **25**(6):1509–1596, 1996.

[6] O. Kunz and W. Wagner. The GERG-2008 wide-range equation of state for natural gases and other mixtures: An expansion of GERG-2004. *J. Chem. Eng. Data*, **57**(11):3032–3091, 2012.

[7] A. Reimer, T. van Westen, and J. Gross. Physically based equation of state for Mie ν-6 fluids. *J. Chem. Phys.*, **158**(16), 2023.

[8] C. G. Gray and K. E. Gubbins. *Theory of Molecular Fluids*. Oxford University Press, 1984. ISBN 9780198556022, 9780191919251.

[9] J.-P. Hansen and I. R. McDonald. *Theory of Simple Liquids: With Applications to Soft Matter*. Academic Press, Oxford, 4th edition, 2013. ISBN 9780123870322.

[10] S. Zhou and J. R. Solana. Progress in the Perturbation Approach in Fluid and Fluid-Related Theories. *Chem. Rev.*, **109**(6):2829–2858, 2009.

[11] K. E. Gubbins. The theory of non-electrolyte solutions: An historical review. *Mol. Phys.*, **111**(24):3666–3697, 2013.

[12] K. E. Gubbins. Perturbation theories of the thermodynamics of polar and associating liquids: A historical perspective. *Fluid Phase Equilib.*, **416**:3–17, 2016.

[13] W. G. Chapman, K. E. Gubbins, G. Jackson, and M. Radosz. New reference equation of state for associating liquids. *Ind. Eng. Chem. Res.*, **29**(8):1709–1721, 1990.

[14] T. Lafitte, A. Apostolakou, C. Avendaño, A. Galindo, C. Adjiman, E. Müller, and G. Jackson. Accurate perturbation theory for chains of Mie soft-core segments (SAFT-VR Mie) for the description

of vapour-liquid equilibria and derivative properties. *J. Chem. Phys.*, **139**:154504, 2013.

[15] J. Gross and G. Sadowski. Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules. *Ind. Eng. Chem. Res.*, **40**(4):1244–1260, 2001.

[16] T. van Westen and J. Gross. Accurate first-order perturbation theory for fluids: uf-theory. *J. Chem. Phys.*, **154**(4), 2021.

[17] T. van Westen and J. Gross. Accurate thermodynamics of simple fluids and chain fluids based on first-order perturbation theory and second virial coefficients: uv-theory. *J. Chem. Phys.*, **155**(24), 2021.

[18] J. S. Rowlinson and F. Swinton. *Liquids and liquid mixtures: Butterworths monographs in chemistry*. Butterworth-Heinemann, 2013.

[19] L. L. Lee. *Molecular thermodynamics of nonideal fluids*. Butterworth-Heinemann, 2016.

[20] A. R. Goodwin and S. I. Sandler. *Mixing and Combining Rules*. Wiley-Blackwell, 2010.

[21] G. Galliéro, C. Boned, A. Baylaucq, and F. Montel. The van der Waals one-fluid model for viscosity in Lennard–Jones fluids: Influence of size and energy parameters. *Fluid Phase Equilib.*, **245**(1):20–25, 2006.

[22] J. Gross and J. Vrabec. An equation-of-state contribution for polar components: dipolar molecules. *AIChE J.*, **52**(3):1194–1204, 2006.

[23] J. Gross and G. Sadowski. Application of the perturbed-chain SAFT equation of state to associating systems. *Ind. Eng. Chem. Res.*, **41**(22):5510–5515, 2002.

[24] H. A. Lorentz. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Annalen der Physik*, **248**(1):127–136, 1881.

[25] D. Berthelot. Sur le mélange des gaz. *Compt. Rendus*, **126**(3):15, 1898.

[26] D. Chandler and J. D. Weeks. Equilibrium structure of simple liquids. *Phys. Rev. Lett.*, **25**(3):149, 1970.

[27] H. C. Andersen, J. D. Weeks, and D. Chandler. Relationship between the hard-sphere fluid and fluids with realistic repulsive forces. *Phys. Rev. A*, **4**(4):1597, 1971.

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

[29] J. A. Barker and D. Henderson. Perturbation theory and equation of state for fluids. II. A successful theory of liquids. *J. Chem. Phys.*, **47**(11):4714–4721, 1967.

[30] C. G. Gray, K. E. Gubbins, and C. G. Joslin. *Theory of Molecular Fluids: Volume 2: Applications*,

volume 10. Oxford University Press, 2011.

[31] R. Roth. Fundamental measure theory for hard-sphere mixtures: a review. *J. Phys.: Condens. Matter*, **22**(6):063102, 2010.

[32] T. Boublík. Hard-sphere equation of state. *J. Chem. Phys.*, **53**(1):471–472, 1970.

[33] G. Mansoori, N. F. Carnahan, K. Starling, and T. Leland Jr. Equilibrium thermodynamic properties of the mixture of hard spheres. *J. Chem. Phys.*, **54**(4):1523–1525, 1971.

[34] T. Boublík. Background correlation functions in the hard sphere systems. *Mol. Phys.*, **59**(4):775–793, 1986.

[35] M. Cranmer. Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. 2023.

[36] P. Rehner, G. Bauer, and J. Gross. Feos: An open-source framework for equations of state and classical density functional theory. *Ind. Eng. Chem. Res.*, **62**(12):5347–5357, 2023.

[37] A. Hemmen, A. Z. Panagiotopoulos, and J. Gross. Grand Canonical Monte Carlo Simulations Guided by an Analytic Equation of State—Transferable Anisotropic Mie Potentials for Ethers. *J. Phys. Chem. B*, **119**(23):7087–7099, 2015.

[38] V. K. Shen and J. R. Errington. Determination of fluid-phase behavior using transition-matrix Monte Carlo: Binary Lennard-Jones mixtures. *J. Chem. Phys.*, **122**(6):064508, 2005. doi:10.1063/1.1844372.

[39] J. R. Errington and V. K. Shen. Direct evaluation of multicomponent phase equilibria using flat-histogram methods. *J. Chem. Phys.*, **123**(16):164103, 2005. doi:10.1063/1.2064628.

[40] A. M. Ferrenberg and R. H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, **61**(23):2635–2638, 1988. doi:10.1103/physrevlett.61.2635.

[41] A. Z. Panagiotopoulos, V. Wong, and M. A. Floriano. Phase Equilibria of Lattice Polymers from Histogram Reweighting Monte Carlo Simulations. *Macromolecules*, **31**(3):912–918, 1998. doi:10.1021/ma971108a.

[42] B. Akhouri and J. Solana. Square-well mixtures revisited: computer simulation, mixing rules and one-fluid theory. *Mol. Simul.*, **46**(2):102–110, 2020.

[43] P. Albright, J. Sengers, J. Nicoll, and M. Ley-Koo. A crossover description for the thermodynamic properties of fluids in the critical region. *Int. J. Thermophys.*, **7**(1):75–85, 1986.

[44] A. Van Pelt, G. Jin, and J. Sengers. Critical scaling laws and a classical equation of state. *Int. J. Thermophys.*, **15**(4):687–697, 1994.

[45] S. Kiselev. Cubic crossover equation of state. *Fluid Phase Equilib.*, **147**(1-2):7–23, 1998.

[46] C. McCabe and S. B. Kiselev. Application of crossover theory to the SAFT-VR equation of state: SAFT-VRX for pure fluids. *Ind. Eng. Chem. Res.*, **43**(11):2839–2851, 2004.

[47] J. Gross and G. Sadowski. Application of perturbation theory to a hard-chain reference fluid: an equation of state for square-well chains. *Fluid Phase Equilib.*, **168**(2):183–199, 2000.

[48] B. Akhouri and J. Solana. Equilibrium thermodynamic properties of Lennard–Jones fluid mixtures from a single-component effective fluid model. *Mol. Simul.*, **49**(11):1117–1124, 2023.

[49] J. Vrabec, A. Lotfi, and J. Fischer. Vapour liquid equilibria of Lennard-Jones model mixtures from the NpT plus test particle method. *Fluid Phase Equilib.*, **112**(2):173–197, 1995.

[50] V. Harismiadis, N. Koutras, D. Tassios, and A. Z. Panagiotopoulos. How good is conformal solutions theory for phase equilibrium predictions?: Gibbs ensemble simulations of binary Lennard-Jones mixtures. *Fluid Phase Equilib.*, **65**:1–18, 1991.

[51] K. Fotouh and K. Shukla. Excess properties of ternary fluid mixtures from simulation, perturbation theory and van der Waals one-fluid theory: size and energy parameter effects. *Chem. Eng. Sci.*, **52**(14):2369–2382, 1997.

[52] Dortmund Data Bank. Dortmund Data Bank. `https://www.ddbst.com`, 2022.

[53] T. Esper, G. Bauer, P. Rehner, and J. Gross. PCP-SAFT parameters of pure substances using large experimental databases. *Ind. Eng. Chem. Res.*, **62**(37):15300–15310, 2023.

[54] P. Rehner, A. Bardow, and J. Gross. Modeling mixtures with PCP-SAFT: Insights from large-scale parametrization and group-contribution method for binary interaction parameters. *Int. J. Thermophys.*, **44**(12):179, 2023.

6 Primitive and Non-Primitive Model Electrolytes: Comparing Ion-Related Helmholtz Energies Using Molecular Simulations

The content of this chapter is a literal quote of the publication:

A. Reimer, I. Reisch and J. Gross. Primitive and non-primitive model electrolytes: Comparing ion-related Helmholtz energies using molecular simulations. *J. Chem. Phys.*, **162**, 124503 (2025), doi: 10.1063/5.0257401

6.1 Abstract

Two main frameworks are commonly used to describe electrolyte solutions: the non-primitive model, which rigorously accounts for all interactions between ions and solvent molecules, and the primitive model, which treats the solvent as a dielectric continuum, only describing ion-ion interactions explicitly. The primitive model offers simple Helmholtz energy expressions, including the Debye-Hückel (DH) equation, the primitive mean spherical approximation (MSA) and the Born theory of solvation. In this work, we evaluate the accuracy of primitive model approaches by comparing their Helmholtz energies with data from molecular simulations obtained for non-primitive model electrolyte solutions. We model electrolyte solutions as mixtures of equally sized, charged, and (non-polarizable) dipolar Lennard-Jones particles. Using thermodynamic integration, we isolate the Helmholtz energy contributions related to solvent-solvent, ion-solvent, and ion-ion interactions. Molecular simulations are performed across two temperatures and two densities, a range of charges, dipole moments, and ion mole fractions ($0.005 \leq x^{\text{ions}} \leq 0.05$). Our results show that while the primitive model expressions provide a qualitatively reasonable description of electrolyte solutions, they systematically underestimate the Helmholtz energy contributions associated with ion-solvent and ion-ion interactions. Achieving quantitative agreement requires empirical adjustments to the Born radius. Notably, the optimized Born radii are significantly larger than the actual ion sizes used in the molecular simulations, questioning the primitive model's applicability. This work presents rigorous benchmarks for the use of MSA, DH, and Born theories, along with molecular simulation data for non-primitive model electrolytes. These benchmarks provide insights for refining existing models and advancing the development of new equations of state for

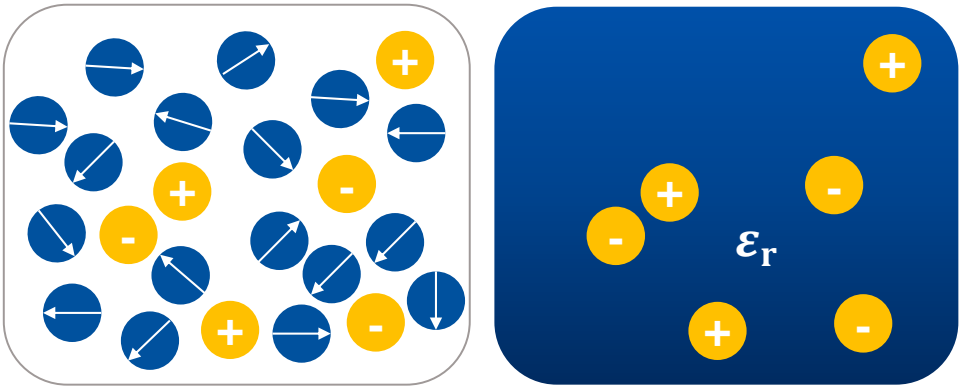


Figure 6.1: Fluid representation in the non-primitive model (left) and in the primitive model (right) for electrolyte systems. Yellow circles indicate cations (plus symbols) and anions (minus symbols). Blue circles (left) represent solvent molecules with the arrows indicating dipole moments. The blue shaded area (right) represents the solvent as dielectric continuum with relative permittivity ε_r.

electrolyte solutions.

6.2 Introduction

Understanding the thermodynamic behaviour of electrolytes is important for a wide range of scientific and industrial applications, including water treatment, battery development, electrolysis, and modelling biochemical reactions. Equations of state (EoS) formulated in terms of the Helmholtz energy or the Gibbs energy summarize all thermodynamic information of a system in a single analytic model. The use of EoS is already common practice for a wide range of applications, such as the design and optimization of processes and products in the chemical, pharmaceutical, and energy industries. However, many of the commonly used EoS are not suitable for describing systems with ions. The development of equations of state capable of describing ions (eEoS) is challenging because of the long-range nature of the electrostatic interactions and is an active area of research. For a review see ref. 1.

Two different frameworks can be distinguished for the theoretical description of electrolyte solutions: the non-primitive (solvent-explicit) model and the primitive (solvent-implicit) model. The representation of electrolyte solutions within both frameworks is schematically illustrated in figure 6.1.

In the non-primitive model (Figure 6.1, left), the interactions between ions and dipolar solvent molecules are treated explicitly. These intermolecular interactions determine the

spatial and orientational arrangement of molecules (i.e. molecular structure) and thus the thermodynamic properties of the electrolyte solution. In the non-primitive model, the total electrostatic contribution to the Helmholtz energy is defined by the sum of contributions related to ion-ion, ion-dipole and dipole-dipole interactions. The mathematical description of these contributions, however, is demanding. Examples of eEoS based on the non-primitive model are the perturbation theory of Theiss and Gross[2] and the non-primitive mean spherical approximation[3–5].

Within the primitive model framework (Figure 6.1, right), the solvent is regarded as a dielectric continuum with a relative permittivity ε_r. The relative permittivity can be related to the ability of a solvent to shield ions by aligning its polar molecules around them.[6] Only ion-ion interactions are considered explicitly. An approximate analytical expression for describing the effect of these interactions on the Helmholtz energy was originally derived by Debye and Hückel (DH)[7]; another prominent analytical expression was derived from integral equation theory, using the mean spherical approximation (MSA) closure.[8,9] Based on Born's theory of solvation[10], another analytic expression was derived, which is generally considered an approximate model for the effect of ion-solvent interactions on the Helmholtz energy.

Various eEoS[11–31] combine the primitive and the non-primitive frameworks by explicitly modelling short-ranged (e.g., dispersive and solvent-solvent) interactions while using either the DH equation or the MSA result to account for ion-ion interactions. Some of these eEoS include the Born theory to account for ion-solvent (ion-dipole) interactions. This approach can achieve good results, but is physically questionable for several reasons. First, the primitive model expressions are derived under the assumption of low ion concentrations, which makes their applicability to electrolyte solutions with elevated ion concentrations insufficient. Second, the use of the Born theory in eEoS is controversial.[32,33] Moreover, the primitive model expressions require input parameters such as the relative permittivity ε_r and ion size parameters, both of which can be defined in multiple ways[34–38]

Despite these concerns, eEoS using primitive modelling approaches are more popular than their non-primitive counterparts. This is primarily due to their mathematical simplicity, their ease of parametrization and their robustness for extrapolations (noting that extrapolative capabilities are rather limited for most eEoS). Although the non-primitive model provides a physically rigorous and well-defined framework for calculating the electrostatic Helmholtz energy, its practical application is hindered by the complexity of deriving analytical formulations. Since both models aim to describe the same underlying system, a connection or transferability between them should exist. This study aims to identify this connection. Understanding the relationship between the Helmholtz energy contributions in the primitive and non-primitive modelling frameworks could address the limitations of eEoS based on the primitive model,

clarify the definition of their inputs, and guide the development of new physically rigorous eEoS. Insights from both approaches could be integrated, such as developing an eEoS based on the non-primitive model that incorporates the proper limiting behaviour described by DH theory.

Systematic comparisons of the primitive and non-primitive models remain scarce in literature. Notable exceptions include the studies by Kournopoulos *et al.* [39] and Simonin [33]. Kournopoulos *et al.* [39] compare the chemical potential derived from primitive model expressions with that from non-primitive molecular simulations for aqueous NaCl. They isolate the ion-ion contribution by simulating the system with non-polar water molecules. The resulting chemical potential is compared to that obtained using the MSA or the DH equation. Additionally, they compare the sum of the chemical potential from MSA and Born to that obtained from molecular simulations using a polar water force-field. The authors observe good qualitative agreement between the results from the molecular simulations and the primitive model expressions. However, they note that achieving quantitative agreement requires empirical adjustment of the Born size parameter. Simonin [33] compares the chemical potential contribution calculated using the Born theory with that obtained using the non-primitive MSA solution for ion-dipole interactions. The author finds that the Born theory is not well-suited for describing the contribution of ion-solvent interaction to the chemical potential, as it significantly overestimates their contribution in most systems.
In this study, we perform a thorough comparison between the Helmholtz energy contributions associated with ion-ion and ion-dipole interactions obtained from molecular simulations of non-primitive electrolytes and those calculated using classical primitive model expressions, specifically the DH equation, the primitive MSA, and the Born theory of solvation. To facilitate this comparison, we consider simple model electrolyte systems composed of equally sized charged and dipolar Lennard-Jones particles. Molecular simulations for non-primitive model electrolytes are performed to calculate the Helmholtz energy contributions due to ion-dipole, dipole-dipole, and ion-ion interactions, which requires a decomposition of the Ewald summation into various contributions. We investigate a broad range of electrolyte systems with varying ion concentrations, ion strengths, solvent dipole moments, as well as different temperatures and densities. Additionally, we explore the impact of different definitions of the relative permittivity and the ion size parameter, both of which are required as inputs for the primitive model expressions. This study provides rigorous benchmarks for the MSA, DH, and Born equations, which will be of value for improving existing models of electrolyte solutions, particularly in the context of eEoS. Specifically, we present new molecular simulation data for the Helmholtz energy and potential energy contributions of model electrolytes offering valuable insights for the development of eEoS in both the primitive and non-primitive modelling frameworks.

6.3 Molecular Model

We model electrolyte solutions as a mixture containing a fixed number of positively charged Lennard-Jones (LJ) particles, negatively charged LJ particles and dipolar LJ particles in a system with constant volume V and temperature T.

LJ anions (-) and LJ cations (+) are modelled with a point charge of magnitude $q = q_+ = -q_-$ placed in their LJ centre. The ion mole fraction is defined as $x^{\text{ions}} = N^{\text{ions}}/N$, where $N^{\text{ions}} = N_+ + N_-$ is the total number of ions and N is the total number of particles in the system. Due to electro-neutrality the number of anions equals the number of cations $N_+ = N_-$.

The dipolar, non-polarizable solvent particles, also referred to as Stockmayer fluid, are modelled based on the ion-extended-dipole model. In this model, each dipole moment consists of two point charges of opposite sign and same magnitude q_{d}, which are placed symmetrically at a distance of $l = 0.1\sigma$ from the centre of mass of the particle. The dipole moment follows from

$$\mu = q_{\text{d}} l \tag{6.1}$$

Drunsel *et al.* [40] showed that the ion-extended dipole model (with $l = 0.1\sigma$) is well suited to describe point dipoles.

The short-range (non-electrostatic) pair-potential between all particles is described by the LJ potential

$$u^{\text{LJ}}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{6.2}$$

where r is the distance between the two particles, σ is the size parameter and ϵ is the energy parameter. All particles are characterized by the same size parameter σ and the same energy parameter ϵ.

The electrostatic interaction between any two charges q_i and q_j on different particles is described by Coulomb's law

$$u^{\text{qq}}(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r} \tag{6.3}$$

where ε_0 is the vacuum permittivity. If ion-ion interactions are regarded in eq. (6.3), i.e. if q_i and q_j are charges of two ions, then we denote the resulting pair potential as u^{cc}. For pair potentials between (four charges) of two dipoles we refer to the resulting pair potential as

u^{dd}. And for ion-dipole pairs, as u^{cd}. The full potential energy of the model electrolyte system can be written as

$$U(\mathbf{r}^N, \boldsymbol{\omega}^N) = U^{\mathrm{LJ}}(\mathbf{r}^N) + U^{\mathrm{ES}}(\mathbf{r}^N, \boldsymbol{\omega}^N) \tag{6.4}$$

where $\mathbf{r}^N$ denotes the positions of all particles, $\boldsymbol{\omega}^N$ denotes the orientations of all dipoles, U^{LJ} is the sum of all LJ pair interactions and

$$U^{\mathrm{ES}}(\mathbf{r}^N, \boldsymbol{\omega}^N) = U^{\mathrm{dd}}(\mathbf{r}^N, \boldsymbol{\omega}^N) + U^{\mathrm{cd}}(\mathbf{r}^N, \boldsymbol{\omega}^N) + U^{\mathrm{cc}}(\mathbf{r}^N) \tag{6.5}$$

is the total electrostatic potential energy with $U^{\mathrm{dd}}, U^{\mathrm{cd}}$ and U^{cc} denoting the sums over all dipole-dipole, ion-dipole and ion-ion interactions. For brevity, we henceforth omit the dependency on $\mathbf{r}^N$ and $\boldsymbol{\omega}^N$ in these equations.

In the following sections, thermodynamic variables and properties are made dimensionless. Dimensionless Helmholtz energy contributions are indicated by a lowercase letter $a = \beta A/N$ with $\beta = (k_{\mathrm{B}}T)^{-1}$ and Boltzmann's constant k_{B}. Other dimensionless quantities are denoted by an asterisk: e.g. temperature $T^* = k_{\mathrm{B}}T/\epsilon$, density $\rho^* = \rho\sigma^3$, squared dipole moment $\mu^{*2} = \beta\mu^2/(4\pi\varepsilon_0\sigma^3)$ and squared charge $q^{*2} = \beta q^2/(4\pi\varepsilon_0\sigma)$. Note that in the literature, an alternative definition of the dimensionless dipole moment of dipolar LJ fluids is often used, which is given by $\tilde{\mu}^{*2} = \mu^2/(4\pi\varepsilon_0\epsilon\sigma^3) = T^*\mu^{*2}$.

6.4 Non-Primitive Model: Helmholtz Energy Contributions

The residual Helmholtz energy $a^{\mathrm{res}}(\rho^*, T^*, x^{\mathrm{ions}}, q^{*2}, \mu^{*2})$ of the model electrolyte systems considered in this work with respect to the ideal gas is given by

$$a^{\mathrm{res}} = a - a^{\mathrm{ig}}(NVT) = -\frac{1}{N}\ln Q^{\mathrm{res}}_{\mathrm{NVT}} \tag{6.6}$$

where $Q^{\mathrm{res}}_{\mathrm{NVT}}$ denotes the configurational canonical partition function. Using thermodynamic integration[41], the Helmholtz energy can be decomposed into two additive contributions

$$a^{\mathrm{res}} = a^{\mathrm{LJ}} + a^{\mathrm{el}} \tag{6.7}$$

where $a^{\mathrm{LJ}}(\rho^*, T^*)$ is the Helmholtz energy of a LJ reference system without any electrostatic interactions, and $a^{\mathrm{el}}(\rho^*, T^*, x^{\mathrm{ions}}, q^{*2}, \mu^{*2})$ is the Helmholtz energy contribution due to turning on the electrostatic (dipole-dipole, ion-dipole, and ion-ion) interactions. For a proper comparison to results from a primitive model approach (which does not include effects due

to dipole-dipole interactions) we separate the contribution due to dipole-dipole interactions using a two-step thermodynamic integration path, yielding

$$a^{\mathrm{el}} = a^{\mathrm{dd}} + a^{\mathrm{cc+cd}} \tag{6.8}$$

where $a^{\mathrm{dd}}(\rho^*, T^*, x^{\mathrm{ions}}, \mu^{*2})$ is the Helmholtz energy contribution related to dipole-dipole interactions and $a^{\mathrm{cc+cd}}(\rho^*, T^*, x^{\mathrm{ions}}, q^{*2}, \mu^{*2})$ is the ion-related Helmholtz energy contribution.

The two-step thermodynamic integration is described in detail in section 6.4.1. The full electrostatic Helmholtz energy contribution a^{el} can also be directly determined in a single thermodynamic integration step as proposed by Theiss and Gross[42]. This method is used to validate results obtained using the two-step thermodynamic integration, with details provided in section 6.4.2.

6.4.1 Two-Step Thermodynamic Integration

The non-primitive model Helmholtz energy contributions a^{dd} and $a^{\mathrm{cc+cd}}$ are defined using the two-step thermodynamic integration path, illustrated in figure 6.2. The integration path starts with a LJ reference system (figure 6.2, left). In the first step, the charges of the dipoles (and thereby the dipole moments μ) are turned on leading to a system of LJ and dipolar LJ particles (figure 6.2, centre). In the second step, the ion charges q are turned on, resulting in the fully charged electrolyte system (figure 6.2, right).

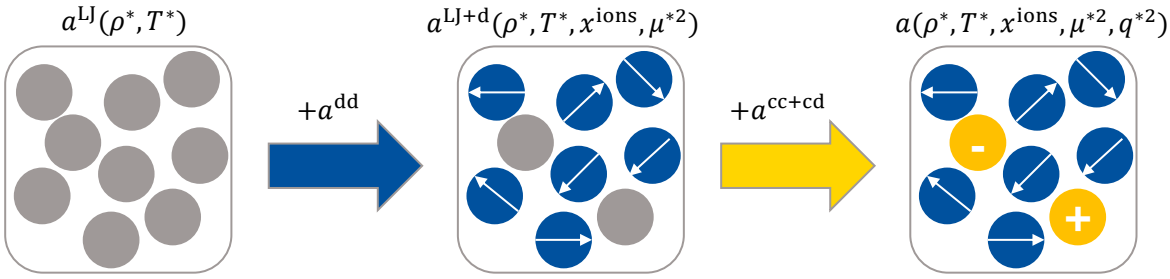


Figure 6.2: Schematic representation of the two step thermodynamic integration path defining the non-primitive model Helmholtz energy contributions.

To perform the first integration step, we linearly couple the potential energy of the LJ fluid and the potential energy of the LJ and dipolar LJ mixture (LJ + d) using a coupling parameter $0 \leq \lambda_1 \leq 1$,

$$U^{\mathrm{LJ+d}}_{\lambda_1} = U^{\mathrm{LJ}} + \lambda_1 U^{\mathrm{dd}} \tag{6.9}$$

The dipole-dipole Helmholtz energy contribution follows from integration over λ_1 as

$$a^{\mathrm{dd}} := \int_0^1 \left(\frac{\partial a_{\lambda_1}^{\mathrm{LJ+d}}}{\partial \lambda_1} \right)_{\mathrm{NVT}} \mathrm{d}\lambda_1 = \frac{\beta}{N} \int_0^1 \langle U^{\mathrm{dd}} \rangle_{\mathrm{NVT},\lambda_1} \mathrm{d}\lambda_1 \tag{6.10}$$

The angle brackets denote the ensemble average

$$\langle U^{\mathrm{dd}} \rangle_{\mathrm{NVT},\lambda_1} = \frac{1}{Z_{\lambda_1}} \int U^{\mathrm{dd}} e^{-\beta U_{\lambda_1}^{\mathrm{LJ+d}}} \mathrm{d}\mathbf{r}^N \mathrm{d}\omega^N \tag{6.11}$$

with the configurational integral

$$Z_{\lambda_1} = \int e^{-\beta U_{\lambda_1}^{\mathrm{LJ+d}}} \mathrm{d}\mathbf{r}^N \mathrm{d}\omega^N \tag{6.12}$$

The charge-dipole and charge-charge Helmholtz energy contribution is obtained in a second integration step starting from the dipolar LJ fluid as reference

$$U_{\lambda_2} = U^{\mathrm{LJ}} + U^{\mathrm{dd}} + \lambda_2 (U^{\mathrm{cc}} + U^{\mathrm{cd}}) \tag{6.13}$$

with $0 \leq \lambda_2 \leq 1$. The Helmholtz energy contribution emerging from ionic charges (step 2 of figure 6.2) is calculated as

$$a^{\mathrm{cc+cd}} := \frac{\beta}{N} \int_0^1 \langle U^{\mathrm{cc}} + U^{\mathrm{cd}} \rangle_{\mathrm{NVT},\lambda_2} \mathrm{d}\lambda_2 \tag{6.14}$$

Summation of eq. (6.10) and eq. (6.14) yields the total electrostatic contribution a^{el} as defined in eq. (6.8).

6.4.2 One-Step Thermodynamic Integration

To ensure the reliability of the results obtained using the two-step integration method, we use an alternative, one-step integration method for validation. In this approach, the charges of both ions and dipoles are turned on simultaneously. This enables the computation of the total electrostatic Helmholtz energy contribution a^{el} in a single thermodynamic integration step.

For this integration path, we define

$$U_\lambda = U^{\mathrm{LJ}} + \lambda U^{\mathrm{ES}} \tag{6.15}$$

with $0 \leq \lambda \leq 1$. The Helmholtz energy of the electrolyte system with respect to the Helmholtz energy of the LJ reference fluid follows as

$$a^{\mathrm{el}} = a - a^{\mathrm{LJ}} = \frac{\beta}{N} \int_0^1 \langle U^{\mathrm{ES}} \rangle_{\mathrm{NVT},\lambda} \mathrm{d}\lambda \tag{6.16}$$

Validation of the two-step integration method is performed by verifying that the total electrostatic contribution obtained from the one-step integration matches the sum of the dipole-dipole and ion-related contributions calculated using the two-step approach,

$$(a^{\mathrm{el}})_{\mathrm{TI-1}} - (a^{\mathrm{dd}} + a^{\mathrm{cc+cd}})_{\mathrm{TI-2}} \approx 0 \tag{6.17}$$

with equality holding within the statistical uncertainty of the results.

It is possible to decompose the electrostatic Helmholtz energy contributions from the one-step thermodynamic integration into three parts (a^{dd}, a^{cd}, a^{cd}). Whereas the sum of all Helmholtz energy contributions from a one-step thermodynamic integration and a two-step integration must be equal (eq. (6.17)), the dipolar contributions a^{dd} (for example) from the two thermodynamic integration procedures will differ because the structures of the underlying fluids differ.

6.5 Primitive Model: Helmholtz Energy Contributions

In eEoS based on the primitive model, the ion-related electrostatic Helmholtz energy contribution $a^{\mathrm{cc+cd}}$ is typically expressed as either

$$a^{\mathrm{cc+cd}} \approx a^{\mathrm{DH}} + a^{\mathrm{Born}} \equiv a^{\mathrm{DH+Born}} \tag{6.18}$$

which we refer to as the DH+Born approach, or

$$a^{\mathrm{cc+cd}} \approx a^{\mathrm{MSA}} + a^{\mathrm{Born}} \equiv a^{\mathrm{MSA+Born}} \tag{6.19}$$

denoted as the MSA+Born approach. a^{DH} and a^{MSA} are the Helmholtz energy contributions calculated using the DH equation[7] and the primitive MSA[8,9], respectively. Both are analytical models for the Helmholtz energy contribution related to ion-ion interactions. The analytic expression a^{Born}, based on the Born theory of solvation[10] is included to account for ion-dipole interactions.

In the following, we present the primitive model expressions applied to the molecular model introduced in section 6.3 and discuss the definitions of ion size parameters and permittivity as inputs.

Debye-Hückel Equation

The DH equation was derived by Debye and Hückel[7] based on a solution of the linearized Poisson-Boltzmann equation. With the model restrictions made in section 6.3, the DH equation for the reduced Helmholtz energy can be written as

$$a^{\mathrm{DH}} = \frac{A^{\mathrm{DH}}}{N k_{\mathrm{B}} T} = -\frac{x^{\mathrm{ions}} q^{*2} \kappa^* \chi(\kappa^*)}{3\varepsilon_{\mathrm{r}}} \tag{6.20}$$

with the dimensionless inverse screening length

$$\kappa^* = \kappa\sigma = \sqrt{\frac{4\pi\rho^*}{\varepsilon_r} x^{\mathrm{ions}} q^{*2}} \tag{6.21}$$

and

$$\chi(\kappa^*) = \frac{3}{(\kappa^*\sigma^{\mathrm{DH}*})^3}\left[\ln(1+\kappa^*\sigma^{\mathrm{DH}*}) - (\kappa^*\sigma^{\mathrm{DH}*}) + \frac{1}{2}(\kappa^*\sigma^{\mathrm{DH}*})^2\right] \tag{6.22}$$

where $\sigma^{\mathrm{DH}*} = \sigma^{\mathrm{DH}}/\sigma$ is a dimensionless ion size parameter and ε_{r} is the relative permittivity. Below, we test obvious choices of how to define $\sigma^{\mathrm{DH}*}$.

Mean-Spherical Approximation

The MSA[8] expression for the Helmholtz energy follows from solving the Ornstein-Zernike equation with a closure based on the low-density limit of the potential of mean force. The analytic solution of the MSA for the restricted primitive model (equally sized charged hard-

spheres) was derived by Waisman and Lebowitz[43,44] and can be written as[45]

$$a^{\mathrm{MSA}} = -\frac{3(\kappa^* \sigma^{\mathrm{M}*})^2 + 6(\kappa^* \sigma^{\mathrm{M}*}) + 2 - 2(1 + 2\kappa^* \sigma^{\mathrm{M}*})^{3/2}}{12\pi \rho^* \sigma^{\mathrm{M}*}} \tag{6.23}$$

with κ^* defined in eq. (6.21) and the dimensionless MSA ion size parameter $\sigma^{\mathrm{M}*}$.

Born Theory of Solvation

The term 'Born theory' (or 'Born term') is used in various contexts in the literature[32,33,46]. In this work, the Born theory specifically refers to the Helmholtz energy contribution given by

$$a^{\mathrm{Born}} = -\left(1 - \frac{1}{\varepsilon_{\mathrm{r}}}\right) \frac{x^{\mathrm{ions}} q^{*2}}{\sigma^{\mathrm{B}*}}, \tag{6.24}$$

where $\sigma^{\mathrm{B}*} = \sigma^{\mathrm{B}}/\sigma$ is the dimensionless Born size parameter. The Born size parameter σ^{B} can be interpreted as the diameter of the cavity that must be opened in a solvent to insert an ion.[39] Equation (6.24) corresponds to the formulation used in many eEoS, including SAFT-VRE[22] and ePC-SAFT advanced[23].

Ions Size Parameters in Primitive Model Equations

All primitive model expressions introduced in this section depend on specific ion size parameters (σ^{DH}, $\sigma^{\mathrm{M}*}$, $\sigma^{\mathrm{B}*}$, respectively), which are typically treated as adjustable parameters in eEoS. For a rigorous comparison between the non-primitive and primitive model Helmholtz energy contributions, we evaluate several obvious definitions of these size parameters.

First, we assume that all ion size parameters are equal to the LJ parameter σ, which is applied for both ions and solvent molecules. This assumption is revisited in section 6.7.2.2.

The primitive model expressions used in this work were originally derived for hard-sphere ions. However, in our electrolyte model, non-electrostatic interactions are described by the LJ potential, which allows partial overlapping of particles, particularly at high temperatures. To address this, we also evaluate the primitive model equations using a temperature-dependent effective hard-sphere diameter as the ion size parameter.

We investigate a third approach, inspired by the findings of Rashin and Honig[47], who reported that Born size parameters must be significantly larger (approximately 7%) than actual ionic diameters to accurately reproduce experimental hydration energies. Therefore, in our work,

the Born size parameter is empirically adjusted to minimize deviations between the predictions from primitive model expressions and our results from non-primitive model molecular simulations.

Relative Permittivity in Primitive Model Equations

The equations for the primitive model as summarized in eqs. (6.20), (6.23), and (6.24) were derived assuming a constant (composition independent) relative permittivity ε_r.[32,36] This permittivity could either be that of the pure solvent, or an effective value accounting for the dipolar fluid between ionic charges.[34–36,46,48] However, in real electrolyte solutions, beyond the infinite dilution limit, the permittivity exhibits a strong dependence on the ion concentration. Some studies have attempted to derive modified primitive model equations that incorporate concentration-dependent permittivities.[34,35] Others have simply substituted the concentration-dependent permittivity directly into the existing primitive model equations without a formal theoretical derivation.[49,50] Even tough they are empirical, these approaches (because they are parametrized) were shown to improve both, the accuracy and applicability of the primitive model equations.

In this work, we evaluate predictions from the primitive model expressions eqs. (6.20), (6.23), and (6.24) for two different, unambiguous definitions of the relative permittivity: (1) the permittivity of the pure solvent $\varepsilon_r^{\mathrm{solv}}(\rho^*, T^*, \mu^{*2})$, defined as the permittivity of a system consisting of only dipolar LJ particles with equal particle density ρ^* as the electrolyte solution, and (2) the concentration dependent permittivity of the electrolyte system $\varepsilon_r(\rho^*, T^*, x^{\mathrm{ions}}, \mu^{*2}, q^{*2})$.

In figure 6.10, results for both permittivity definitions are compared for two exemplary systems. Other definitions of the relative permittivity could be proposed but are not considered here, as the resulting values typically lie between those obtained for (1) and (2).

6.6 Simulation Details

The Helmholtz energy contributions due to dipole-dipole, charge-dipole, and charge-charge interactions defined in eqs. (6.10) and (6.14) are calculated by integrating the average intermolecular energies, $\langle U^{\mathrm{dd}}\rangle_{NVT,\lambda}$,$\langle U^{\mathrm{cd}}\rangle_{NVT,\lambda}$ and $\langle U^{\mathrm{cc}}\rangle_{NVT,\lambda}$, which we sample by means of NVT Monte Carlo (MC) simulations. All simulations were carried out with a total of $N = 2000$ particles in a cubic box of volume $V = L^3$, where L denotes the box length. We equilibrate the simulations for at least $5 \cdot 10^4$ MC cycles and consider at least $5 \cdot 10^4$ cycles for sampling (where one cycle comprises N MC moves). One-step thermodynamic integration simulations with 10^5

MC cycles for equilibration and 10^5 cycles for sampling were performed for validation. We perform translations of a random particle (dipole or ion) with a probability of 50% , rotations of dipole particles with 45% probability and additional translations of randomly chosen ions with a probability of 5% aiming to improve the statistics for the ion-ion contribution. The step sizes for translation and rotation are adjusted to achieve 50% acceptance rates for the moves.

Lennard-Jones interaction are explicitly evaluated up to a cut-off radius $r_c = 3.0\sigma$ and standard long-range tail-corrections are applied to account for the remainder. The LJ parameters are set to $\sigma = 1\text{Å}$ and $\varepsilon/k_B = 1\text{K}$.

The electrostatic energy contributions are evaluated with the standard Ewald summation[51] with $\alpha = 5.6$ as the Ewald parameter and seven reciprocal vectors to sum up the Fourier term. We compute the full electrostatic energy U^{el}, the electrostatic energy due to ion-ion interactions U^{cc} and the energy due to dipole-dipole interactions U^{dd} as

$$\begin{aligned} U^{\text{el}} &= U^{\text{real}} + U^{\text{fourier}} - U^{\text{self}} \\ U^{\text{cc}} &= U^{\text{real,cc}} + U^{\text{fourier,cc}} - U^{\text{self,cc}} \\ U^{\text{dd}} &= U^{\text{real,dd}} + U^{\text{fourier,dd}} - U^{\text{self,dd}} \end{aligned} \tag{6.25}$$

where the superscript "real" indicates the real-space contribution of the pair-interactions, "fourier" stands for the Fourier-space contribution of the Ewald sum and "self" indicates the correction for considering the interactions of charges with themselves. For the full electrostatic energy U^{el} the Ewald contributions are defined as

$$U^{\text{real}} = \frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{N^{\text{q}}} \sum_{j>i}^{N^{\text{q}}} q_i q_j \frac{\operatorname{erfc}\left(\alpha |\mathbf{r}_{ij}|/L\right)}{|\mathbf{r}_{ij}|} \tag{6.26}$$

$$U^{\text{fourier}} = \frac{1}{2\varepsilon_0 L^3} \sum_{k\neq 0} \frac{1}{k^2} \left| \sum_{i=1}^{N^{\text{q}}} q_i \exp\left(i\mathbf{k}\cdot\mathbf{r}_i\right) \right|^2 \exp\left(-\frac{k^2 L^2}{4\alpha^2}\right) \tag{6.27}$$

$$U^{\text{self}} = \underbrace{\frac{1}{4\pi\varepsilon_0} \frac{\alpha N^{\text{ions}} q^2}{\sqrt{\pi} L}}_{=U^{\text{self,cc}}} + \underbrace{\frac{N^{\text{dipoles}}}{4\pi\varepsilon_0} \left(\frac{2\alpha\mu^2}{l\sqrt{\pi}L} - \frac{\mu^2}{l^3} \operatorname{erf}\left(\frac{\alpha}{L} l\right) \right)}_{=U^{\text{self,dd}}} \tag{6.28}$$

where $N^{\text{q}} = 2N^{\text{dipoles}} + N^{\text{ions}}$ is the number of all charges in the system. To compute the component-specific electrostatic energy contributions U^{cc} and U^{dd}, eqs. (6.26) and (6.27) are summed only over the charges of ions or dipoles, respectively. We obtain the ion-dipole

contributions U^{cd} from

$$U^{\mathrm{cd}} = U^{\mathrm{el}} - U^{\mathrm{cc}} - U^{\mathrm{dd}} \tag{6.29}$$

The direct calculation of

$$U^{\mathrm{cd}} = U^{\mathrm{real,cd}} + U^{\mathrm{fourier,cd}} \tag{6.30}$$

is also possible but requires the computationally expensive evaluation of a double sum over all charges in the calculation of the Fourier contribution $U^{\mathrm{fourier,cd}}$.[52] The simulation code used in this work is an extension of the implementation of Theiss and Gross[42,53].

The integrals defining the Helmholtz energy contributions of eq. (6.10), eq. (6.14) and eq. (6.16) are calculated numerically based on Simpson's rule, with 21 equidistant values of λ_1, λ_2 and λ, respectively.

The relative permittivity ε_{r} is only sampled for $\lambda = 1$, as

$$\varepsilon_{\mathrm{r}} = 1 + \frac{4}{3V}\pi\beta\left(\langle \mathbf{M}^2\rangle - \langle \mathbf{M}\rangle^2\right) \tag{6.31}$$

where the total dipole moment of the simulation box is defined as

$$\mathbf{M} = \sum_{i=1}^{N_{\mathrm{dipoles}}} \boldsymbol{\mu}_i \tag{6.32}$$

and $\boldsymbol{\mu}_i$ defines the dipole vector $\boldsymbol{\mu}_i = \mu\hat{\boldsymbol{\mu}}_i$, where $\hat{\boldsymbol{\mu}}_i$ is the orientational unit vector of the dipole of particle i.[52]

95% confidence intervals of sampled mean electrostatic energies, Helmholtz energies and relative permittivities, are computed using a Bootstrap method with a resampled data size of 500.[54]

Table 6.1 gives an overview of all model systems examined in this study. We consider two different temperatures $T^* = 1.8$ and $T^* = 4.0$ and ion mole fractions $0.005 \leq x^{\mathrm{ions}} \leq 0.05$. The calculated relative permittivity of the pure solvent is also included.

Molecular simulation results for the Helmholtz energy contributions obtained using the two-step thermodynamic integration path defined in section 6.4.1, together with the concentration dependent permittivities, are provided in tables D.1–D.3. The average electrostatic energy contributions $\beta U^{\mathrm{cc}}/N$, $\beta U^{\mathrm{cd}}/N$ and $\beta U^{\mathrm{dd}}/N$ for the full electrolyte systems ($\lambda_2 = 1$) are listed in the supplementary material [of the online version of the publication]. Results for the

Helmholtz energy contributions obtained with the one-step simulation path are also provided in the supplementary material [of the online version].

Table 6.1: Specification of all simulated systems, including the relative permittivity of the solvent $\varepsilon_r^{solv}(\mu^{*2}, \rho^*, T^*)$ obtained from molecular simulations of the pure solvent. Uncertainties in the last digit, based on the 95 % confidence interval, are shown in parentheses.

ρ^*	T^*	$\tilde{\mu}^{*2}$	μ^{*2}	q^{*2}	ε_r^{solv}
0.8	1.8	2.25	1.25	15, 30, 50	12.02(21)
0.8	1.8	3.6	2	15, 30, 50	29.4(20)
0.8	1.8	5.4	3	50, 100	87(8)
0.2	4	2	0.5	15	1.47(1)
0.2	4	4	1	15, 25	2.07(2)
0.2	4	5	1.25	15, 25	2.45(2)
0.8	4	2	0.5	15, 25	3.65(4)
0.8	4	4	1	25	8.37(12)
0.8	4	5	1.25	15, 25, 50	11.84(16)
0.8	4	8	2	25, 50	27.2(12)
0.8	4	16	4	25, 50	198(42)

6.7 Results

6.7.1 Assessment of Molecular Simulation Results

All results for the total electrostatic Helmholtz energy contribution a^{el} (defined in eq. (6.8)) were validated with results obtained using using the one-step thermodynamic integration method described in section 6.4.2. The mean absolute relative deviation of the Helmholtz energy between the results of the two methods was 0.095%.

As an additional check for correctness of our thermodynamic integration procedure, we plot the averages $\langle \beta U^{dd}/N \rangle_{\lambda_1}$ and $\langle \beta \left(U^{cc} + U^{cd}\right)/N \rangle_{\lambda_2}$ against the coupling parameters λ_1 and λ_2 in figure 6.3. According to the Gibbs-Bogoliubov inequality[41], these averages must decrease monotonically with increasing coupling parameters. Although the correct behaviour is observed for the system shown in figure 6.3, deviations from this monotonic behaviour were observed in certain systems, particularly for $\langle \beta \left(U^{cc} + U^{cd}\right) \rangle_{\lambda_2}$. Such deviations are unphysical and indicate that the simulated system is not well equilibrated. We identified two phenomena that cause deviations from the Gibbs-Bogoliubov inequality:

1. Ion clustering: In systems with low relative permittivity or high ionic charges, ions

tend to form neutral pairs or larger aggregates for λ_2-values above a certain threshold. This clustering behaviour was also reported by Theiss and Gross[42] and Eggebrecht and Ozler[55] for hard-sphere model electrolyte systems. Cluster formation is favoured under conditions of low density, small dipole moments, high temperatures or increased ion mole fractions, all of which reduce the relative permittivity. Ions within clusters are partially shielded from interactions with dipolar solvent molecules, leading to a sharp decrease in the magnitude of the ion-dipole energy $|\langle U^{\mathrm{cd}}\rangle_{\lambda_2}|$. This leads to a violation of the Gibbs-Bogoliubov inequality.

2. Ion shielding: In systems characterized by a combination of high permittivity and low ionic charges, we observed a decrease in the magnitude of the ion-ion energy $|\langle U^{\mathrm{cc}}\rangle_{\lambda_2}|$ while increasing the coupling parameter λ_2. This behaviour is likely due to the dominance of dipole-ion interactions over ion-ion interactions in such systems, leading to a clustering of dipolar molecules around the ions, effectively shielding ion-ion interactions.

Both phenomena described above cause long-lasting non-equilibrium configurations in the molecular simulations. All state points for which we observed a violation of the monotonic behaviour required by the Gibbs-Bogoliubov inequality were excluded from further analysis. A total of 107 state points were included in the evaluation.

Besides particle clustering, finite-size effects can significantly affect the quality of results from molecular simulations of electrolyte systems.

One indicator of such effects is the ion-ion energy contribution $\langle U^{\mathrm{cc}}\rangle_{\lambda_2=0}$, obtained at $\lambda_2 = 0$. As shown in figure 6.3, the ensemble-averaged dipole-dipole energy contribution $\langle U^{\mathrm{dd}}\rangle_{\lambda_1=0}$ is approximately zero. This is because, in simulations with $\lambda_1 = 0$, the sampled configurations correspond to a pure LJ system. The equal number of positive (like-charged) and negative (unlike-charged) contributions to the dipole-dipole energy results in the effective cancellation of the sampled dipole-dipole energy.

For $\lambda_2 = 0$ the system's configurations are those of a system consisting of LJ particles dissolved in a dipolar LJ solvent. The sampled ion-dipole interaction energy within this systems also effectively cancels, $\langle U^{\mathrm{cd}}\rangle_{\lambda_2=0} \approx 0$, as the number of positive and negative contributions to U^{cd} are equal. However, a non-zero ion-ion energy contribution, $\langle U^{\mathrm{cc}}\rangle_{\lambda_2=0}$, is obtained. The total number of like-charged (anion-anion or cation-cation) interactions, which contribute positively to $\langle U^{\mathrm{cc}}\rangle_{\lambda_2=0}$, is given by $N^{\mathrm{ions}}(N^{\mathrm{ions}}-2)/4$ whereas the number of unlike (anion-cation) interactions scales as $(N^{\mathrm{ions}}/2)^2$. Consequently, in systems with finite size, a negative non-zero energy contribution from ion-ion interactions persists.

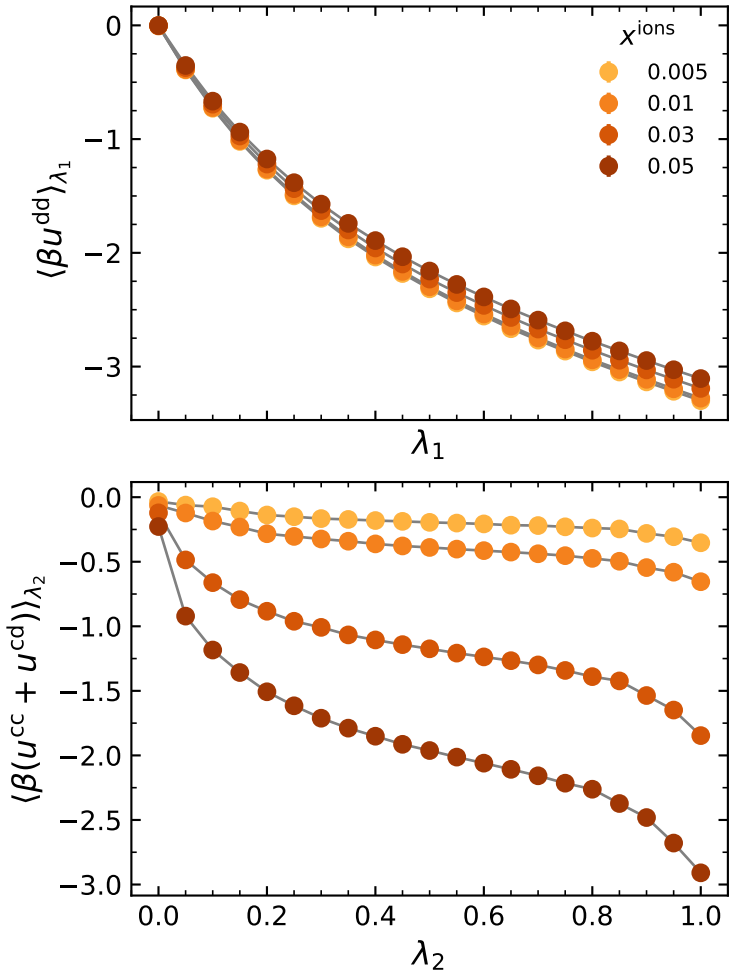


Figure 6.3: Average energy contributions from molecular simulations with different ion mole fractions at $\mu^{*2} = 2.0$, $q^{*2} = 50$, $\rho^* = 0.8$, $T^* = 4.0$ versus coupling parameter λ_1 (first step of the thermodynamic integration, upper plot) and λ_2 (second step, lower plot). The symbols are simulation results, the lines are a guide to the eye.

To assess the impact of finite-size effects on the calculated Helmholtz energy contribution $a^{\text{cc+cd}}$, a finite-size study was conducted. The results are shown in figure 6.4. The upper diagram shows the Helmholtz energy contribution $a^{\text{cc+cd}}$ obtained from molecular simulations with overall particle numbers ranging from $N = 400$ to $N = 2400$. The lower diagram shows the ensemble averages of the ion-ion energy at $\lambda_2 = 0$ for the same range of particle numbers. As expected, $\langle U^{\text{cc}} \rangle_{\lambda_2=0}$ tends to become less negative with increasing system size. For the final Helmholtz energy contributions $a^{\text{cc+cd}}$, however, no system size dependency is visible, supporting the reliability of the calculated results.

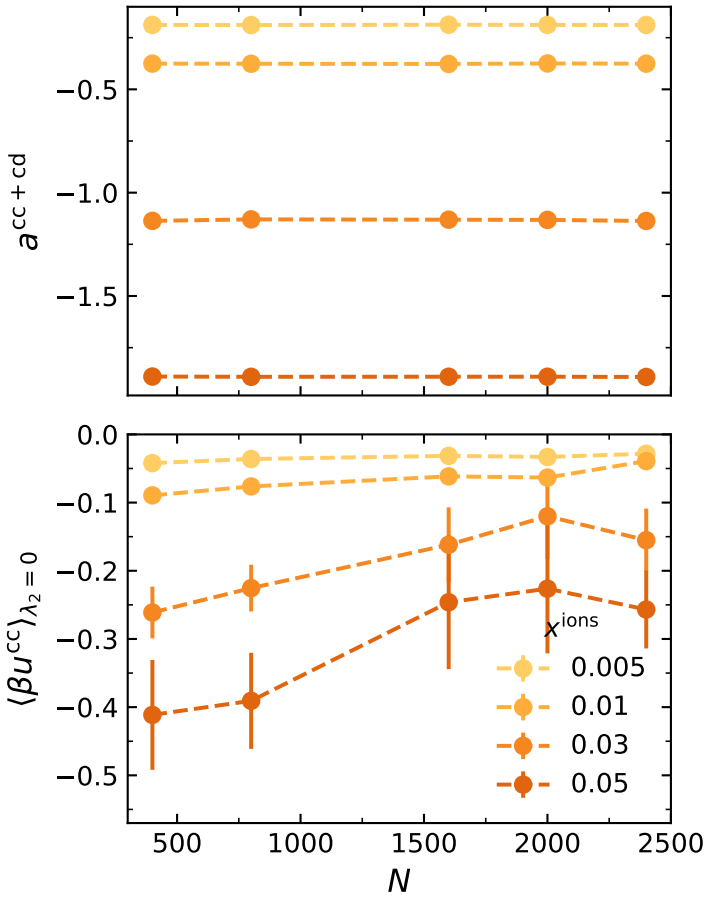


Figure 6.4: Finite-size effect study. Ion related Helmholtz energy contribution a^{cc+cd} and potential energy due to charge-charge interactions at $\lambda_2 = 0$ from molecular simulations with different numbers of particles $400 \leq N \leq 2400$ at $\mu^{*2} = 2.0$, $q^{*2} = 50$, $\rho^* = 0.8$, $T^* = 4.0$. The lines are a guide to the eye. The upper diagram shows no significant finite-size effect in the electrostatic Helmholtz energy contributions. The lower diagram analyzes the energy-increment for $\lambda_2 = 0$ (see figure 6.3), which has to disappear for infinite-sized systems.

6.7.2 Comparison of Helmholtz Energy Contributions from the Primitive and Non-Primitive Models

In figure 6.5, we compare estimates obtained from the DH equation together with the Born theory of solvation, both evaluated with the relative permittivity of the solvent ε_r^{solv}, to our results a^{cc+cd} from molecular simulation. For all systems shown, $a^{DH+Born}$ qualitatively captures the trends of a^{cc+cd} with ion concentration and ion charge but overestimates its magnitude. This overestimation is further emphasized in figure 6.6 where $a^{DH+Born}$ is compared to our simulation results of a^{cc+cd} for all simulated state points. For each state point, we observe that $a^{DH+Born} < a^{cc+cd}$.

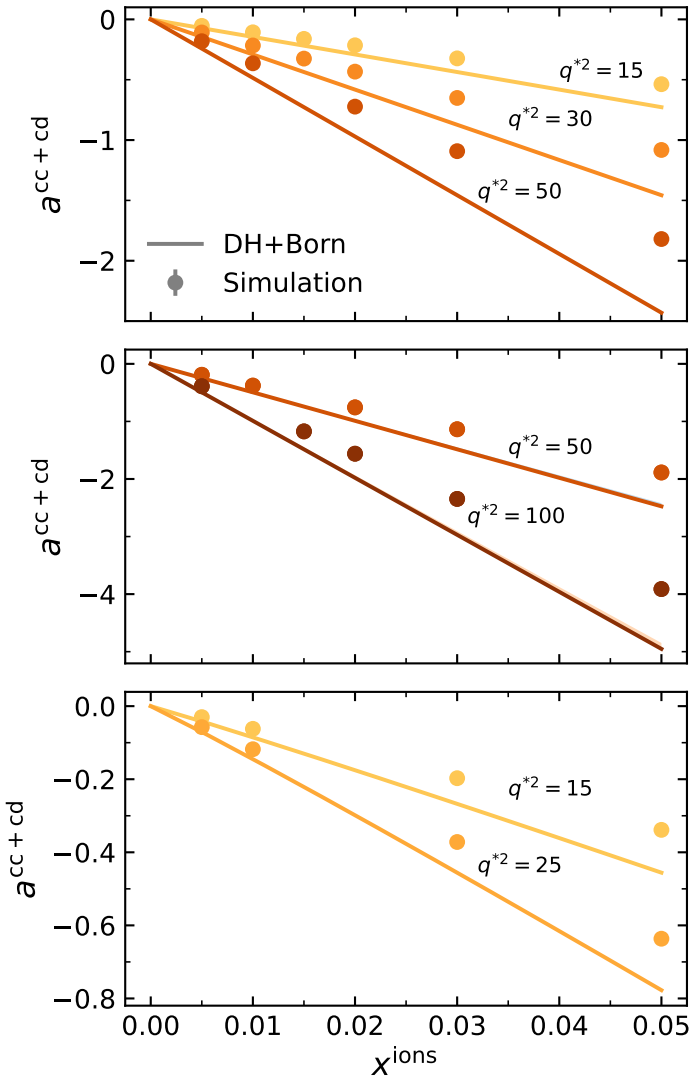


Figure 6.5: Molecular simulation results for the charge-charge and charge-dipole Helmholtz energy contribution (symbols) compared to Helmholtz energy estimates obtained using the DH equation and the Born theory, evaluated based on the relative permittivity of the solvent (lines). Top: $T^* = 1.8$, $\rho^* = 0.8$, $\mu^{*2} = 2$, middle: $T^* = 1.8$, $\rho^* = 0.8$, $\mu^{*2} = 3$, bottom: $T^* = 4.0$, $\rho^* = 0.2$, $\mu^{*2} = 1.0$

In figure 6.7, the relative deviations between a^{cc+cd} and $a^{DH+Born}$ are shown schematically for systems with varying ion mole fraction at temperature $T^* = 4$ and densities $\rho^* = 0.8$ (top) and $\rho^* = 0.2$ (bottom). For all systems considered in this work, the relative deviations between the primitive and the non-primitive Helmholtz energy contributions decrease with increasing ion charge q^{*2}. Contrary to our expectation that the results would agree better at low ion concentrations, no clear dependency on x^{ions} can be observed. The relative deviations across all simulated state points range from 12% to 45%.

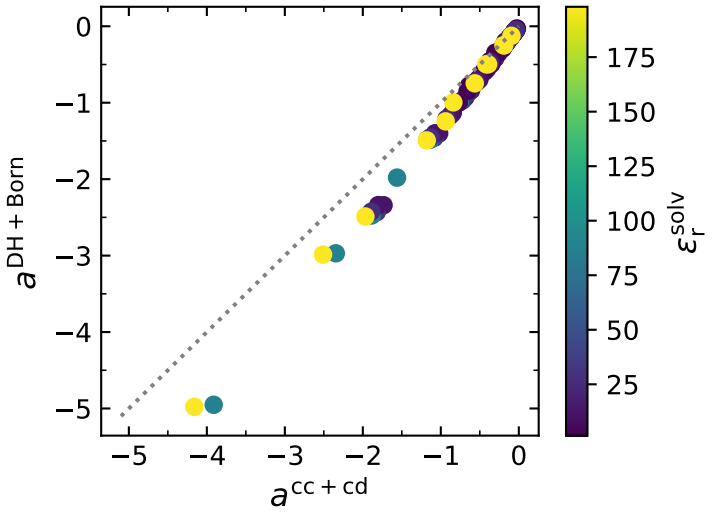


Figure 6.6: Parity plot, comparing the ion related Helmholtz energy contribution obtained using the DH+Born approach $a^{\mathrm{DH+Born}}$ to molecular simulation results for the non-primitive model $a^{\mathrm{cc+cd}}$ for all simulated systems.

We quantify the overall deviation of the primitive and non-primitive model results with the mean absolute relative deviation (MARD) across all simulated state points,

$$\mathrm{MARD} = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{a_i^{\mathrm{DH+Born}} - a_i^{\mathrm{cc+cd}}}{a_i^{\mathrm{cc+cd}}}\right| \times 100\% \tag{6.33}$$

where m is the number of all calculated Helmholtz energy contributions (here, $m = 107$) and the index i denotes a state point defined by $x^{\mathrm{ions}}, T^*, \rho^*, \mu^{*2}, q^{*2}$ and the permittivity of the solvent $\varepsilon_{\mathrm{r}}^{\mathrm{solv}}$. We obtain a MARD of 31.5%.

As discussed in section 6.2, the Helmholtz energy contribution related to ion-ion interactions is commonly described using either the DH equation or the primitive MSA result. In figure 6.8, we compare the Helmholtz energies calculated based on both models for all state points considered in this work. All points are relatively close to the dashed line that marks $a^{\mathrm{MSA}} = a^{\mathrm{DH}}$.

In figure 6.9, the Helmholtz energy contributions a^{DH} (solid lines) and a^{MSA} (dashed lines) are plotted versus x^{ions} for system with varying ion charges q^{*2}. At low ion mole fractions, the predictions from both equations agree closely; however, at higher mole fractions, the MSA yields slightly more negative values compared to the DH equation. Since the DH+Born approach already predicts more negative values than those observed for $a^{\mathrm{cc+cd}}$, the MSA+Born approach deviates even further from $a^{\mathrm{cc+cd}}$. The MARD between $a^{\mathrm{MSA+Born}}$ and $a^{\mathrm{cc+cd}}$ across all state points is 31.7%.

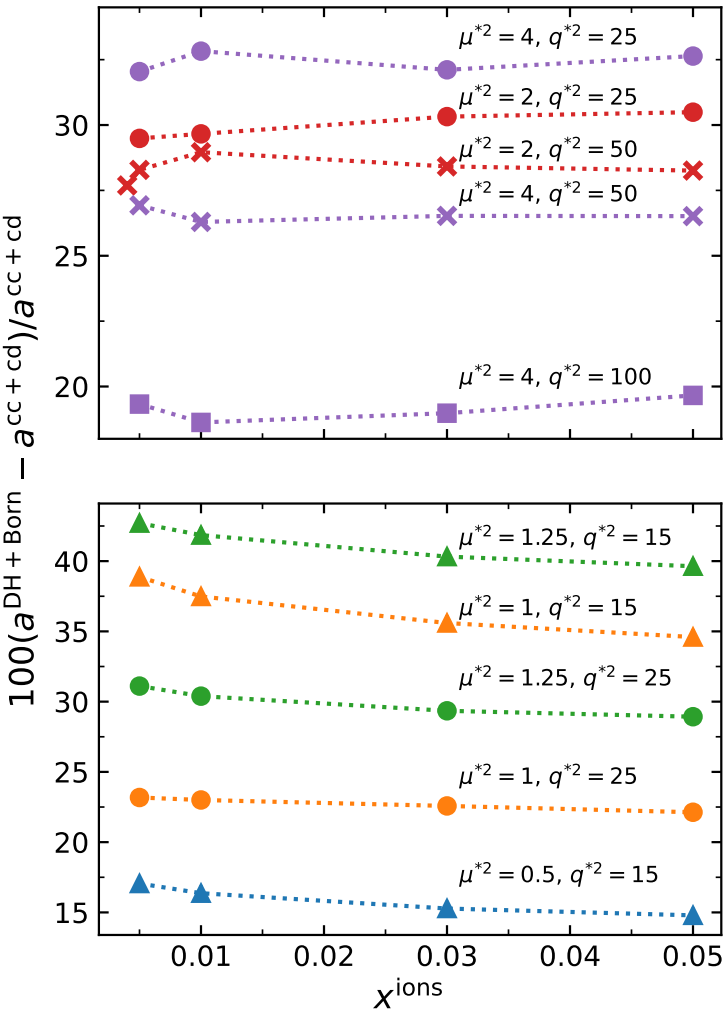


Figure 6.7: Relative deviations between the charge-charge and charge-dipole Helmholtz energy contribution obtained with molecular simulations $a^{\mathrm{cc+cd}}$ and calculated with the DH+Born approach $a^{\mathrm{Born+DH}}$ at temperature $T^* = 4$ and densities $\rho^* = 0.8$ (upper plot) and $\rho^* = 0.2$ (lower plot). The primitive model equations are evaluated with the permittivity of the solvent $\varepsilon_{\mathrm{r}}^{\mathrm{solv}}$. Lines are a guide to the eye.

For the absolute majority of systems considered in this work, we find that the Helmholtz energy contribution from the DH/MSA equations is multiple orders of magnitude smaller than the Helmholtz energy contribution obtained with the Born theory. See, for example the vastly different values for a^{cc} in figure 6.8 compared to values for $a^{\mathrm{cc+cd}}$ in figure 6.6. The Born term dominates the full electrostatic Helmholtz energy contribution $a^{\mathrm{cc+cd}}$, whereas the impact of the ion-ion expressions (DH or MSA) is marginal. This observation aligns with the findings of Kournopoulos *et al.* [39], who reported that for their test system (NaCl in water), the Born theory contributes several orders of magnitude more to free energy of solvation than DH/MSA.

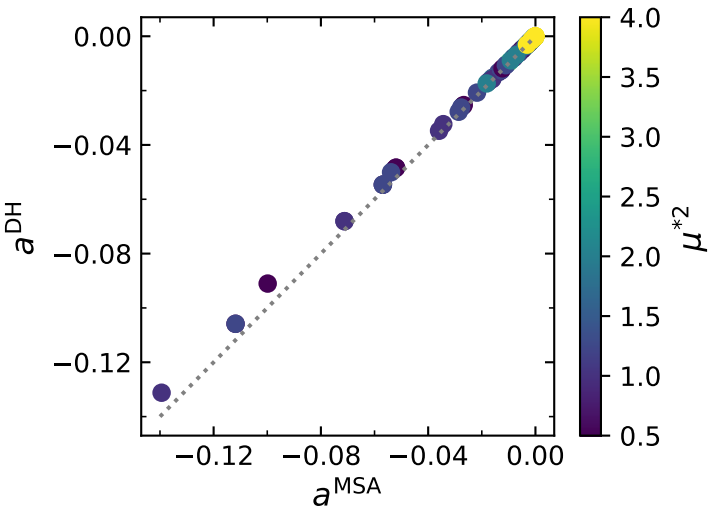


Figure 6.8: Helmholtz energies calculated with MSA and DH (evaluated with solvent permittivities) for all simulated state points.

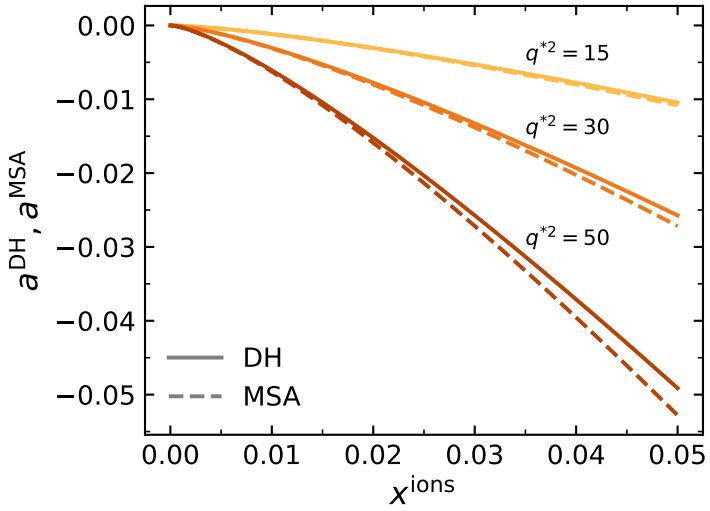


Figure 6.9: Charge-charge Helmholtz energy contribution calculated with the DH equation (solid lines) and the MSA (dashed lines) at temperature $T^* = 1.8$ and density $\rho^* = 0.8$.

Given that the MSA+Born results closely follow the trends observed in the DH+Born approach, we will focus solely on the DH+Born approach in the following analysis.

6.7.2.1 Impact of the Relative Permittivity

So far, evaluations of the primitive model equations were conducted using the permittivity of the solvent $\varepsilon_r^{solv}(T^*, \rho^*, \mu^{*2})$, as provided in table 6.1. In this section, we investigate the effects of using the full, concentration-dependent permittivity $\varepsilon_r(T^*, \rho^*, \mu^{*2}, q^{*2}, x^{ions})$ in the calculations.

In figure 6.10 the relative permittivity obtained from molecular simulations is shown as a function of the ion mole fraction for two representative systems. As the ion concentration increases, the orientational ordering of the dipole molecules around the ions decreases, leading to a significant decrease in permittivity ε_r.

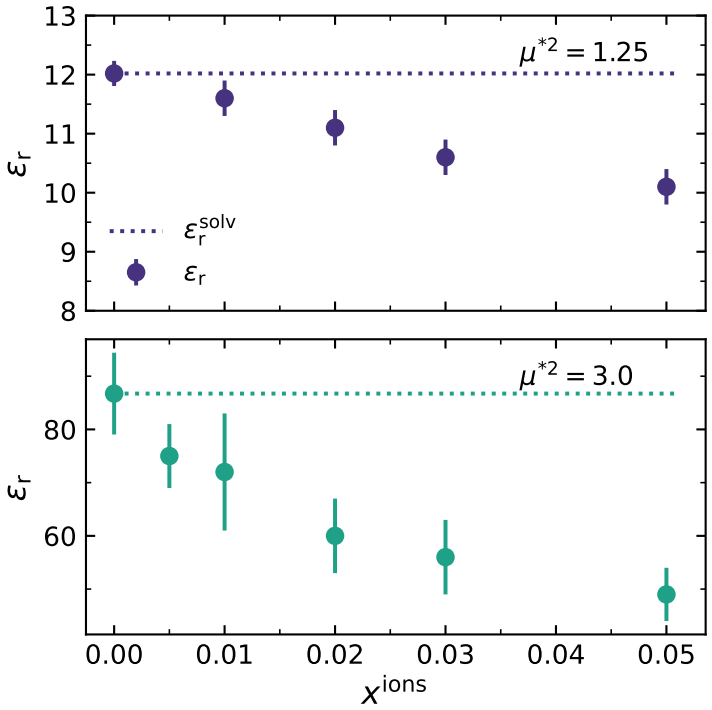


Figure 6.10: Relative permittivities $\varepsilon_r(x^{ions}, \rho^*, T^*, \mu^{*2}, q^{*2})$ (symbols) and $\varepsilon_r^{solv}(\rho^*, T^*, \mu^{*2})$ (dotted lines) from molecular simulations for two electrolyte systems with temperature $T^* = 1.8$, density $\rho^* = 0.8$, squared ion charges $q^{*2} = 50$ and squared dipole moments $\mu^{*2} = \{1.25, 3.0\}$.

In figure 6.11 the DH equation and the Born theory are evaluated for the same systems using the permittivity of the solvent, ε_r^{solv} (lines), and the concentration-dependent permittivity, ε_r (symbols). Although figure 6.10 shows a strong ion concentration dependence of permittivity, the resulting difference in the primitive model Helmholtz energy contributions is relatively small. Specifically for the Born theory, using the concentration dependent permittivity leads to

marginal increase in the Helmholtz energy compared to using the permittivity of the solvent. At high permittivity values, this difference vanishes, since $\lim_{\varepsilon_r \to \infty} \frac{\varepsilon_r - 1}{\varepsilon_r} = 1$. For the DH equation, more pronounced differences are observed, particularly at elevated ion concentrations. Since the DH equation approximately scales as $a^{\mathrm{DH}} \sim 1/\varepsilon_r$, using the concentration-dependent permittivity ε_r instead of $\varepsilon_r^{\mathrm{solv}}$ leads to larger magnitudes of a^{DH}.

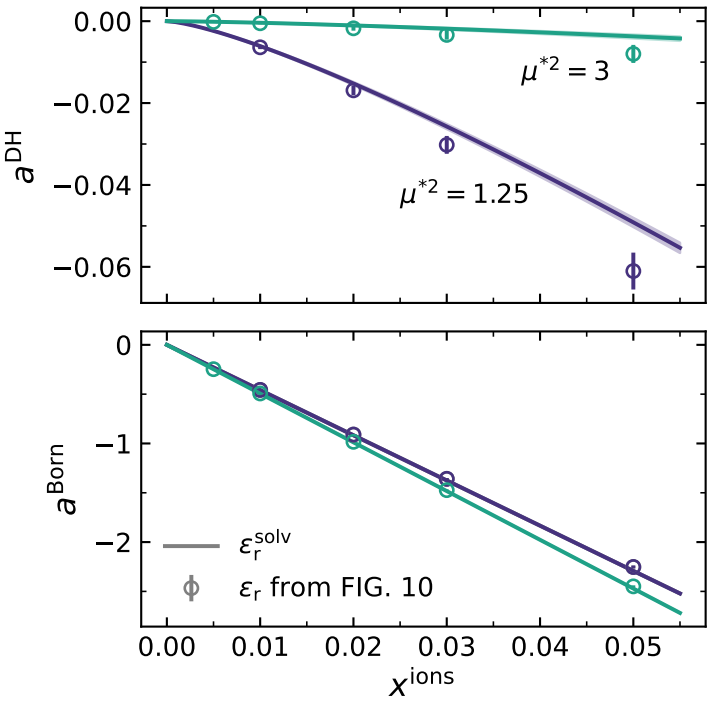


Figure 6.11: DH equation and Born theory evaluated with $\varepsilon_r^{\mathrm{solv}}(\rho^*, T^*, \mu^{*2})$ (lines) and $\varepsilon_r(x^{\mathrm{ions}}, \rho^*, T^*, \mu^{*2}, q^{*2})$ (symbols) for two electrolyte systems with temperature $T^* = 1.8$, density $\rho^* = 0.8$, squared ion charges $q^{*2} = 50$ and squared dipole moments $\mu^{*2} = \{1.25, 3.0\}$. The shaded areas and the error bars visualize the uncertainty ranges, corresponding to the DH equation and Born theory evaluated with the respective permittivity ± its 95%-confidence interval.

In comparing the Helmholtz energy contributions from the primitive and non-primitive models, we focus on the sum of the DH equation and the Born theory. Because the Born term dominates this sum for all systems considered in this work, the choice of permittivity definition becomes rather unimportant for the comparison. The MARD between $a^{\mathrm{DH+Born}}(\varepsilon_r)$ and $a^{\mathrm{cc+cd}}$ over all simulation results is 30.62% and thus slightly lower than the MARD obtained using the solvent permittivities (which was 31.5%).

6.7.2.2 Impact of the Born Size Parameter

Until this section, we have set the Born and DH size parameters to be identical and constant, $\sigma^{\mathrm{B}} = \sigma^{\mathrm{DH}} = \sigma$. In the following, we explore two alternative definitions of the size parameters:

1. Temperature dependent diameters: In this approach, both the Born and DH size parameters are defined as

$$\sigma^{\mathrm{B}}(T^*) = \sigma^{\mathrm{DH}}(T^*) = d(T^*) \tag{6.34}$$

   where $d(T^*)$ is the effective hard-sphere diameter for a LJ fluid proposed by van Westen [56] (Eq. (21) in Ref. 56). For the temperatures considered in this work, the effective diameters are $d(T^* = 4) = 0.96068\sigma$ and $d(T^* = 1.8) = 0.994978\sigma$.

2. Optimized Born size parameter: In this approach, the Born size parameter is adjusted to reproduce our molecular simulation results. Because the contribution of the DH term plays a subordinate role compared to the Born term, the DH size parameter remains $\sigma^{\mathrm{DH}}(T^*) = d(T^*)$. The Born size parameter is defined as

$$\sigma^{\mathrm{B}}(T^*, q^{*2}) = \alpha(q^{*2}) d(T^*) \tag{6.35}$$

   where $\alpha(q^{*2})$ is a dimensionless parameter obtained by minimizing the relative deviations between $a^{\mathrm{cc+cd}}$ and $a^{\mathrm{DH+Born}}$ individually for each ion (defined by q^{*2}).

Figure 6.12 compares $a^{\mathrm{DH+Born}}$ evaluated with different size parameters (eq. (6.34), eq. (6.35)) to $a^{\mathrm{cc+cd}}$. The results for $\sigma^{\mathrm{B}} = \sigma^{\mathrm{DH}} = d(T^*)$ (circles in figure 6.12) deviate more from the simulation results than those obtained with $\sigma^{\mathrm{B}} = \sigma^{\mathrm{DH}} = \sigma$ (diamonds). The overall MARD when using $\sigma^{\mathrm{B}} = \sigma^{\mathrm{DH}} = d(T^*)$ increases to 34.85% if solvent permittivities are used and to 33.93% if concentration-dependent permittivities are applied (whereas the MARD was 31.5% for all size parameters set to σ).

Using a temperature-dependent diameter $d(T^*)$ instead of σ for calculating $a^{\mathrm{DH+Born}}$ has only a minor impact on the results for $T^* = 1.8$ (unfilled symbols). However, it shifts the results for $T^* = 4$ (filled symbols) to more negative values. The relationship between $a^{\mathrm{DH+Born}}$ and $a^{\mathrm{cc+cd}}$ for $\sigma^{\mathrm{DH}} = \sigma^{\mathrm{B}} = d(T^*)$ can be roughly approximated by $a^{\mathrm{DH+Born}} \approx 1.34 \times a^{\mathrm{cc+cd}}$. A study with more temperatures would be needed to determine whether this relationship holds in general.

Inspired by the observed linear relationship between $a^{\mathrm{DH+Born}}$ and $a^{\mathrm{cc+cd}}$ when using temperature dependent size parameters, we compute optimized Born size parameters using the Ansatz function eq. (6.35). The results for $a^{\mathrm{DH+Born}}$ evaluated with $\sigma^{\mathrm{DH}} = d(T^*)$ and

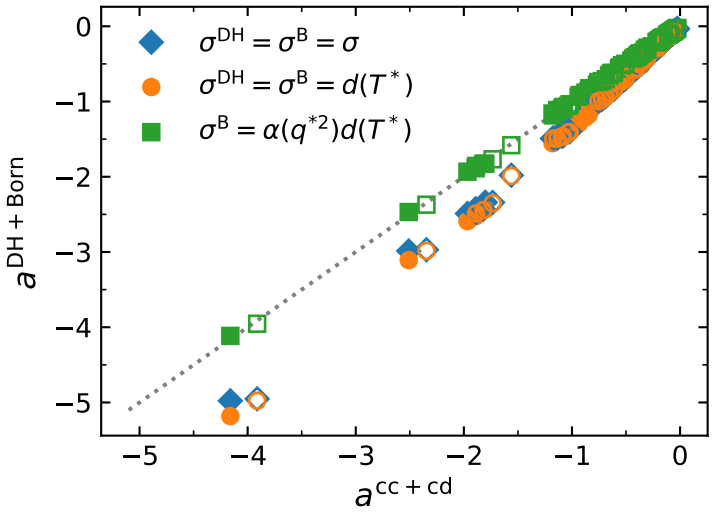


Figure 6.12: Helmholtz energy from primitive model expressions evaluated with different ion size parameters and solvent permittivities. Filled symbols indicate results for $T^* = 4$, unfilled symbols correspond to $T^* = 1.8$.

$\sigma^{B} = \alpha(q^{*2})d(T^*)$ are represented by the squares in figure 6.12. The optimized $\alpha(q^{*2})$—values, shown in figure 6.13, can be reasonably approximated by a linear function of the squared charge q^{*2}. Using these optimized values to calculate σ^{B}, the MARD between $a^{Born+DH}$ and a^{cc+cd} is reduced to 2.13%. Similarly, optimizing with concentration-dependent permittivities yields slightly lower $\alpha(q^{*2})$—values and a MARD of 2.0%.

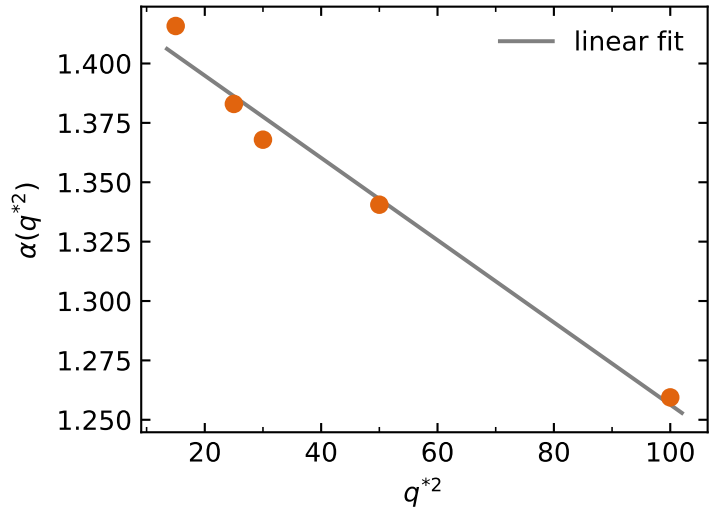


Figure 6.13: Linear scaling factor for the optimized Born size parameter of eq. (6.35) obtained with ε_r^{solv}. Symbols indicate optimization results, the line is a linear fit.

The findings of this section indicate that (1) using a temperature-dependent effective hard-

sphere diameter does not solve the large discrepancies between the Helmholtz energy contributions of the primitive and non-primitive models, and (2) adjusting parameters, particularly the Born size parameter, is both necessary and effective to reduce these discrepancies. However, the physical basis of such empirical adjustments is questionable, especially considering that the optimal values obtained for the Born size parameter are approximately 20–45% larger than the actual ion sizes used in the molecular simulations.

6.8 Conclusion

This work presents a rigorous comparison between the primitive model and the non-primitive model for electrolyte solutions. Molecular simulation data of model electrolytes based on the non-primitive model are generated, including Helmholtz energy contributions related to dipole-dipole, ion-ion and ion-dipole interactions, as well as relative permittivities for two temperatures and over a wide range of densities, ion concentrations, ion charges and dipole moments. The simulation results are compared to the electrostatic Helmholtz energy contributions predicted by classical primitive model expressions, in particular the Debye-Hückel (DH) equation, the primitive MSA and the Born theory of solvation. These comparisons are made for ion concentrations exceeding the low concentration limit for which the theories were originally formulated.

The comparison shows only qualitative agreement between the predictions based on the primitive model and our simulation results. The numerical values of the Helmholtz energy contributions differ significantly, with the primitive model approaches consistently leading to more negative values. Different definitions of the input parameters required for the primitive model equations, including ion size and relative permittivity, are investigated. Achieving quantitative agreement between the Helmholtz energy contributions required empirical adjustments to the Born size parameter, although the physical justification for these adjustments remains unclear, as the optimized parameters were significantly larger than the actual ion sizes used in the molecular simulations.

An additional insight into the comparison of primitive and non-primitive models could come from evaluating whether the DH limiting law is reproduced at low ion concentrations in the non-primitive framework. This requires isolating the ion-ion contribution to the Helmholtz energy and simulating large particle systems. The thermodynamic integration leading to Helmholtz energies using molecular simulations, as applied in this study was found to be unsuited for this purpose. Further refinement of the simulation method is required and remains the scope of future work.

Our findings highlight the need for further development of non-primitive approaches that can accurately capture the effect of charge-charge and charge-dipole interactions on the Helmholtz energy. The simulation data generated, together with the methods developed in this work for separating Helmholtz energy contributions related to dipole-dipole and ion-solvent interactions, should be valuable for the future development of accurate and physically rigorous eEoS.

References

[1] G. M. Kontogeorgis, A. Schlaikjer, M. D. Olsen, B. Maribo-Mogensen, K. Thomsen, N. von Solms, and X. Liang. A review of electrolyte equations of state with emphasis on those based on cubic and cubic-plus-association (CPA) models. *Int. J. Thermophys.*, **43**(4):54, 2022.

[2] M. Theiss and J. Gross. Perturbation approaches for describing dipolar fluids and electrolyte solutions. *J. Chem. Phys.*, **153**(4), 2020.

[3] L. Blum. Solution of the mean spherical approximation for hard ions and dipoles of arbitrary size. *J. Stat. Phys.*, **18**(5):451–474, 1978.

[4] L. Blum and d. Wei. Analytical solution of the mean spherical approximation for an arbitrary mixture of ions in a dipolar solvent. *J. Chem. Phys.*, **87**(1):555–565, 1987.

[5] D. Wei and L. Blum. The mean spherical approximation for an arbitrary mixture of ions in a dipolar solvent: Approximate solution, pair correlation functions, and thermodynamics. *J. Chem. Phys.*, **87**(5):2999–3007, 1987.

[6] R. A. Robinson and R. H. Stokes. *Electrolyte Solutions*. Courier Corporation, 2002.

[7] P. Debye and E. Hückel. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Phys. Z.*, **24**(9):185–206, 1923.

[8] L. Blum. Mean spherical model for asymmetric electrolytes: I. Method of solution. *Mol. Phys.*, **30**(5):1529–1535, 1975.

[9] L. Blum and J. Høye. Mean spherical model for asymmetric electrolytes. 2. Thermodynamic properties and the pair correlation function. *J. Phys. Chem.*, **81**(13):1311–1316, 1977.

[10] M. Born. Volumen und hydratationswärme der ionen. *Z. Phys.*, **1**(1):45–48, 1920.

[11] W. Fürst and H. Renon. Representation of excess properties of electrolyte solutions using a new equation of state. *AICHE J.*, **39**(2):335–343, 1993. doi:10.1002/aic.690390213.

[12] A. Galindo, A. Gil-Villegas, G. Jackson, and A. N. Burgess. SAFT-VRE: Phase Behavior of Electrolyte Solutions with the Statistical Associating Fluid Theory for Potentials of Variable Range. *J. Phys. Chem. B*, **103**(46):10272–10281, 1999. doi:10.1021/jp991959f.

[13] B. Behzadi, B. Patel, A. Galindo, and C. Ghotbi. Modeling electrolyte solutions with the SAFT-VR equation using Yukawa potentials and the mean-spherical approximation. *Fluid Phase Equilib.*, **236**(1-2):241–255, 2005. doi:10.1016/j.fluid.2005.07.019.

[14] L. F. Cameretti, G. Sadowski, and J. M. Mollerup. Modeling of Aqueous Electrolyte Solutions with Perturbed-Chain Statistical Association Fluid Theory. *Ind. Eng. Chem. Res.*, **44**(23):8944–8944, 2005. doi:10.1021/ie051055i.

[15] Y. Lin, K. Thomsen, and J.-C. de Hemptinne. Multicomponent equations of state for electrolytes. *AICHE J.*, **53**(4):989–1005, 2007. doi:10.1002/aic.11128.

[16] C. Held, L. F. Cameretti, and G. Sadowski. Modeling aqueous electrolyte solutions. *Fluid Phase Equilib.*, **270**(1-2):87–96, 2008. doi:10.1016/j.fluid.2008.06.010.

[17] R. Inchekel, J.-C. de Hemptinne, and W. Fürst. The simultaneous representation of dielectric constant, volume and activity coefficients using an electrolyte equation of state. *Fluid Phase Equilib.*, **271**(1-2):19–27, 2008. doi:10.1016/j.fluid.2008.06.013.

[18] B.-S. Lee and K.-C. Kim. Modeling of aqueous electrolyte solutions based on perturbed-chain statistical associating fluid theory incorporated with primitive mean spherical approximation. *Korean J. Chem. Eng.*, **26**(6):1733–1747, 2009. doi:10.1007/s11814-009-0286-4.

[19] C. Held, T. Reschke, S. Mohammad, A. Luza, and G. Sadowski. ePC-SAFT revised. *Chem. Eng. Res. Des.*, **92**(12):2884–2897, 2014. doi:10.1016/j.cherd.2014.05.017.

[20] J. M. Schreckenberg, S. Dufal, A. J. Haslam, C. S. Adjiman, G. Jackson, and A. Galindo. Modelling of the thermodynamic and solvation properties of electrolyte solutions with the statistical associating fluid theory for potentials of variable range. *Mol. Phys.*, **112**(17):2339–2364, 2014. doi:10.1080/00268976.2014.910316.

[21] B. Maribo-Mogensen, K. Thomsen, and G. M. Kontogeorgis. An electrolyte CPA equation of state for mixed solvent electrolytes. *AICHE J.*, **61**(9):2933–2950, 2015. doi:10.1002/aic.14829.

[22] D. K. Eriksen, G. Lazarou, A. Galindo, G. Jackson, C. S. Adjiman, and A. J. Haslam. Development of intermolecular potential models for electrolyte solutions using an electrolyte SAFT-VR Mie equation of state. *Mol. Phys.*, **114**(18):2724–2749, 2016. doi:10.1080/00268976.2016.1236221.

[23] M. Bülow, M. Ascani, and C. Held. ePC-SAFT advanced - Part I: Physical meaning of including a concentration-dependent dielectric constant in the born term and in the Debye-Hückel theory. *Fluid Phase Equilib.*, **535**:112967, 2021. doi:10.1016/j.fluid.2021.112967.

[24] M. Ascani and C. Held. Prediction of salting-out in liquid-liquid two-phase systems with ePC-SAFT: Effect of the Born term and of a concentration-dependent dielectric constant. *Zeitschrift für Anorganische und Allgemeine Chemie*, **647**(12):1305–1314, 2021. doi:10.1002/zaac.202100032.

[25] M. A. Selam, I. G. Economou, and M. Castier. A thermodynamic model for strong aqueous electrolytes based on the eSAFT-VR Mie equation of state. *Fluid Phase Equilib.*, **464**:47–63, 2018.

[26] D. K. Eriksen, G. Lazarou, A. Galindo, G. Jackson, C. S. Adjiman, and A. J. Haslam. Development of intermolecular potential models for electrolyte solutions using an electrolyte SAFT-VR Mie equation of state. *Mol. Phys.*, **114**(18):2724–2749, 2016.

[27] S. Herzog, J. Gross, and W. Arlt. Equation of state for aqueous electrolyte systems based on the

semirestricted non-primitive mean spherical approximation. *Fluid Phase Equilib.*, **297**(1):23–33, 2010.

[28] J. Rozmus, J.-C. de Hemptinne, A. Galindo, S. Dufal, and P. Mougin. Modeling of strong electrolytes with ePPC-SAFT up to high temperatures. *Ind. Eng. Chem. Res.*, **52**(29):9979–9994, 2013.

[29] G. Das, S. Hlushak, and C. McCabe. A SAFT-VR+ DE equation of state based approach for the study of mixed dipolar solvent electrolytes. *Fluid Phase Equilib.*, **416**:72–82, 2016.

[30] Y. Lin, K. Thomsen, and J.-c. de Hemptinne. Multicomponent equations of state for electrolytes. *AICHE J.*, **53**(4):989–1005, 2007.

[31] A. Schlaikjer, K. Thomsen, and G. M. Kontogeorgis. eCPA: An ion-specific approach to parametrization. *Fluid Phase Equilib.*, **470**:176–187, 2018.

[32] J.-P. Simonin. Further reflections about the "Born" term used in thermodynamic models for electrolytes. *J. Mol. Liq.*, **380**:121713, 2023.

[33] J.-P. Simonin. On the "Born" term used in thermodynamic models for electrolytes. *J. Chem. Phys.*, **150**(24), 2019.

[34] G. M. Silva, X. Liang, and G. M. Kontogeorgis. How to account for the concentration dependency of relative permittivity in the Debye–Hückel and Born equations. *Fluid Phase Equilib.*, **566**:113671, 2023.

[35] I. Y. Shilov and A. K. Lyashchenko. The role of concentration dependent static permittivity of electrolyte solutions in the Debye–Huckel theory. *J. Phys. Chem. B*, **119**(31):10087–10095, 2015.

[36] G. M. Kontogeorgis, B. Maribo-Mogensen, and K. Thomsen. The Debye-Hückel theory and its importance in modeling electrolyte solutions. *Fluid Phase Equilib.*, **462**:130–152, 2018.

[37] J.-K. Hyun and T. Ichiye. Understanding the Born radius via computer simulations and theory. *J. Phys. Chem. B*, **101**(18):3596–3604, 1997.

[38] M. Bucher and T. L. Porter. Analysis of the Born model for hydration of ions. *J. Phys. Chem.*, **90**(15):3406–3411, 1986.

[39] S. Kournopoulos, M. S. Santos, S. Ravipati, A. J. Haslam, G. Jackson, I. G. Economou, and A. Galindo. The contribution of the ion–ion and ion–solvent interactions in a molecular thermodynamic treatment of electrolyte solutions. *J. Phys. Chem. B*, **126**(47):9821–9839, 2022.

[40] F. Drunsel, W. Zmpitas, and J. Gross. A new perturbation theory for electrolyte solutions. *J. Chem. Phys.*, **141**(5), 2014.

[41] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*.

Elsevier, 2023.

[42] M. Theiss and J. Gross. Nonprimitive model electrolyte solutions: Comprehensive data from Monte Carlo simulations. *J. Chem. Eng. Data*, **65**(2):634–639, 2020.

[43] E. Waisman and J. L. Lebowitz. Mean spherical model integral equation for charged hard spheres. II. Results. *J. Chem. Phys.*, **56**(6):3093–3099, 1972.

[44] E. Waisman and J. L. Lebowitz. Mean spherical model integral equation for charged hard spheres I. Method of solution. *J. Chem. Phys.*, **56**(6):3086–3093, 1972.

[45] L. L. Lee. *Molecular Thermodynamics of Nonideal Fluids*. Butterworth-Heinemann, 2016.

[46] G. M. Silva, B. Maribo-Mogensen, X. Liang, and G. M. Kontogeorgis. Improving the Born equation: Origin of the Born radius and introducing dielectric saturation effects. *Fluid Phase Equilib.*, **576**:113955, 2024. doi:https://doi.org/10.1016/j.fluid.2023.113955.

[47] A. A. Rashin and B. Honig. Reevaluation of the Born model of ion hydration. *J. Phys. Chem.*, **89**(26):5588–5593, 1985.

[48] G. M. Silva, X. Liang, and G. M. Kontogeorgis. The true Hückel equation for electrolyte solutions and its relation with the Born term. *J. Mol. Liq.*, **368**:120554, 2022.

[49] B. Maribo-Mogensen, G. M. Kontogeorgis, and K. Thomsen. Comparison of the Debye–Hückel and the Mean Spherical Approximation Theories for Electrolyte Solutions. *Ind. Eng. Chem. Res.*, **51**(14):5353–5363, 2012.

[50] M. Bülow, M. Ascani, and C. Held. ePC-SAFT advanced-Part I: Physical meaning of including a concentration-dependent dielectric constant in the born term and in the Debye-Hückel theory. *Fluid Phase Equilib.*, **535**:112967, 2021.

[51] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, **369**(3):253–287, 1921.

[52] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford university press, 2017.

[53] M. Theiss and J. Gross. Dipolar hard spheres: Comprehensive data from Monte Carlo simulations. *J. Chem. Eng. Data*, **64**(2):827–832, 2019.

[54] B. Efron. Computers and the theory of statistics: thinking the unthinkable. *SIAM review*, **21**(4):460–480, 1979.

[55] J. Eggebrecht and P. Ozler. Multipolar electrolyte solution models. I. Computer simulation of the charged and dipolar hard sphere mixture. *J. Chem. Phys.*, **93**(3):2004–2015, 1990.

[56] T. van Westen. Algebraic second virial coefficient of the Mie m- 6 intermolecular potential based on perturbation theory. *J. Chem. Phys.*, **154**(23), 2021.

7 Conclusion and Outlook

A major long-term objective in molecular thermodynamics is the development of a single, analytic equation of state capable of accurately predicting the properties of a broad range of substances and their mixtures, including those with electrostatic interactions. This thesis contributes to this objective by introducing new thermodynamic models and methodological approaches that enhance the predictive capabilities of existing theories. The results demonstrate that combining perturbation theories with data-driven methods provides a promising route toward reliable, accurate, and physically interpretable molecular models.

7.1 Findings and Achievements of This Work

In the first part of this thesis (chapter 3), a new equation of state (EoS) for Mie-ν-6 fluids, referred to as the uv-B_3-theory, is developed as an extension of the uv-theory. By incorporating a new analytic model for the third virial coefficient B_3, the uv-B_3-theory reproduces the third-order virial expansion in the low density limit. This extension significantly improves accuracy, particularly at low temperatures, where the strongly negative values of B_3 are not captured by the original theory. In comparison to state-of-the-art (semi-)empirical models for Lennard-Jones fluids, the uv-B_3-theory shows comparable accuracy across a wide range of thermodynamic properties. In addition, the new EoS is more generally applicable, as it covers varying repulsive exponents, exhibits robust behaviour in the metastable and unstable fluid regions, and provides a more rigorous foundation for extensions to mixtures and non-spherical molecules. This study demonstrates the potential of the uv-theory as a framework for developing equations of state that combine high accuracy with the strengths of physically based models.

In the second part of this thesis (chapter 4), a new analytic model for the second virial coefficient B_2 of Lennard-Jones chains is developed. This model can serve as a basis for extending the uv- and uv-B_3-theories to chain molecules. A comprehensive molecular simulation study provides detailed insights into single-chain properties and the scaling of B_2 with chain length and temperature. The results show that, for long chains, the temperature for single-chain collapse, the Boyle temperature, and the upper critical solution temperature

for polymer–solvent phase separation converge. Three scaling regimes of B_2 are identified, and both the scaling behaviour and the magnitude of B_2 within these regimes are essentially unaffected by single-chain collapse. This suggests that the residual Helmholtz energy of dilute polymer–solvent systems is insensitive to changes in chain conformation, implying that thermodynamic models for polymer-solvent phase equilibria can be developed based on molecular models that do not exhibit single-chain collapse. The analytic B_2-model is developed based on a third-order perturbation theory using Barker–Henderson chains as the reference fluid. Additional empirical terms are introduced to accurately capture the correct Boyle temperatures and the low-temperature scaling of B_2. Future development of equations of state for chainlike molecules can benefit from the proposed model and the insights obtained in this study.

In chapter 5, the modelling of mixtures using perturbation theories is examined. An algebraic correction factor is introduced to improve the accuracy of perturbation theories that employ a one-fluid approximation. The functional form of the correction factor is identified through symbolic regression using reference data from binary square-well mixtures only. The resulting expression improves equation-of-state predictions not only for binary square-well mixtures but also for multicomponent Lennard-Jones mixtures within both, the PC-SAFT equation of state and the uv-theory. This study demonstrates the potential of combining physically based models with machine learning methods, showing that such approaches can lead to physically interpretable and transferable models. The proposed methodology offers a systematic route to develop further corrections for perturbation theories employing a different one-fluid approximation or describing other interaction potentials.

Chapter 6 focuses on the modelling of electrolytes with perturbation theories. Standard approaches typically replace the ion-related Helmholtz energy contributions with expressions based on the primitive model, including the Debye–Hückel theory, the mean-spherical approximation (MSA), and the Born theory of solvation. In 'primitive models', the solvent is merely considered as a dielectric continuum. In this work, the Helmholtz energy contributions associated with ion–ion and ion–dipole interactions are computed using molecular simulations. Component-specific energy contributions are assessed by applying a two-step thermodynamic integration method and decomposing the Ewald summation into individual contributions. Predictions calculated from the most prominent primitive models reproduce qualitative trends of the simulation results, but deviate significantly from the observed values. Quantitative agreement can be achieved by adjusting the Born size parameter. The adjusted size parameters, however, are much larger than the ion sizes used in the molecular simulations. This indicates that the primitive-model approach is not suitable for predicting the ion-related Helmholtz energy contributions considered. This finding highlights the need for physically

motivated, non-primitive models. The simulation data generated in this work provides a valuable benchmark for their development and validation.

7.2 Advancing Perturbation Theories towards Complex Fluids and Data-Integrated Modelling

A natural direction for further research is to extend the physical basis of the uv- and uv-B_3-theories to more complex systems. Such extensions could include non-spherical molecules, polar interactions, ionic interactions, association and quantum effects. Building on the proposed B_2 model and repulsive chains as reference fluid, the equations of state could be extended to Lennard-Jones chain molecules. Association and polar interactions could be incorporated by adding the corresponding Helmholtz energy contributions, using either analytic expressions from existing models such as the PCP-SAFT equation of state or newly derived models. This thesis demonstrates the need for new, non-primitive Helmholtz energy models to accurately capture ionic interactions. Such models can be validated against the molecular simulation benchmarks generated in this thesis. The presented simulation procedure could be used to generate additional data to support the development of data-driven models. Quantum corrections have already been integrated into the uv-theory by van Westen *et al.* [1]. The same approach could be applied to the uv-B_3-theory, which is more accurate at low temperatures where quantum effects become significant.

Once robust models for these systems are established, the parameters of the pair potentials can be adjusted to experimental data for real substances. The models could also be applied within the entropy scaling framework to predict dynamic fluid properties. Initial results for Mie fluids have been reported by Saric *et al.* [2], who correlated uv-B_3-theory entropy predictions with molecular simulation data for viscosities, diffusion coefficients, and thermal conductivities, demonstrating the potential for further investigation. Beyond bulk properties, the molecular model of the uv-theory could also be extended to inhomogeneous fluids using classical density functional theory, which requires reformulating the Helmholtz energy contributions as functionals of the density profiles.

Another avenue for further advancing the uv-theory framework is its combination with data-driven approaches. Such a hybrid strategy could merge the physical rigour and broad applicability of perturbation theories with the accuracy of empirical models. A natural entry point for data-based refinements is the empirical interpolation function, the u-fraction. In the original uv-theory, the u-fraction is represented by a simple ansatz function with a few adjustable parameters fitted to reference data of the model fluid. For applications to real

substances, the ansatz function could be tailored to each component. Possible adaptations include reparametrization, substitution with a more suitable functional form, or the addition of empirical correction terms. For components well described by the original theory or with limited reference data, the original u-fraction could be retained. For well-measured substances, advanced machine-learning techniques, such as symbolic regression, could be employed to generate highly accurate models. This approach is particularly promising for substances that are difficult to describe with physically based equations of state but for which extensive experimental data are available, such as water or carbon dioxide.

References

[1] T. van Westen, G. Bauer, and J. Gross. Corresponding-states framework for classical and quantum fluids—Beyond Feynman–Hibbs. *J. Chem. Phys.*, **162**(3), 2025.

[2] D. Saric, I. H. Bell, G. Guevara-Carrion, and J. Vrabec. Influence of repulsion on entropy scaling and density scaling of monatomic fluids. *J. Chem. Phys.*, **160**(10), 2024.

Appendices

A Supporting Information: Physically Based Equation of State for Mie ν-6 Fluids

The content of this chapter is a literal quote of the appendix of publication:

A. Reimer, T. van Westen and J. Gross. Physically based equation of state for Mie ν-6 fluids. *J. Chem. Phys.*, **158**, 164506 (2023), doi: 10.1063/5.0141856

A.1 Contributions to the *uv*-B_3-Theory

A.1.1 Reference Helmholtz Energy a_0

We describe the Helmholtz energy of the WCA reference fluid a_0 based on a first-order Mayer-f expansion[1–3] about a hard-sphere fluid of effective diameter d, according to

$$a_0 = a_d^{\mathrm{hs}} + \Delta a_0 \tag{A.1}$$

where a_d^{hs} the hard sphere contribution and Δa_0 denotes the Mayer-f-perturbation term. The hard-sphere contribution is calculated with the equation of state of Carnahan and Starling[4],

$$a_d^{\mathrm{hs}} = \frac{4\eta - 3\eta^2}{(1-\eta)^2} \tag{A.2}$$

with the packing fraction $\eta = \frac{\pi}{6}\rho^* d^{*3}$. We use the hard-sphere diameter proposed by van Westen and Gross[5],

$$d^*(T^*, \nu) = \frac{d}{\sigma} = r_{\mathrm{m}}^* \left(\frac{1}{1 + c\sqrt{T^*}} \right)^{2/\nu} \tag{A.3}$$

where $r_{\mathrm{m}}^* = r_{\mathrm{m}}/\sigma$ and

$$c = \left(\frac{\nu}{6}\right)^{-\frac{\nu}{12-2\nu}} - 1 \tag{A.4}$$

is an empirical constant chosen such that eq. (A.3) reproduces $d^*(T^*=1,\nu)=1$, a result predicted by a Boltzmann factor criterion[6]. The Mayer-f-perturbation term is defined as

$$\Delta a_0 = -2\pi\rho^* \int_0^\infty y_d^{\mathrm{hs}}(r^*\sigma)\left(e_0(r^*)-e_d^{\mathrm{hs}}(r^*)\right) r^{*2}\mathrm{d}r^* \tag{A.5}$$

with $r^*=r/\sigma$, y_d^{hs} being the cavity-correlation function of the hard-sphere fluid with effective diameter d, and the Boltzmann factors $e_0(r^*)=\exp(-\beta u_0(r^*))$ and $e_d^{\mathrm{hs}}(r^*)=\exp(-\beta u_d^{\mathrm{hs}}(r^*))$. An algebraic expression for Δa_0, based on the Mean Value Theorem, was proposed by van Westen and Gross[5]:

$$\Delta a_0 = -\frac{2}{3}\pi\rho^*\left[\frac{1-\eta_A/2}{(1-\eta_A)^3}\left(r_{\mathrm{m}}^{*3}-q^{*3}\right)-\frac{1-\eta_B/2}{(1-\eta_B)^3}\left(r_{\mathrm{m}}^{*3}-d^{*3}\right)\right] \tag{A.6}$$

with effective packing fractions η_A and η_B, and the temperature-dependent dimensionless diameter

$$q^*(T^*) = r_{\mathrm{m}}^*\left[1+\sqrt{\frac{2\pi\nu}{6}}T^{*\frac{1}{2}}+C_1T^*+C_2T^{*\frac{3}{2}}+C_3T^{*2}\right]^{-\frac{1}{2\nu}} \tag{A.7}$$

with the empirical coefficients

$$\begin{pmatrix}C_1\\C_2\\C_3\end{pmatrix} = \begin{pmatrix}c_1 & c_2 & & \\ c_3 & c_4 & c_5 & c_6\\ & c_7 & c_8 & c_9\end{pmatrix}\begin{pmatrix}1\\(\nu-7)\\(\nu-7)^2\\(\nu-7)^3\end{pmatrix} \tag{A.8}$$

The coefficients c_1-c_9 are listed in table A.1. The diameter of eq. (A.7) is a fit to the hard-sphere diameter that exactly maps the second virial coefficient of the reference fluid onto that of hard spheres, $B_{20}=B_2^{\mathrm{hs}}$. The effective packing fractions are calculated based on the polynomials

$$\eta_A(\eta,\tau) = \eta + A_1(\tau)\eta + A_2(\tau)\eta^2 + A_3(\tau)\eta^3 + A_4(\tau)\eta^4 \tag{A.9}$$

and

$$\eta_B(\eta,\tau) = \eta + B_1(\tau)\eta + B_2(\tau)\eta^2 + B_3(\tau)\eta^3 \tag{A.10}$$

Table A.1: Model constants c_i from van Westen and Gross[5] for the diameter q, defined in eq. (A.7).

i	c_i
1	1.92840364363978
2	0.44316589626508
3	0.52012081614176
4	0.18252675923441
5	0.0110319989659929
6	-7.97813995328348E-05
7	0.0129885156087242
8	6.41039871789327E-03
9	1.85866741090323E-05

with the dimensionless length scale

$$\tau(T^*, \nu) = r_m^*(\nu) - d^*(T^*, \nu) \tag{A.11}$$

The parameters in eqs. (A.9) and (A.10) are defined as

$$A_i = \left(a_{i1} + \frac{a_{i2}}{\nu}\right)\tau + \left(a_{i3} + \frac{a_{i4}}{\nu}\right)\tau^2 \tag{A.12}$$

and

$$B_i = b_{i1}\tau + b_{i2}\tau^2 \tag{A.13}$$

We evaluated the integral in eq. (A.5) numerically for $10 \leq \nu \leq 20$, $0.3 \leq T^* \leq 20$ and packing fractions $0 \leq \eta = \frac{\pi}{6}\rho^* d^{*3} \leq 0.52$ using the algebraic approximation of the cavity correlation function of hard spheres proposed by de Souza and Ben-Amotz[7] and Ben-Amotz and Herschbach[8],

$$y_d^{\mathrm{hs}}(r^*) \approx \exp\left(A + Br^* + Cr^{*3}\right) \tag{A.14}$$

Table A.2: Model constants a_{ij} for the or the effective packing fractions η_A and η_B, defined in eqs. (A.9)–(A.13)

i	a_{i1}	a_{i2}	a_{i3}	a_{i4}
1	2.64043218	-1.2184421	-22.90786387	0.96433414
2	-16.75643936	30.83929771	73.08711814	-166.57701616
3	19.53170162	-88.87955657	-76.51387192	443.68942745
4	-3.77740877	83.04694547	21.62502721	-304.8643176
i	b_{i1}		b_{i2}	
1	2.19821588		-20.45005484	
2	-13.47050687		56.65701375	
3	12.90119266		-42.71680606	

with

$$A = \frac{3-\eta}{(1-\eta)^3} - 3$$
$$B = \frac{-3\eta(2-\eta)}{(1-\eta)^3}$$
$$C = \ln\left(\frac{2-\eta}{2(1-\eta)^3}\right) - \frac{\eta(2-6\eta+3\eta^2)}{(1-\eta)^3}$$

The 22 model constants a_{ij} and b_{ij} of eqs. (A.12) and (A.13) were fitted simultaneously to the results from numerical integration. The 8225 reference data points from numerical integration are described with a mean absolute relative deviation (MARD) of 0.20%. The correlated constants are provided in table A.2.

A.1.2 First-Order Perturbation Contribution Δa_1^u

The first-order perturbation contribution, eq. (3.8), can be written as

$$\Delta a_1^u = 2\pi\rho^*\beta^* I^u(\rho^*, T^*, \nu) \tag{A.15}$$

with $\beta = 1/k_B T$, and the dimensionless correlation integral

$$I^u(\rho^*, T^*, \nu) = \int_0^\infty g_0(r^*\sigma) w^*(r^*) r^{*2} \mathrm{d}r^* \tag{A.16}$$

Table A.3: Model constants c_{ij} from van Westen and Gross[5] for the correlation integral I^u as defined in eqs. (A.17) and (A.19).

i	c_{i1}	c_{i2}	c_{i3}	c_{i4}	c_{i5}	c_{i6}
1	-0.2622378162	0.6585817423	5.5318022309	0.6902354794	-3.6825190645	-1.7263213318
2	-0.1899241690	-0.5555205158	9.1361398949	0.7966155658	-6.1413017045	4.9553415149
3	0.1169786415	-0.2216804790	-2.0470861617	-0.3742261343	0.9568416381	10.1401796764
4	0.5852642702	2.0795520346	19.0711829725	-2.3403594600	2.5833371420	432.3858674425
5	-0.6084232211	-7.2376034572	19.0412933614	3.2388986513	75.4442555789	-588.3837110653
6	0.0512327656	6.6667943569	47.1109947616	-0.5011125797	-34.8918383146	189.5498636006

where we introduced the pair-correlation function of the WCA reference fluid $g_0(r)$ and the dimensionless potential $w^*(r^*) = w(r)/\varepsilon$. We use the same algebraic expression for Δa_1^u as proposed in the original *uv*-theory[5],

$$I^u(\rho^*, T^*, \nu) = \frac{q^{*3} - r_{\mathrm{m}}^{*3}}{3} - \alpha(r_{\mathrm{m}}^*) + C\frac{C_1\rho^* + C_2\rho^{*2} + C_3\rho^{*3}}{1 + C_4\rho^* + C_5\rho^{*2} + C_6\rho^{*3}} \tag{A.17}$$

with the mean field constant

$$\alpha(r_{\mathrm{m}}^*) = -\int_{r_{\mathrm{m}}^*}^{\infty} u^*(r^*)r^{*2}\mathrm{d}r^* = C\left(\frac{r_{\mathrm{m}}^{*-3}}{3} - \frac{r_{\mathrm{m}}^{*3-\nu}}{\nu-3}\right) \tag{A.18}$$

where $u^*(r^*) = u(r^*)/\varepsilon$. The prefactor to the Mie potential C was defined in eq. (3.2). This ansatz function incorporates the exact low-density limit of eq. (A.16) and an empirically modelled density dependence. The coefficients are defined as

$$C_i(T^*, \nu) = \left(c_{i1} + \frac{c_{i2}}{\nu} + \frac{c_{i3}}{\nu^2}\right) + \left(c_{i4} + \frac{c_{i5}}{\nu} + \frac{c_{i6}}{\nu^2}\right)\tau(T^*, \nu) \tag{A.19}$$

with τ from eq. (A.11). The model constants c_{ij} were optimized by correlating Monte-Carlo (MC) simulation data for I^u of van Westen and Gross[5]. The respective values of c_{ij} are provided in table A.3.

A.1.3 Second Virial Coefficient B_2

Elliott *et al.*[9] proposed an inverse temperature series with empirical corrections for the second virial coefficient B_2 of a LJ fluid. This semi-empirical approach was recently revised and extended to Mie ν-6 fluids by van Westen[10], referred to as Revised-series-approximation (RSAP). We use the RSAP based on the WCA division of the Mie-potential to model $\Delta B_2 =$

$B_2 - B_{20}$ in the *uv*-B_3-theory. The perturbation contribution to the second virial coefficient from the RSAP is given by

$$\Delta B_2^* = \frac{\Delta B_2}{\sigma^3} = \left(B_{20}^* - \frac{2}{3}\pi r_{\mathrm{m}}^{*3} - c_1 \right) Y - c_2 \sum_{i=1}^{15} \frac{\beta^{*i}}{i!i} - c_3\beta^* - c_4\beta^{*2} \tag{A.20}$$

with $Y = \exp(\beta^*) - 1$ and the model constants

$$c_i = c_{i0} + \frac{c_{i1}}{\nu} + \frac{c_{i2}}{\nu^2} + \frac{c_{i3}}{\nu^3} \tag{A.21}$$

The coefficients c_{ij} were fitted to results for B_2 obtained by numerical integration. The correlated coefficients are given in table A.4.

Table A.4: Model constants c_{ij} from van Westen [10] for the second virial coefficient as defined in eqs. (A.20) and (A.21).

	$j=0$	$j=1$	$j=2$	$j=3$
c_{1j}	-0.063550989	6.206829830	-37.45829549	40.72849774
c_{2j}	1.519053409	13.14989643	85.35058674	374.1906360
c_{3j}	0.693456220	9.459946180	-53.28984218	315.8199084
c_{4j}	0.007492596	0.546171170	7.979562575	-119.6126395

The second virial coefficients of the WCA reference fluid follows from the expression of the reference fluid's Helmholtz energy as

$$B_{20}^* = \lim_{\rho^* \to 0} \left(\frac{\partial a_0}{\partial \rho^*} \right)_{N,T^*} = \frac{2}{3}\pi q^{*3} \tag{A.22}$$

A.1.4 Third Virial Coefficient B_3

Our model for the third virial coefficient B_3 of the Mie ν-6 fluids is presented in section 3.3.4. In figure A.1 we show the first and second order derivatives of B_3^* with respect to the inverse temperature β^* for different repulsive exponents. Our model (solid lines) is compared to results from numerical integration via Fast Fourier transforms (FFT) [11] (circles). Both derivatives are described rather well by the model.
The *uv*-B_3-theory requires the difference $\Delta B_3 = B_3 - \tilde{B}_{30}$ as input, where $\tilde{B}_{30}$ defines the third

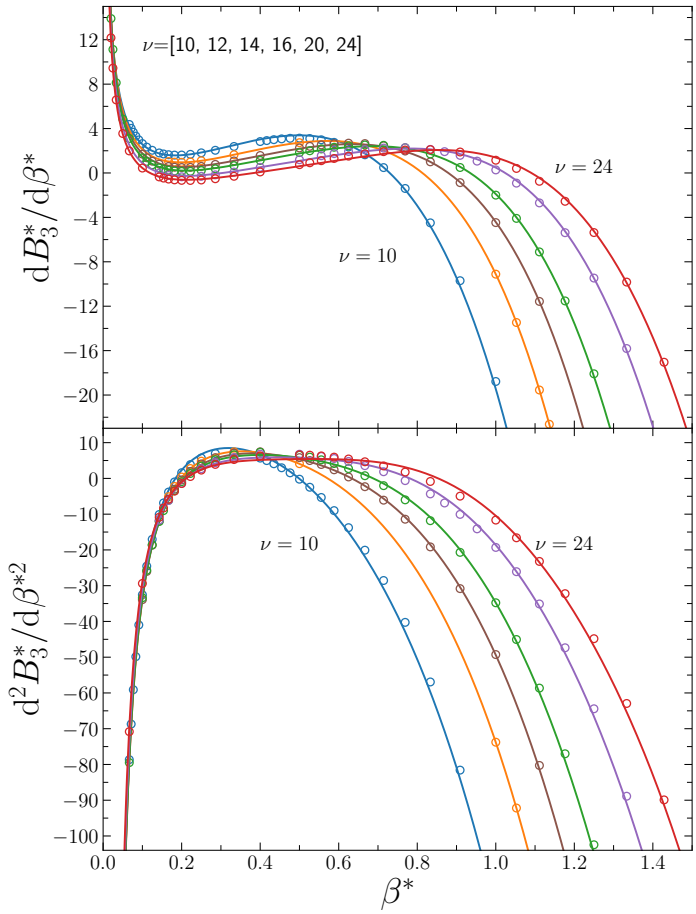


Figure A.1: First and second derivative of the third virial coefficient of Mie ν-6 fluids with respect to the inverse temperature β^*. Comparison of the model of eqs. (3.19)–(3.24) (lines) with results obtained by numerical integration of B_3 using FFT[11] (circles).

virial coefficient of the WCA reference fluid predicted by our description of a_0,

$$\begin{aligned}\tilde{B}_{30}^* &= \frac{\tilde{B}_{30}}{\sigma^6} = \lim_{\rho^*\to 0}\left(\frac{\partial^2 a_0}{\partial \rho^{*2}}\right)_{N,T^*} \\ &= 10\left(\frac{\pi}{6}d^{*3}\right)^2 - \frac{5}{9}\pi^2 d^{*3}\left[(1+A_1)(r_\mathrm{m}^{*3}-q^{*3}) - (1+B_1)(r_\mathrm{m}^{*3}-d^{*3})\right]\end{aligned} \tag{A.23}$$

with the parameters A_1 and B_1 from eqs. (A.12) and (A.13). We use a tilde for $\tilde{B}_{30}$ because the description of B_{30} based on a_0 of eqs. (A.1) and (A.6) is approximate. The tilde is not used for B_{20}, because our description of a_0 by eq. (A.6) gives an exact description of eq. (A.5) to first-order in density (within the uncertainty of our description of q^* by the fit of eq. (A.7)[5]).

A.1.5 First-Order Perturbation Contribution to the Virial Coefficients, ΔB_{21}^u and ΔB_{31}^u

The contribution of the first-order perturbation term Δa_1^u to the second and third virial coefficients are obtained from eq. (A.15), as

$$\begin{aligned}\Delta B_{21}^{u*} = \frac{\Delta B_{21}^u}{\sigma^3} &= \lim_{\rho^*\to 0}\left(\frac{\partial \Delta a_1^u}{\partial \rho^*}\right)_{N,T^*} = 2\pi\beta^* \lim_{\rho\to 0} I^u(\rho^*, T^*, \nu) \\ &= 2\pi\beta^*\left(\frac{q^{*3} - r_{\mathrm{m}}^{*3}}{3} - \alpha(r_{\mathrm{m}}^*)\right)\end{aligned} \tag{A.24}$$

and

$$\Delta B_{31}^{u*} = \frac{\Delta B_{31}^u}{\sigma^6} = \lim_{\rho^*\to 0}\left(\frac{\partial^2 \Delta a_1^u}{\partial \rho^{*2}}\right)_{N,T^*} = 4\pi\beta^* \mathcal{C} C_1 \tag{A.25}$$

with the model constant C_1 from eq. (A.19).

References

[1] H. C. Andersen, J. D. Weeks, and D. Chandler. Relationship between the Hard-Sphere Fluid and Fluids with Realistic Repulsive Forces. *Phys. Rev. A*, **4**(4):1597–1607, 1971. doi:10.1103/physreva.4.1597.

[2] K. Gubbins, W. Smith, M. Tham, and E. Tiepel. Perturbation theory for the radial distribution function. *Mol. Phys.*, **22**(6):1089–1105, 1971.

[3] J. A. Barker. Cluster integrals and statistical mechanics of solutions. *Proc. R. Soc. A*, **241**(1227):547–553, 1957.

[4] N. F. Carnahan and K. E. Starling. Equation of State for Nonattracting Rigid Spheres. *J. Chem. Phys.*, **51**(2):635–636, 1969. doi:10.1063/1.1672048.

[5] T. van Westen and J. Gross. Accurate thermodynamics of simple fluids and chain fluids based on first-order perturbation theory and second virial coefficients: *uv*-theory. *J. Chem. Phys.*, **155**(24):244501, 2021. doi:10.1063/5.0073572.

[6] D. Ben-Amotz and G. Stell. Hard sphere perturbation theory for fluids with soft-repulsive-core potentials. *J. Chem. Phys.*, **120**(10):4844–4851, 2004. doi:10.1063/1.1647520.

[7] L. E. de Souza and D. Ben-Amotz. Optimized perturbed hard sphere expressions for the structure and thermodynamics of Lennard-Jones fluids. *Mol. Phys.*, **78**(1):137–149, 1993. doi:10.1080/00268979300100131.

[8] D. Ben-Amotz and D. R. Herschbach. Hard fluid model for solvent-induced shifts in molecular vibrational frequencies. *J. Phys. Chem.*, **97**(10):2295–2306, 1993.

[9] J. R. Elliott, A. J. Schultz, and D. A. Kofke. Combined temperature and density series for fluid-phase properties. II. Lennard-Jones spheres. *J. Chem. Phys.*, **151**(20):204501, 2019. doi:10.1063/1.5126281.

[10] T. van Westen. Algebraic second virial coefficient of the Mie *m*-6 intermolecular potential based on perturbation theory. *J. Chem. Phys.*, **154**(23):234502, 2021. doi:10.1063/5.0050659.

[11] A. J. Schultz. Reduced Second Virial Coefficient from the Lennard–Jones Potential. <https://www.etomica.org/apps/virial/b3>, 2005. doi:10.1017/cbo9780511815928.046.

B Supporting Information: Effect of Chain Conformation on the Free Energy of Dilute Polymer Solutions: Monte Carlo Simulations and Perturbation Theory for the Second Virial Coefficient of Lennard-Jones Chains

The content of this chapter is a literal quote of the Supporting Information of the publication:

A. Reimer, J. Gross and T. van Westen. Effect of Chain Conformation on the Free Energy of Dilute Polymer Solutions: Monte Carlo Simulations and Perturbation Theory for the Second Virial Coefficient of Lennard–Jones Chains. *Macromolecules*, **58**, 3, 1498–1511 (2025), doi:10.1021/acs.macromol.4c02827

B.1 Perturbation Series for the Second Virial Coefficient of Two Chain Molecules

In the following, $\langle \ldots \rangle_\lambda$ is used as a short-hand notation to denote an average over the conformations of single, isolated chains,

$$\langle X(\mathbf{12}) \rangle_\lambda \tag{B.1}$$

The the total perturbation contribution to the second virial coefficient of Eq. (4.10) can be written as the following integral over the coupling parameter λ

$$\Delta B_2 = \frac{\beta}{2} \int_0^1 \int \left\langle \exp\left[-\beta U_\lambda^{\text{inter}}(12)\right] \Phi_\lambda(12) \right\rangle_\lambda \mathrm{dR} \mathrm{d}\lambda \tag{B.2}$$

with

$$\beta \Phi_\lambda(12) = \beta W^{\text{inter}}(12) + \beta \left(W^{\text{intra}}(12) - \left\langle W^{\text{intra}}(12) \right\rangle_\lambda \right) \tag{B.3}$$

Taylor expanding the integrand of Eq. (B.2) about 0 leads to the following perturbation series

$$\Delta B_2 = \sum_i^{\infty} \Delta B_{2i} \tag{B.4}$$

with

$$\Delta B_{21} = \frac{\beta}{2} \int \left\langle \exp\left[-\beta U_0^{\text{inter}}(12)\right] \Phi_0(12) \right\rangle_0 \mathrm{dR} \tag{B.5}$$

$$\Delta B_{22} = \frac{\beta}{4} \int \left\langle \exp\left[-\beta U_0^{\text{inter}}(12)\right] \left(\Phi_0(12)^2 + \Phi_0(12)'\right) \right\rangle_0 \mathrm{dR} \tag{B.6}$$

$$\Delta B_{23} = \frac{\beta}{12} \int \left\langle \exp\left[-\beta U_0^{\text{inter}}(12)\right] \left(\Phi_0(12)^3 + 3\Phi_0(12)\Phi_0(12)' + \Phi_0(12)''\right) \right\rangle_0 \mathrm{dR} \tag{B.7}$$

The derivatives of Φ_λ to the coupling parameter λ (denoted by the primes) essentially describe the first few terms of the Taylor series of the intramolecular distribution function of a single isolated chain $f_\lambda(\mathbf{\Omega}_i)$ in λ. They can be derived as

$$\Phi_0(12)' = \left(\frac{\partial \Phi_\lambda(12)}{\partial \lambda}\right)_0 = \beta^2 \left\langle \left(W^{\text{intra}}(12) - \left\langle W^{\text{intra}}(12)\right\rangle_0\right)^2 \right\rangle_0 \tag{B.8}$$

and

$$\Phi_0(12)'' = \left(\frac{\partial^2 \Phi_\lambda(12)}{\partial \lambda^2}\right)_0 = -\beta^3 \left\langle \left(W^{\text{intra}}(12) - \left\langle W^{\text{intra}}(12)\right\rangle_0\right)^3 \right\rangle_0 \tag{B.9}$$

The approximate perturbation terms $\Delta\tilde{B}_{2i}$ from the main text are derived from the above by assuming the conformations of single isolated chains do not depend on λ; then all derivatives Φ_0', Φ_0'', and so on, equal zero.

B.2 Simulation Details

B.2.1 Rosenbluth Method

For chains of 16 segments or less, chain conformations were generated using the method of Rosenbluth and Rosenbluth.[1–4] In this method, a chain is grown one segment at a time, with the direction of each new segment being selected from a pre-defined number of k trial directions. In this work we used $k = 6$ throughout. The trial directions are generated with a probability corresponding to the Boltzmann factor of the intra-molecular energy of the new bond. A single trial direction is then selected from the set of k trial directions with a probability corresponding to the Boltzmann factor of the non-bonded intra-molecular potential energy of the new segment (according to Eq. (4.1), with a cut-off $r_c = 3.5\sigma$ and standard long-range tail corrections).

The Rosenbluth method is known to introduce a bias. To correct for this, and to recover correct Boltzmann statistics, each generated chain of index i is given a statistical weight equal to the Rosenbluth weight R_i, defined as the sum over the Boltzmann weights of all $k^{(m-1)}$ chains that could have been grown for that chain. The average of a single chain property X over N generated chains is thus calculated as

$$\langle X \rangle = N^{-1} \sum_{i}^{N} X_i P_i \tag{B.10}$$

with

$$P_i = R_i / \sum_{i}^{N} R_i \tag{B.11}$$

For calculating the average of a property Y that depends on the conformation of two chains (e.g. the second virial coefficient) one simply substitutes the number of independent chain pairs N_{pair} for N and the Rosenbluth weight of the pair $R_{1,i}R_{2,i}$ for R_i. Although the Rosenbluth method is formally correct, recovering Boltzmann statistics of long chain molecules is difficult due to poor overlap of the Rosenbluth and Boltzmann distributions.[5] For this reason we employ the Rosenbluth method only for generating conformations of chains of up to 16 segments.

Single-chain properties were averaged over at least $2 \cdot 10^5$ independent chain conformations. To compute the second virial coefficient, the average of the Mayer-f function $\langle f_M(\mathbf{12})\rangle_{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2}$ of Eq. (4.5) was sampled using the full Mie potential (i.e. no cut-off) over a grid in the intermolecular distance $r^* = r/\sigma = [0, r^*_{\max}]$, along 10 random directions $\boldsymbol{r}/r$, for 10^5 independent chain pairs and 40 random relative orientations of the chains. The orientations were generated by rotating chain 2 (specifically, it's centre of mass and the relative positions of the segments with respect to the center of mass) with respect to a randomly chosen (x, y, or z) axis of the Cartesian reference system. The maximum grid-distance was defined as $r^*_{\max} = m_1 + m_2 + r^*_{\text{cut}}$, where r^*_{cut} is the distance at which $f_{\text{M}}(r)r^{*2}$ of a Mie monomer equals 0.0001. We used a grid spacing $\Delta r^* = 0.05$.

All computed results were averaged over at least five independent calculations, and the standard error was calculated as the root-mean-square deviation of the independent calculations with respect to the average value. For the second virial coefficient, 95% confidence intervals were estimated by assuming the standard error follows a student-t distribution (with five degrees of freedom).

B.2.2 Mayer Sampling Monte Carlo Calculations for B_2 of Long Chain Molecules

Due to the aforementioned limitations of the Rosenbluth method for long chain molecules, and for improved efficiency, we computed the second virial coefficient of chains with more than 16 segments using Mayer sampling Monte Carlo (MSMC)[6] with overlap sampling[7,8] using the virial extension[9] of the ZENO[10] software tool.

MSMC is an efficient method for calculating cluster integrals for virial coefficients. MSMC employs free energy perturbation methods, where the target integral is computed by evaluating the ratio of the integral to that of a known reference system based on simulation averages.

We performed ZENO simulations with 10^8 Monte Carlo production steps (increased to 10^9 for certain long-chain, low-temperature combinations; see Supplementary Material [of the online version of the publication] for details) using a virial reference diameter of 4. An example input file is provided in the Supplementary Material [of the published article]. ZENO's built-in uncertainty calculation was used to obtain the standard errors of the results. For further details on the calculation methods, we refer to the work of Singh and Kofke[6] and Bansal *et al.*[9].

B.2.3 Canonical Monte Carlo Simulations

To sample single-chain properties of long Lennard-Jones chains ($m = \{16, 32, 64, 128, 256\}$), we performed NVT-MC simulations of a single chain. In each MC cycle we applied m MC moves chosen from configurational bias regrowth moves,[11,12] single-bead axial (crank-shaft) rotations of both inner ($2 \leq i \leq m-1$) and outer ($i = 1$ and $i = m$) segments,[13,14] reptation moves, where an end segment is randomly attached at the other side of the chain, rigid rotations of part of the chain around a randomly chosen pivot segment, and end-bridging moves, where an outer segment a is connected to an inner segment c by detaching the inner segment's neighbour b and using it as a bridge for the new abc bond, with relative probabilities 0.04, 0.45, 0.04, 0.02, and 0.45. For the configurational bias regrowth move, we chose a random segment as the starting segment and regrew the molecule in a random direction (forward or backward) using six trial orientations for each new segment. The crank-shaft moves of inner segments were combined with a configurational bias scheme, using six trial angles for the rotated segment. For more detailed information on the other MC moves we refer to Taylor.[15]

We typically used 10^5 MC cycles for initialisation and 10^6 MC cycles for production. The Lennard-Jones interactions between segments (Eq. (4.1)) were truncated at $r_c = 3.5\sigma$, and standard long-range tail corrections (assuming the segment-segment pair correlation function equals unity for distances beyond r_c) were applied. Since the correlation-hole effect[16,17] implies that standard tail corrections are more approximate for intramolecular segment-segment interactions as compared to intermolecular segment-segment interactions, we performed some simulations for a LJ 32-mer with a longer cut-off $r_c = 6\sigma$. The sampled intramolecular energies match those for $r_c = 3.5\sigma$ to within one standard deviation.

NVT-MC simulations were also used to compute the perturbation contributions to the second virial coefficient, defined in Eq. (4.36) Therefore we simulated N chains interacting by the reference part of the intermolecular potential (Eq. (4.7)) in a cubic box of volume V at temperature T. Periodic boundary conditions were applied. We typically used at least 10^6 MC cycles for equilibration and 10^6 MC cycles for calculating ensemble averages. Each MC cycle consisted of N MC moves, comprising translation of the centre-of-mass of a chain,[4] rotation,[4] (partial) configurational bias regrowth,[11,12] and crank-shaft moves,[13,14] with relative probabilities 0.3, 0.1, 0.1, 0.5. During equilibration, the step-size for translation and rotation moves was adjusted to obtain approximately 40% acceptance. The volume of the box was chosen large enough for the conformations of the chains to be independent of the other chains in the box and, thus, to be suitable for generating second virial coefficients. Typically, a volume corresponding to a dimensionless segment density $\rho_s^* = m\rho\sigma^3 = 0.001$ was sufficiently large.

This was verified by comparing sampled single-chain properties (which were sampled each MC cycle) to those sampled using the Rosenbluth method of Sec. B.2.1 applied to the same molecular model (see Figs. B.1 and B.4).

During the production part of the simulation, 10 snapshots of the simulation box were stored (each taken after $N_{\mathrm{prod}}/10$ MC cycles); in a post-processing step each snapshot was used to generate the perturbation contributions to the second virial coefficient (see Eqs. (4.36)) for all $N(N-1)/2$ distinct chain pairs that can be generated from that snapshot. The integration of the conformational averages $\langle \ldots \rangle_0$ over the distance vector between the chains centres of mass $\boldsymbol{r}_{12}$ was divided in three parts:[18]

1. one part of 10 equidistant grid-points in r_{12} for $r < R_g/2$,
2. one part for 150 equidistant grid-points in r_{12} for $R_g/2 \leq r_{12} < 2R_g$,
3. and another long-ranged (LR) part of 20 equidistant grid-points in the inverse distance $x = R_g/r_{12}$ for $x < 1/2$. For the first-order perturbation contribution, the integral representation of this long-ranged part can be written

$$\Delta\tilde{B}_{21}^{\mathrm{LR}} = 2\pi\beta R_g^3 \int_0^{\frac{1}{2}} \left\langle \exp\left[-\beta U_{12,0}(12)\right] W_{12}(12) \right\rangle_0 x^{-4} \mathrm{d}x \tag{B.12}$$

The higher-order perturbation contributions are of similar form.

Compared to the integration done in Sec. B.2.1, this procedure places more emphasis on the range of r where the integrands $\langle \ldots \rangle_0 r_{12}^2$ change most rapidly, and minimizes computational effort. We generated the perturbation contributions B_{21}, B_{22}, and B_{23} for chains of 1, 2, 3, 4, 8, 16, and 32 segments. Statistical uncertainties were estimated by dividing the production part of the simulation into five blocks and calculating the standard deviation of the block averages with respect to the average over all five blocks.

B.3 Supplemental Figures

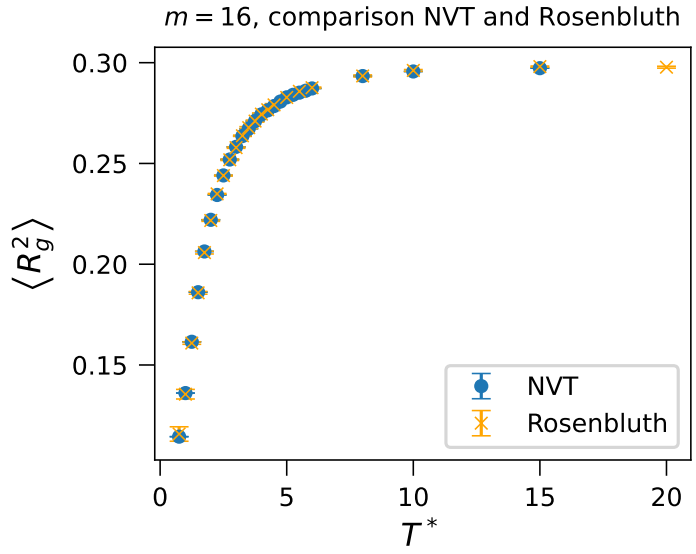


Figure B.1: Molecular simulation results for the radius of gyration of freely-jointed chains comprising 16 Lennard-Jones segments. Comparison of results obtained using NVT-MC simulations of several chains ($N > 1$) in a large box and Rosenbluth MC simulations of a single chain. Error bars denote one standard deviation.

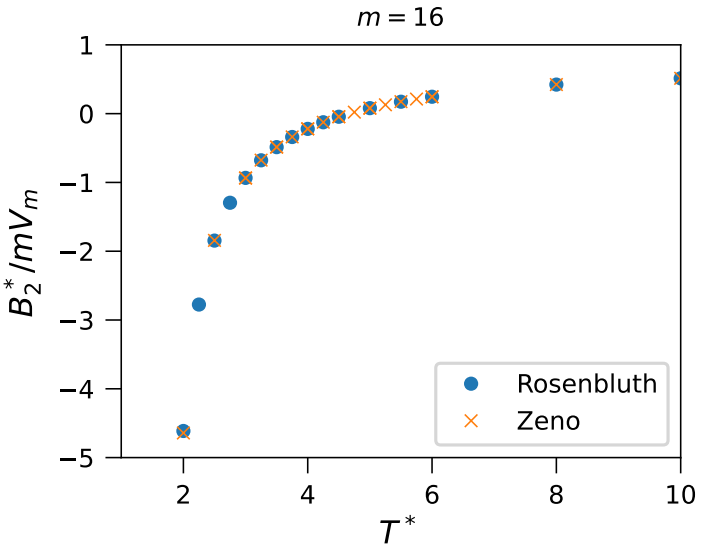


Figure B.2: Molecular simulation results for the second virial coefficient of freely-jointed chains comprising 16 Lennard-Jones segments ($m = 16$). Comparison of results obtained using single-chain conformations generated by Rosenbluth MC of single chains and results obtained from MSMC simulations with the Zeno software package. Estimated uncertainties are smaller than the symbol size.

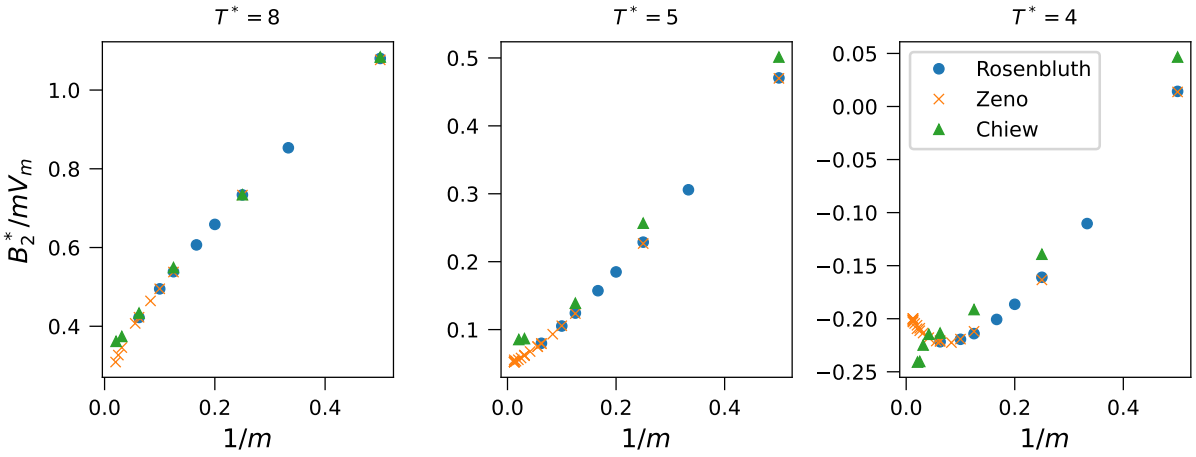


Figure B.3: Molecular simulation results for the second virial coefficient of freely-jointed chains comprising m Lennard-Jones segments at various dimensionless temperatures T^*. Comparison of simulation data generated in this work (Rosenbluth, Zeno) to literature data of Chiew[19]. Estimated uncertainties of the data generated in this work are smaller than the symbol size; for the literature data no uncertainties were reported.

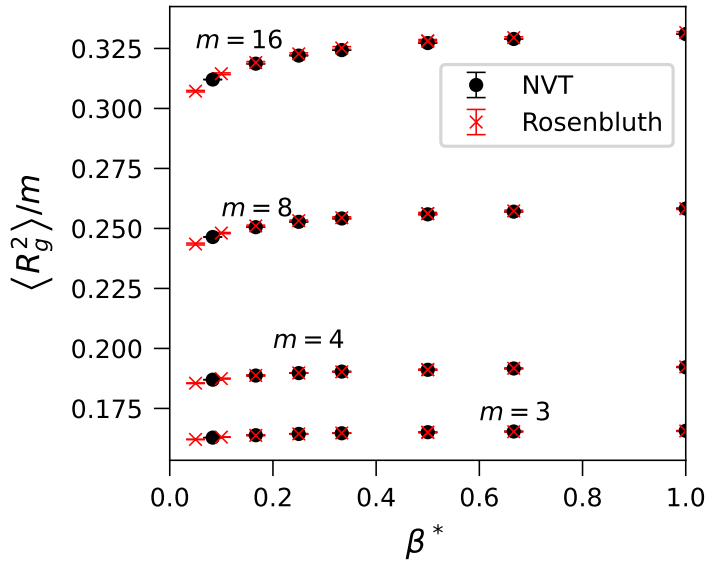


Figure B.4: Molecular simulation results for the radius of gyration of freely-jointed chains comprising m BH segments.

References

[1] M. N. Rosenbluth and A. W. Rosenbluth. Monte-Carlo calculation of the average extension of molecular chains. *J. Chem. Phys.*, **23**(2):356–359, 1955.

[2] D. Frenkel and B. Smit. Unexpected length dependence of the solubility of chain molecules. *Mol. Phys.*, **75**(5):983–988, 1992.

[3] D. Frenkel, G. C. A. M. Mooij, and B. Smit. Novel scheme to study structural and thermal properties of continuously deformable molecules. *J. Phys. Condens. Matter*, **4**(12):3053–3076, 1992.

[4] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Academic Press, San Diego, 2nd edition, 2002.

[5] J. Batoulis and K. Kremer. Statistical properties of biased sampling methods for long polymer chains. *J. Phys. A: Math. Gen.*, **21**(1):127, 1988. doi:10.1088/0305-4470/21/1/020.

[6] J. K. Singh and D. A. Kofke. Mayer Sampling: Calculation of Cluster Integrals using Free-Energy Perturbation Methods. *Phys. Rev. Lett.*, **92**:220601, 2004. doi:10.1103/PhysRevLett.92.220601.

[7] K. M. Benjamin, A. J. Schultz, and D. A. Kofke. Gas-phase molecular clustering of TIP4P and SPC/E water models from higher-order virial coefficients. *Ind. Eng. Chem. Res.*, **45**(16):5566–5573, 2006.

[8] K. M. Benjamin, J. K. Singh, A. J. Schultz, and D. A. Kofke. Higher-order virial coefficients of water models. *J. Phys. Chem. B*, **111**(39):11463–11473, 2007.

[9] A. Bansal, A. J. Schultz, J. F. Douglas, and D. A. Kofke. Probabilistic computations of virial coefficients of polymeric structures described by rigid configurations of spherical particles: A fundamental extension of the ZENO program. *J. Chem. Phys.*, **157**(22):224801, 2022. doi:10.1063/5.0127465.

[10] D. Juba, D. J. Audus, M. Mascagni, J. F. Douglas, and W. Keyrouz. ZENO: Software for calculating hydrodynamic, electrical, and shape properties of polymer and particle suspensions. *J. Res. Natl. Inst. Stand. Technol*, **122**(20):10–6028, 2017.

[11] D. Frenkel, G. C. A. M. Mooij, and B. Smit. Novel scheme to study structural and thermal properties of continuously deformable molecules. *Journal of Physics: Condensed Matter*, **3**:3053, 1991. doi:10.1088/0953-8984/4/12/006.

[12] J. I. Siepmann and D. Frenkel. Configurational bias Monte Carlo: a new sampling scheme for flexible chains. *Mol. Phys.*, **75**:59–70, 1992. doi:10.1080/00268979200100061.

[13] X. Li and Y. C. Chiew. Monte Carlo simulation of Lennard-Jones chains. *J. Chem. Phys.*, **101**:2522,

1994. doi:10.1063/1.467691.

[14] F. A. Escobedo and J. J. de Pablo. Extended continuum configurational bias Monte Carlo methods for simulation of flexible molecules. *J. Chem. Phys.*, **102**:2636, 1995. doi:10.1063/1.468695.

[15] M. P. Taylor, W. Paul, and K. Binder. Applications of the Wang-Landau algorithm to phase transitions of a single polymer chain. *Polym. Sci. Ser. C*, **55**:23–38, 2013. doi:10.1134/S1811238213060040.

[16] K. G. Honnell, J. G. Curro, and K. S. Schweizer. Local structure of semiflexible polymer melts. *Macromolecules*, **23**:3496–3505, 1990.

[17] Y. C. Chiew. Percus-Yevick integral equation theory for athermal hard-sphere chains. II. Average intermolecular correlation functions. *Mol. Phys.*, **73**:359–373, 1991.

[18] J. R. Elliott. One-the-fly second virial coefficient. *J. Phys. Chem. B*, **125**:4494–4500, 2021.

[19] Y. C. Chiew and V. Sabesan. Second virial coefficients of Lennard-Jones chains. *Fluid Phase Equilib.*, **155**:75–83, 1999.

C Supporting Information: Beyond One-Fluid Approximations for the Thermodynamics of Fluid Mixtures

The content of this chapter is a literal quote of the Supporting Information of the publication: A. Reimer, T. Winkler-Markert and J. Gross. Beyond One-Fluid Approximations for the Thermodynamics of Fluid Mixtures. *J. Chem. Phys.*, **164**, 014117 (2026), doi:10.1063/5.0307122

C.1 Reference Fluid – Hard-Sphere Mixture

C.1.1 Compressibility Factor

The properties of the hard-sphere (HS) reference fluid are described using the Boublík, Mansoori, Carnahan, Starling, Leland (BMCSL) equation of state.[1,2] For a hard-sphere mixture containing N^{c}, the compressibility factor is expressed as

$$Z^{\mathrm{hs}} = \frac{1}{\zeta_0}\left(\frac{\zeta_0\zeta_3}{1-\zeta_3} + \frac{3\zeta_1\zeta_2}{(1-\zeta_3)^2} + \frac{3\zeta_2^3-\zeta_3\zeta_2^3}{(1-\zeta_3)^3}\right) \tag{C.1}$$

where

$$\zeta_n = \frac{\pi}{6}\rho\sum_{i=1}^{N^{\mathrm{c}}} x_i\sigma_{ii}^n \tag{C.2}$$

with the number density $\rho = N/V$, the mole fraction $x_i = N_i/N$ of component i, and the hard-sphere diameter σ_{ii}. The attractive contribution to the total compressibility factor is then obtained as

$$Z^{\mathrm{att}} = Z - Z^{\mathrm{hs}} \tag{C.3}$$

C.1.2 Radial Distribution Functions

Boublík[3] developed an analytical model for the radial distribution functions (RDFs) of hard-sphere (hs) mixtures. The model expresses the RDFs in terms of the chemical potentials of the hs components and that of a hard-dumbbell (hd) in a mixture with hard spheres. The model provides reasonable accuracy up to dimensionless particle separations $r^* \leq 1.5$.[4]

C.1.2.1 Pure Fluids

For pure hard-sphere fluids, the RDF model of Boublík[3] can be expressed as

$$g^{\mathrm{hs}}(r^*,\eta) = \begin{cases} 0 & \text{for } r^* < 1 \\ \exp\left(2\beta\Delta\mu^{\mathrm{hs}} - \beta\Delta\mu^{\mathrm{hd}}(r^*)\right) & \text{for } r^* \geq 1 \end{cases} \tag{C.4}$$

where $r^* = r/\sigma$ and $\eta = \frac{\pi}{6}\rho\sigma^3$ is the packing fraction. The residual chemical potential of the pure hs fluid is given by

$$\beta\Delta\mu^{\mathrm{hs}}(\eta) = -\ln(1-\eta) + \frac{7\eta}{(1-\eta)} + \frac{15\eta^2}{2(1-\eta)^2} + \frac{\eta^3(2-\frac{\eta}{3})}{(1-\eta)^3} \tag{C.5}$$

and that of a single hd with reduced site-site distance r^* diluted in the hs fluid is

$$\begin{aligned}\beta\Delta\mu^{\mathrm{hd}}(r^*,\eta) = &-\ln(1-\eta) + \frac{(3R^* + 3S^* + V^*)\eta}{(1-\eta)} \\ &+ \frac{(3R^{*2} + 6S^* + 6V^*)\eta^2}{2(1-\eta)^2} + \frac{V^*\eta^3(2-\frac{\eta}{3})}{(1-\eta)^3}\end{aligned} \tag{C.6}$$

with geometric parameters

$$R^* = 1 + \tfrac{1}{2}r^*, \quad S^* = 1 + r^*, \quad V^* = 1 + \tfrac{3}{2}r^* - \tfrac{1}{2}r^{*3} \tag{C.7}$$

C.1.2.2 Mixtures

For a hard-sphere mixture, the RDF between components i and j can be approximated as[3]

$$g_{ij}^{\mathrm{hs}}(r_{ij}^*, \zeta_0, \zeta_1, \zeta_2, \zeta_3) = \exp\left(\beta\Delta\mu_i^{\mathrm{hs}} + \beta\Delta\mu_j^{\mathrm{hs}} - \beta\Delta\mu_{ij}^{\mathrm{hd}}(r_{ij}^*)\right) \tag{C.8}$$

where $r_{ij}^* = 2r/(\sigma_{ii}+\sigma_{jj})$. The chemical potential of hard-sphere component i in the mixture is given by

$$\begin{aligned}\beta\Delta\mu_i^{\mathrm{hs}} = &-\ln(1-\zeta_3) + \frac{(3\zeta_2 + 3\zeta_1\sigma_{ii} + \zeta_0\sigma_{ii}^2)\sigma_{ii}}{1-\zeta_3} \\ &+ \frac{(9\zeta_2^2 + 6\zeta_1\zeta_2\sigma_{ii})\sigma_{ii}^2}{2(1-\zeta_3)^2} + \frac{(6-\zeta_3)\zeta_2^3\sigma_{ii}^3}{3(1-\zeta_3)^3}\end{aligned} \tag{C.9}$$

The chemical potential of the hard-dumbbell $\Delta\mu_{ij}^{\mathrm{hd}}(r_{ij}^*)$, depends on the size ratio $p = \frac{\sigma_{ii}}{\sigma_{jj}}$. For $r_{ij}^* < \frac{p-1}{p+1}$, it is approximated by the chemical potential of the larger hard-sphere:

$$\beta\Delta\mu_{ij}^{\mathrm{hd}} = \begin{cases} \beta\Delta\mu_j^{\mathrm{hs}} & \text{for } p < 1 \\ \beta\Delta\mu_i^{\mathrm{hs}} & \text{for } p \geq 1 \end{cases} \tag{C.10}$$

For $r_{ij}^* \geq \frac{p-1}{p+1}$, $\Delta\mu_{ij}^{\mathrm{hd}}(r_{ij}^*)$ is computed as

$$\begin{aligned}\beta\Delta\mu_{ij}^{\mathrm{hd}} = &-\ln(1-\zeta_3) + \frac{R_i s + S_i r + V_i \rho}{1-\eta} \\ &+ \frac{R_i^2 s^2 + 2S_i q s + 6V_i r s}{6(1-\eta)^2} + \frac{V_i q s^2 (2-\eta/3)}{9(1-\eta)^3}\end{aligned} \tag{C.11}$$

with

$$s = 6\zeta_2, \quad r = \frac{3}{\pi}\zeta_1, \quad q = \frac{3}{2\pi}\zeta_2 \tag{C.12}$$

and

$$R_{ij} = \frac{1}{2}\left(1 + \frac{(r_{ij}^*(p+1) + (p-1))^2}{4r_{ij}^*(p+1)}\right)\sigma_{jj} \tag{C.13}$$

$$S_{ij} = \pi\left(\frac{1}{2}\left(p^2+1\right) + \frac{1}{4}r_{ij}^*(p+1)^2 + \frac{(p-1)^2}{4r_{ij}^*}\right)\sigma_{jj}^2 \tag{C.14}$$

$$\begin{aligned}V_{ij} = \frac{\pi}{12}\Bigg\{&(p^3+1) + \frac{3}{4}\left[r_{ij}^*(p+1)(p^2+1) + \frac{(p-1)(p^2-1)}{r_{ij}^*}\right] \\ &- \frac{1}{8}\left[r_{ij}^{*3}(p+1)^3 + \frac{3(p+1)(p-1)^2}{r_{ij}^*}\right]\Bigg\}\sigma_{jj}^3\end{aligned} \tag{C.15}$$

C.2 Finding an Analytic Expression for the Correction Factor

C.2.1 Symbolic Regression

Suitable analytic expressions for the correction factor $f_x(\zeta)$ were determined using symbolic regression implemented in the open-source software package PySR.[5] The objective of the regression was to identify simple functional forms for $h_x(\Pi_0, \Pi_1, \Pi_2, \Pi_3)$ that approximate $f_x^{\text{true}} - 1$. The `PySRRegressor` was configured as follows:

```
PySRRegressor(
    niterations=1000,
    binary_operators=["+", "-", "*", "/"],
    unary_operators=[ "square", "cube", "inv", "sqrt",
    "x_minus_1(x)=x-1", "log" ],
    extra_sympy_mappings={"x_minus_1": lambda x: x - 1},
    elementwise_loss="loss(x,y)=(x-y)^2",
    populations=40,
    population_size=60,
    maxsize=12,
    verbosity=0,
    complexity_of_constants=4,
    constraints={'^': (-2, 2)},
    nested_constraints={"log": {"log": 0}}
    )
```

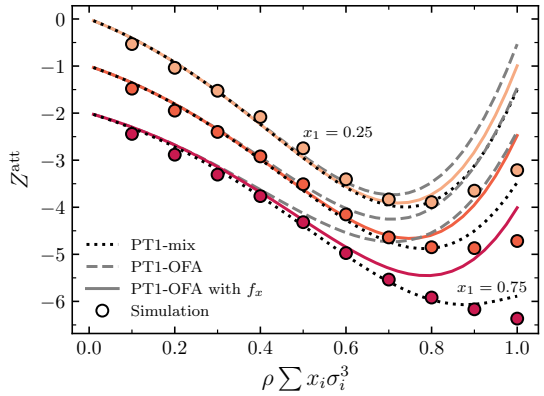
Ten independent regression runs were performed using the full set of variables (features) Π_0–Π_3. An additional ten runs were conducted using only the linearly independent subset of variables (Π_0, Π_1, Π_2), with the complexity of constants reduced to 2.

C.2.2 Selection of the Best Candidate Expression

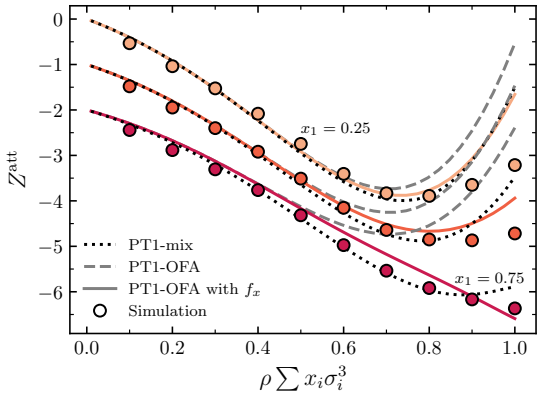
All candidate equations for $h_x(\Pi_0, \Pi_1, \Pi_2, \Pi_3)$ identified within the 20 symbolic regression runs were collected. Duplicates and expressions that do not reduce to zero for pure fluids were removed. To identify the most suitable functional forms, the effect of each remaining candidate expressions on predictions of the residual internal energy U^{res} and the attractive contribution to the compressibility factor Z^{att} for binary square-well mixtures was evaluated. These properties were calculated using first-order perturbation theory (PT1) in three variants: (1) the rigorous formulation evaluated with the mixture RDF model of Boublík[3] (PT1-mix), (2) the one-fluid approximation employing the pure-fluid RDF of Boublík[3] (PT1-OFA), and (3) the one-fluid approximation including the correction factor f_x (PT1-OFA with f_x).

Two binary square-well systems were considered in a first selection step: one with a size ratio $\sigma_{22}/\sigma_{11} = 3$, energy ratio $\varepsilon_{22}/\varepsilon_{11} = 1$, and reduced temperature $T^* = k_B T/\varepsilon_{11} = 1.25$; and

another with $\sigma_{22}/\sigma_{11} = 5$, $\varepsilon_{22}/\varepsilon_{11} = 1$, and $T^* = 2.0$. Candidate expressions were discarded if the PT1-OFA with f_x predictions fell outside the range bounded by the rigorous PT1-mix and the uncorrected PT1-OFA results. This selection criterion ensures that the correction factor improves the OFA predictions without overcompensating for the errors introduced by the OFA. In figure C.1, PT1 property predictions are shown for two candidate expressions. The candidate corresponding to the left figure consistently improves the results, whereas the one in the right figure overcompensates and was subsequently discarded.



(a) Candidate equation leading to consistent improvement.



(b) Candidate equation discarded due to overcompensation.

Figure C.1: Attractive contribution to the compressibility factor for a binary square-well mixture with $\sigma_2/\sigma_1 = 3$, $\varepsilon_2/\varepsilon_1 = 1$, and reduced temperature $T^* = 1.25$. Comparison of PT1-mix (dotted lines), PT1-OFA (dashed lines), and PT1-OFA with f_x (solid lines) results to molecular simulation data from Akhouri and Solana[6]. For clarity, the lines and symbols representing different mole fractions x_1 are shifted downward by 1 relative to the curve above them.

Further selection of candidate expressions was performed based on the mean absolute relative deviation (MARD) between the PT1-mix and the PT1-OFA with f_x predictions,

$$\mathrm{MARD}(X, f_x) = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{X_i^{\text{PT1-OFA with } f_x} - X_i^{\text{PT1-mix}}}{X_i^{\text{PT1-mix}}} \right| \tag{C.16}$$

where $X \in \{\beta U^{\mathrm{res}}/N, Z^{\mathrm{att}}\}$ and $N = 720$ state points, corresponding to the size-asymmetric binary square-well systems considered by Akhouri and Solana[6]. Candidate expressions that did not reduce the MARD for both properties by at least 50% relative to $f_x = 1$ were discarded. Table C.1 summarizes the remaining expressions and the corresponding MARD values. The expression with the lowest complexity among these was selected as the optimal candidate.

Table C.1: Candidate expressions for the correction factor obtained via symbolic regression. MARD values are reported relative to predictions from rigorous first-order perturbation theory (PT1-mix).

Equation for h_x	Equation for f_x $f_x = 1 + h_x$	MARD Z^{att}/%	MARD $\beta U^{\text{res}}/N$/%
0	1	12.32	1.9
$\Pi_0^3(\Pi_2 - 1)$	$1 + \zeta_3^3\left(\frac{\zeta_3^2\zeta_0}{\zeta_2^3} - 1\right)$	4.61	0.79
$\Pi_0^4\Pi_1^2\log(\Pi_2^2)$	$1 + 2\zeta_3^4\left(\frac{\zeta_3\zeta_1}{\zeta_2^2}\right)^2\log\left(\frac{\zeta_3^2\zeta_0}{\zeta_2^3}\right)$	2.92	0.62
$\Pi_0^4\Pi_1^4\sqrt{\Pi_3 - 1}$	$1 + \zeta_3^4\left(\frac{\zeta_3\zeta_1}{\zeta_2^2}\right)^4\sqrt{\left(\frac{\zeta_0\zeta_3}{\zeta_1\zeta_2} - 1\right)}$	3.26	0.67

C.3 Application of the Correction Factor to the uv-Theory and the PC-SAFT Equation of State

To assess the generality of the correction factor f_x, we analyse its transferability to LJ fluids using three different equations of state: Both variants of the uv-theory[7], denoted as uv-BH-theory and uv-WCA-theory and the PC-SAFT equation of state[8].

The uv-theory describes the Helmholtz energy of Mie fluids through interpolation between a rigorous upper and an effective lower bound, which are both based on first-order perturbation theories. Incorporating the correction factor f_x, the reduced Helmholtz energy within the uv-theory is expressed as

$$a = a_0 + f_x \Delta a_1^u + (1 - \phi^u)(\Delta B_2 - f_x \Delta B_{21}^u)\rho \tag{C.17}$$

where a_0 is the Helmholtz energy of the reference fluid (either BH or WCA), Δa_1^u is the first-order perturbation term, and ϕ^u is the interpolation function, denoted as u-fraction. The term within the brackets ensures that the uv-theory reproduces the exact low-density limit, with the residual second virial coefficient $\Delta B_2 = B_2 - B_{20}$, where B_{20} is the virial coefficient of the reference fluid, and $\Delta B_{21}^u = (\partial \Delta a_1^u / \partial \rho)_{\rho=0}$.

The PC-SAFT equation of state[8] models dispersive chain molecules based on a second order perturbation theory, using a reference system of repulsive chains. Including the correction factor f_x, the PC-SAFT equation of state can be written as

$$a = a^{\text{hc}} + a^{\text{disp}} = a^{\text{hc}} + f_x a_1 + a_2, \tag{C.18}$$

where a^{hc} is the hard-chain reference contribution, and a^{disp} is the dispersive term composed of the first- and second-order perturbation contributions, a_1 and a_2, respectively.

The notation for the Helmholtz energy contributions in eqs. (C.17) and (C.18) follows the conventions of the original works.[7,8] In the uv-theory, Δa_1^u, and in the PC-SAFT equation of state, a_1, denote the analytical expressions for the first-order terms in the respective u-expansions.

C.4 Molecular Simulations

All molecular simulations in this work are conducted in cubic boxes with periodic boundary conditions. A global cutoff radius of at least $r_c \geq 3\sigma_{max}$, where $\sigma_{max} = \max_i \sigma_{ii}$ denotes the largest size parameter in the mixture, is applied together with analytical long-range corrections. Argon-like LJ parameters, $\sigma_{11} = 3.402\,\text{Å}$ and $\varepsilon_{11}/k_B = 120\,\text{K}$ are used, with the mass of all particles set to $m = 39.948\,\text{u}$.

C.4.1 Single Phase Properties from Molecular Dynamics Simulations

Molecular dynamics simulations are carried out with the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[9] using the velocity Verlet integration scheme and a dimensionless time step of $\Delta\tau^* = \Delta\tau/\sqrt{m\sigma_{11}^2/\varepsilon_{11}} \approx 0.46454 \times 10^{-3}$. Each simulation consists of an energy minimization, followed by at least 5×10^5 equilibration steps and 5×10^6 production steps. Initial densities are calculated with the uv-WCA-theory.

Binary LJ mixtures are simulated in the canonical ensemble to obtain pressures and energies. The total number of particles is at least $N = 1800$ for mixtures with $\sigma_{22}/\sigma_{11} \leq 2.0$, and $N = 2400$ for mixtures with $\sigma_{22}/\sigma_{11} > 2.0$ and for all simulations at pressures $p\sigma_{11}^3/\varepsilon_{11} \leq 0.2$. The cutoff radius is set globally to $r_c = 3.5\sigma_{max}$, except for systems with $\sigma_{22}/\sigma_{11} = 3.0$ and $p\sigma_{11}^3/\varepsilon_{11} \geq 0.25$, where $r_c = 3.0\sigma_{max}$ is used. A Nosé–Hoover thermostat with a damping parameter of $2000\,\Delta\tau^*$ is applied.[10,11] The simulation settings are verified and validated against molecular simulation results of Ely[12], showing a maximum absolute relative deviation of 1.764% in pressure and 0.1924% in potential energy, when defining the cut-off radii as proposed by the author.

Ternary and quinary LJ mixtures are simulated in the isobaric–isothermal ensemble with $N = 2400$ and $N = 4800$ particles, respectively, to obtain densities and energies. Structural information in terms of the radial distribution functions $g_{ij}(r)$ is also sampled and shown for two representative systems in figures C.2 and C.3. A Nosé–Hoover thermostat and barostat with damping parameters $100\,\Delta\tau^*$ are applied. The cut-off radius in all simulations of ternary and quinary mixtures is set to $r_c = 3\,\sigma_{max}$. The simulation settings are verified and validated

against molecular simulation results for ternary mixtures of Fotouh and Shukla [13], showing a maximum absolute relative deviation of 0.213% in density and potential energy.

The results of the molecular simulations for binary, ternary and quinary LJ mixtures conducted in this work are provided as text files.

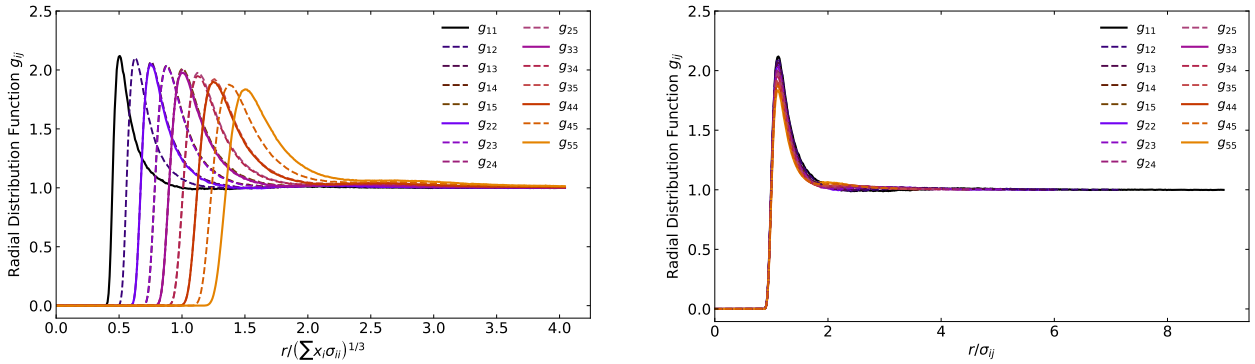


Figure C.2: Radial distribution functions g_{ij} of the gaseous phase for a equimolar quinary mixture with mole fractions $x_i = 0.2$ for all species. Size parameter ratios: $\sigma_{ii}/\sigma_{11} = \{1.00, 1.50, 2.00, 2.50, 3.00\}$; energy parameter ratios: $\varepsilon_{ii}/\varepsilon_{11} = 1$. Thermodynamic state: $k_B T/\varepsilon_{11} = 1.50 \pm 0.02$, $p\,\sigma_{11}^3/\varepsilon_{11} = 0.015 \pm 0.001$, and $\rho \sum_i x_i \sigma_{ii}^3 = 0.151 \pm 0.003$. The same data are presented using two different scaling conventions for the radial coordinate: radius r scaled with $\sigma_x^3 = \sum_i x_i \sigma_{ii}^3$ (left); radius r scaled with the arithmetic mean of the size parameter σ_{ij} (right).

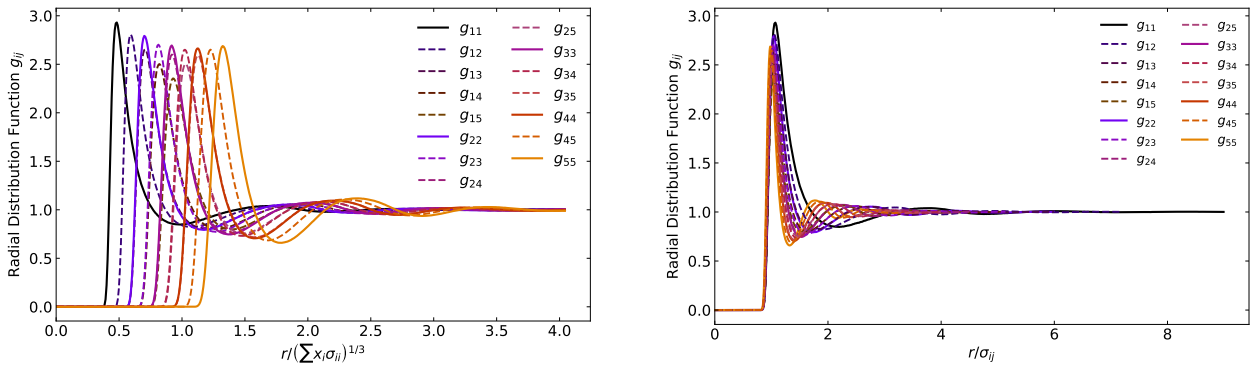


Figure C.3: Radial distribution functions g_{ij} of the liquid-like phase for a quinary equimolar mixture with mole fractions $x_i = 0.2$ for all species. Size parameter ratios: $\sigma_{ii}/\sigma_{11} = \{1.00, 1.50, 2.00, 2.50, 3.00\}$; energy parameter ratios: $\varepsilon_{ii}/\varepsilon_{11} = 1$. Thermodynamic state: $k_B T/\varepsilon_{11} = 3.00 \pm 0.03$, $p\,\sigma_{11}^3/\varepsilon_{11} = 1.953 \pm 0.001$, and $\rho \sum_i x_i \sigma_{ii}^3 = 1.057 \pm 0.002$. The same data are presented using two different scaling conventions for the radial coordinate: radius r scaled with $\sigma_x^3 = \sum_i x_i \sigma_{ii}^3$ (left); radius r scaled with the arithmetic mean of the size parameter σ_{ij} (right).

C.4.2 Phase Equilibrium Properties from Grand Canonical Monte Carlo Simulations

Phase equilibrium properties of binary LJ mixtures are determined using grand canonical Monte Carlo simulations with the transition-matrix scheme (TM-GCMC) and histogram reweighting. A brief outline of the method is provided below, with further details available in Refs. 14–18

The TM-GCMC simulations are performed at constant excess chemical potentials $\boldsymbol{\mu}^{\mathrm{sim}} = (\mu_1, \mu_2)$, volume V, and temperature T. We define the excess chemical potential of a component μ_i in terms of the total chemical potential μ_i^{tot}, the de Broglie wavelength $\Lambda_i(T)$, and the inverse temperature $\beta = (k_{\mathrm{B}}T)^{-1}$ as $\mu_i = \mu_i^{\mathrm{tot}} - \beta^{-1}\ln(\Lambda_i^3(T)/^3)$. To streamline the notation, we omit "excess" in what follows. The particle numbers $\mathbf{N} = (N_1, N_2)$ fluctuate within a predefined domain $\boldsymbol{\Omega}$, representing the region in particle-number space that includes all relevant vapor and liquid states. The extent of $\boldsymbol{\Omega}$ is estimated from vapor–liquid equilibrium predictions based on the uv-WCA theory, following the approach illustrated in Fig. 2 of ref. 14. To efficiently sample the full range of relevant particle numbers, the domain $\boldsymbol{\Omega}$ is stratified into overlapping subdomains of size $5 \leq \Delta N_1 = \Delta N_2 \leq 8$. Each subdomain is simulated independently with 3.0×10^7 equilibration and 6.0×10^7 production steps. Trial moves consist of 50% translations, 20% insertions, 20% deletions, and 10% identity swaps. Moves leading outside the subdomain boundaries are rejected. A transition-matrix sampling scheme is employed to ensure flat-histogram sampling. The normalized probability distribution $P(\mathbf{N}; T, \boldsymbol{\mu}^{\mathrm{sim}})$, representing the likelihood of observing a particle-number state $\mathbf{N}$ under the simulation conditions, is obtained by combining the histograms and the transition matrices from all subdomain simulations.

Histogram reweighting[14] is employed to determine the coexistence chemical potentials $\boldsymbol{\mu}^{\mathrm{coex}}$ and the corresponding probability distribution $P^{\mathrm{coex}}(\mathbf{N}; T, \boldsymbol{\mu}^{\mathrm{coex}})$. The chemical potentials are iteratively adjusted, and the reweighted distribution is calculated at each step as

$$P^{\mathrm{coex}}(\mathbf{N}; T, \boldsymbol{\mu}^{\mathrm{coex}}) = P(\mathbf{N}; T, \boldsymbol{\mu}^{\mathrm{sim}}) \exp\left(\mathbf{N}\beta(\boldsymbol{\mu}^{\mathrm{coex}} - \boldsymbol{\mu}^{\mathrm{sim}})\right) \tag{C.19}$$

with appropriate normalization applied. The iteration is continued until the distribution becomes bimodal, exhibiting two pronounced peaks whose summed probabilities are equal, corresponding to the vapour and liquid phases.

At the coexistence chemical potentials $\boldsymbol{\mu}^{\mathrm{coex}}$, the vapour–liquid equilibrium properties are obtained from the probability distribution $P^{\mathrm{coex}}(\mathbf{N}; T, \boldsymbol{\mu}^{\mathrm{coex}})$. The ensemble-averaged number of particles of component i in phase $\phi \in \{\mathrm{vap}, \mathrm{liq}\}$ is

$$\langle N_i \rangle^{\phi} = \sum_{\mathbf{N} \in \boldsymbol{\Omega}^{\phi}} 2N_i P^{\mathrm{coex}}(\mathbf{N}; T, \boldsymbol{\mu}^{\mathrm{coex}}) \tag{C.20}$$

where $\mathbf{\Omega}^{\phi}$ is the particle-number domain associated with phase ϕ. The phase compositions follow as

$$x_i^{\phi} = \frac{\langle N_i \rangle^{\phi}}{\langle N_1 \rangle^{\phi} + \langle N_2 \rangle^{\phi}} \tag{C.21}$$

The absolute saturation pressure p^{sat} is obtained in the ideal gas limit from the probability distribution at $N_1 = N_2 = 0$, as

$$p^{\mathrm{sat}}(T) = -\frac{k_{\mathrm{B}}T}{V} \ln\left(2P^{\mathrm{coex}}(\mathbf{N} = \mathbf{0}; T, \boldsymbol{\mu}^{\mathrm{coex}})\right) \tag{C.22}$$

The VLE properties can be calculated continuously for all $\boldsymbol{\mu}^{\mathrm{coex}}$. In this work, we provide 20 evaluations for each mixture. An overview of the simulation settings is provided in table C.2.

σ_{22}/σ_{11}	$\varepsilon_{22}/\varepsilon_{11}$	$k_{\mathrm{B}}T/\varepsilon_{11}$	V/σ_{11}^3	$\Delta N_1 = \Delta N_2$	$r_{\mathrm{c}}/\sigma_{\mathrm{max}}$
1.5	1.0	1.0	888.6	5	3.2
1.5	0.66	1.0	734.6	5	3.0
1.5	0.66	7/6	734.6	5	3.0
0.5	1.0	1.0	222.9	5	3.0
2.0	1.0	1.25	1899.8	8	3.0
2.0	2.0	2.0	1899.8	8	3.0
2.0	1.515	1.5	1899.8	8	3.0
1/3	1.0	1.25	222.9	6	3.0

Table C.2: Simulation settings for the TM-GCMC calculations of binary Lennard–Jones mixtures.

C.5 Supplemental Figures

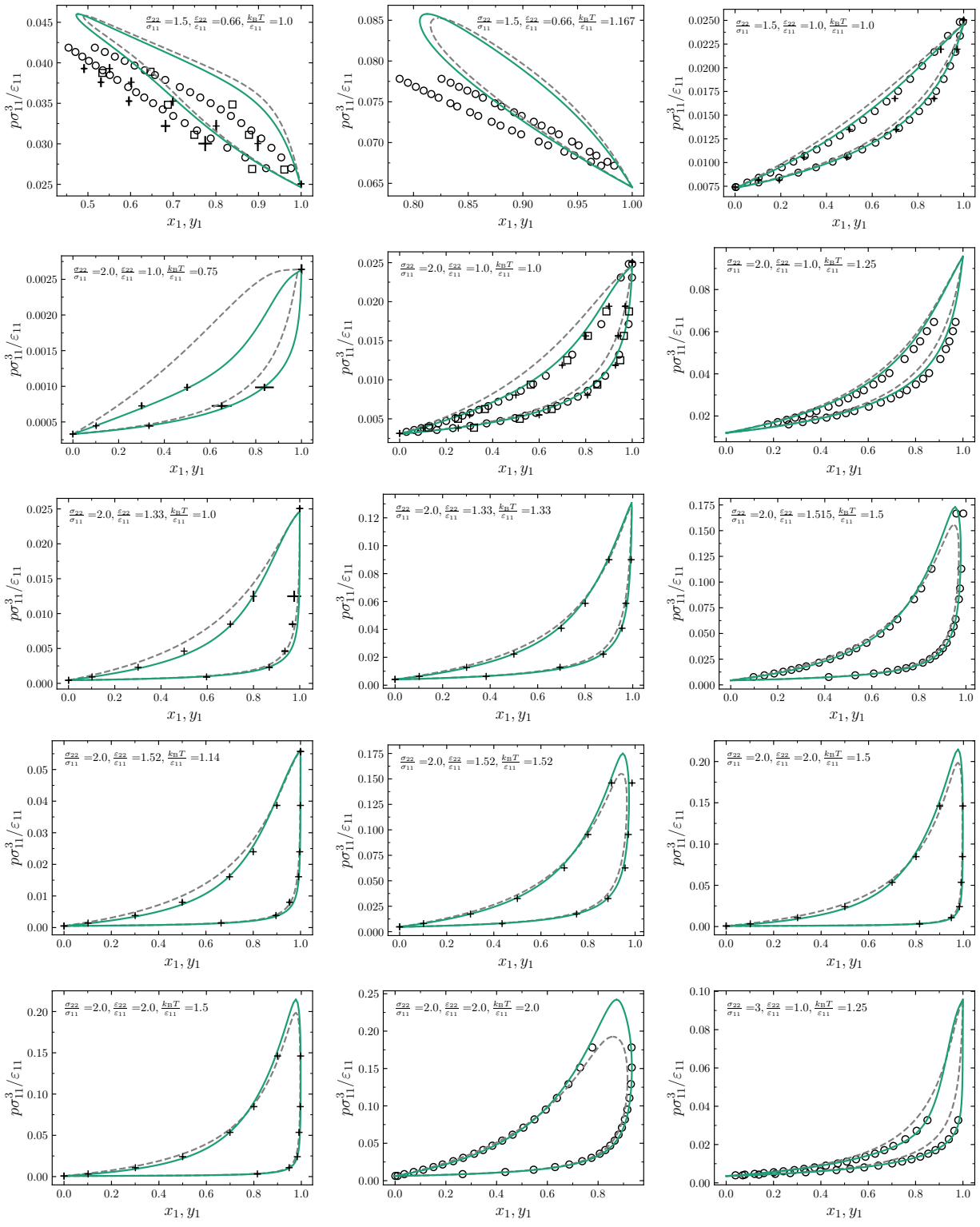


Figure C.4: Vapour–liquid equilibria of binary Lennard-Jones mixtures. Predictions calculated with the **uv-WCA-theory** without the correction factor (dashed lines) and with the correction factor (solid lines), compared to molecular simulation data from Vrabec *et al.* [19] (plus symbols), Harismiadis *et al.* [20] (squares), and this work (circles).

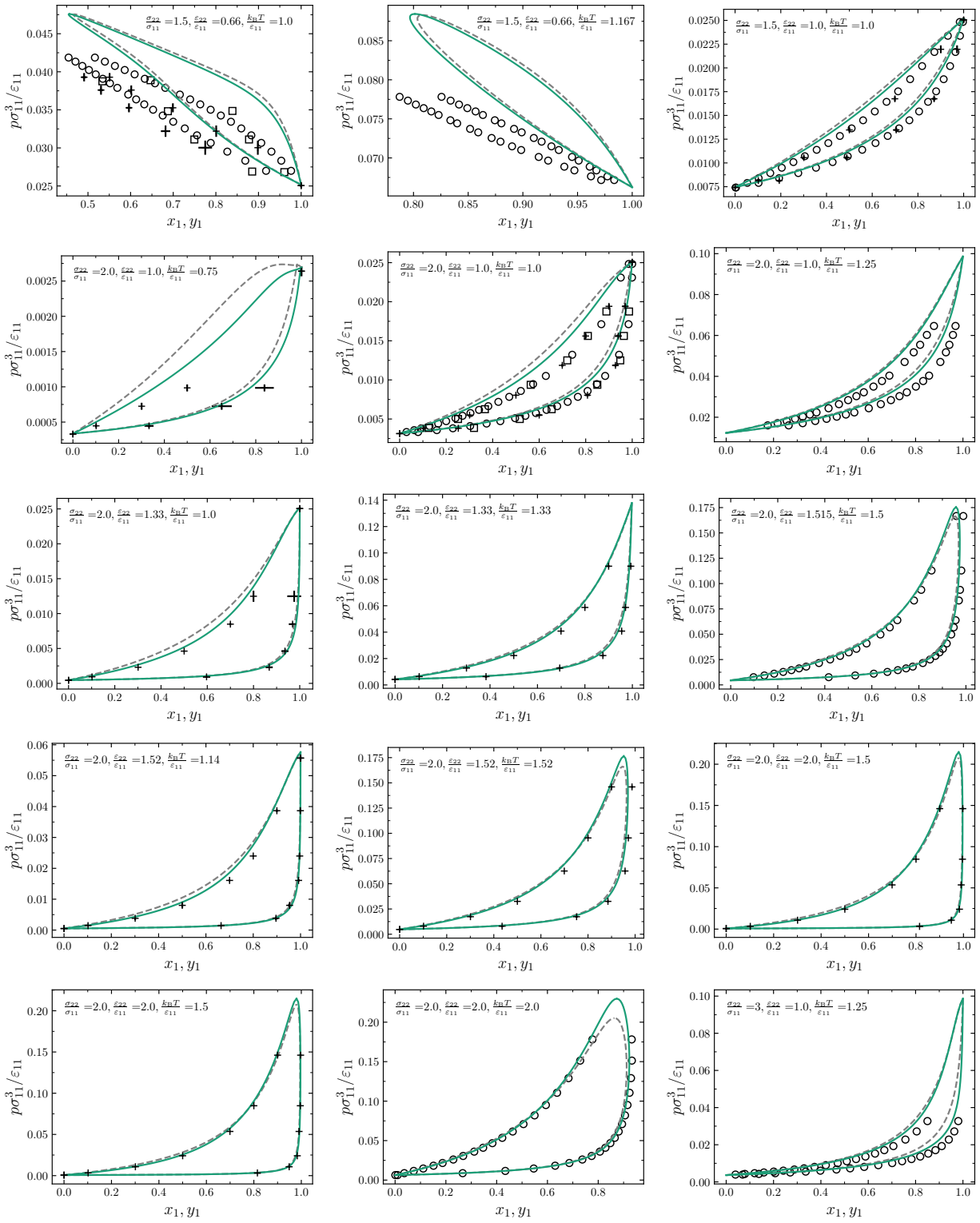


Figure C.5: Vapour–liquid equilibria of binary Lennard-Jones mixtures. Predictions calculated with the **uv-BH-theory** without the correction factor (dashed lines) and with the correction factor (solid lines), compared to molecular simulation data from Vrabec *et al.* [19] (plus symbols), Harismiadis *et al.* [20] (squares), and this work (circles).

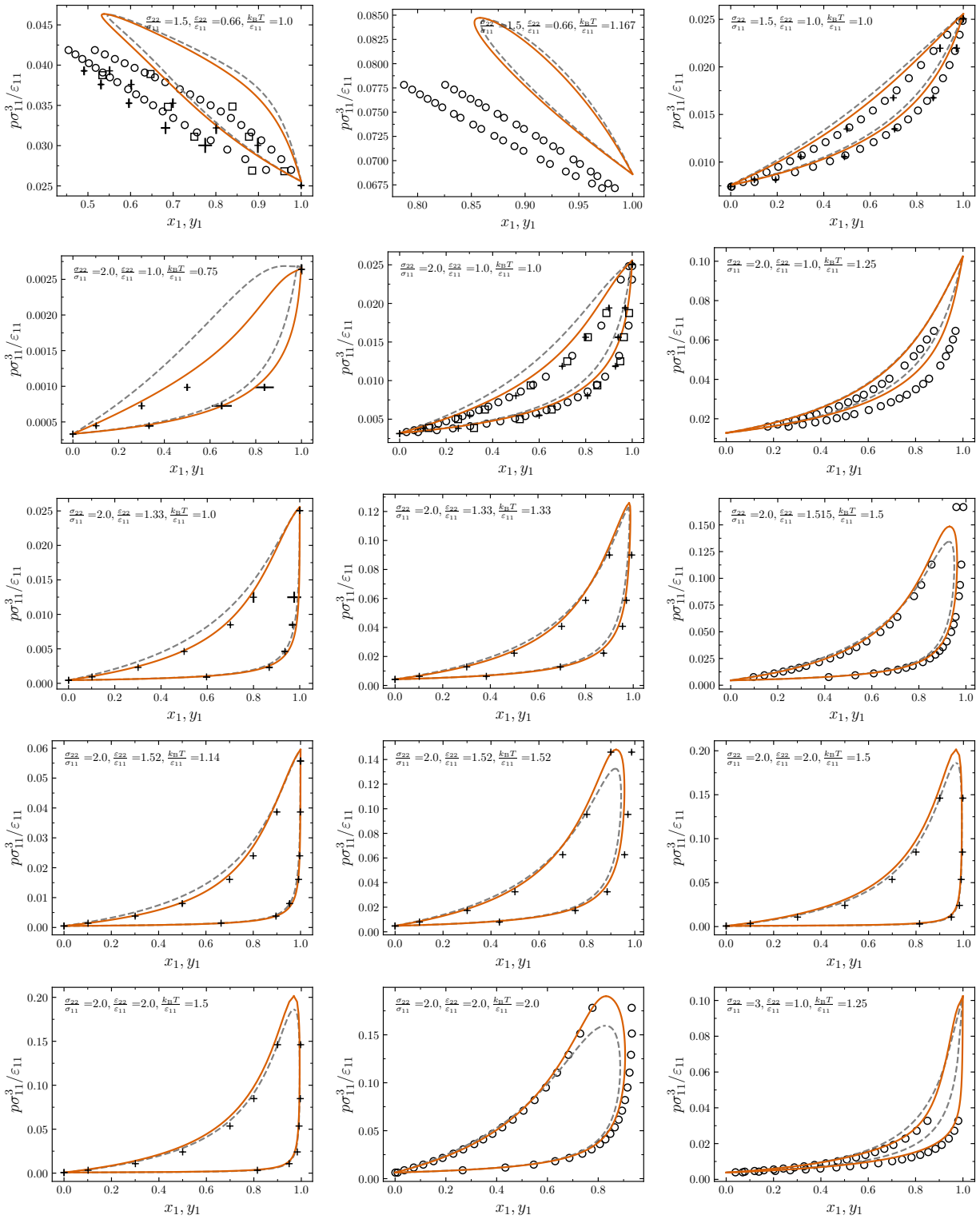
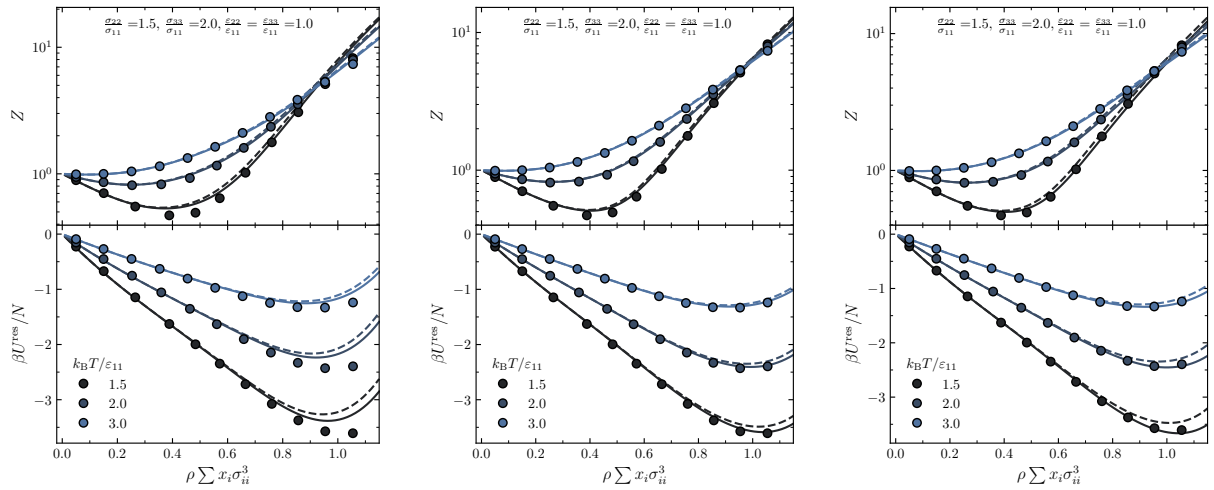


Figure C.6: Vapour–liquid equilibria of binary Lennard-Jones mixtures. Predictions calculated with the **PC-SAFT equation of state** without the correction factor (dashed lines) and with the correction factor (solid lines), compared to molecular simulation data from Vrabec *et al.* [19] (plus symbols), Harismiadis *et al.* [20] (squares), and this work (circles).



(a) PC-SAFT (b) *uv*-BH-theory (c) *uv*-WCA-theory

Figure C.7: Equimolar ternary LJ mixtures. Comparison of molecular simulation results from this work (circles) with equation-of-state predictions calculated with the original theories (dashed lines) and including the correction factor (solid lines).

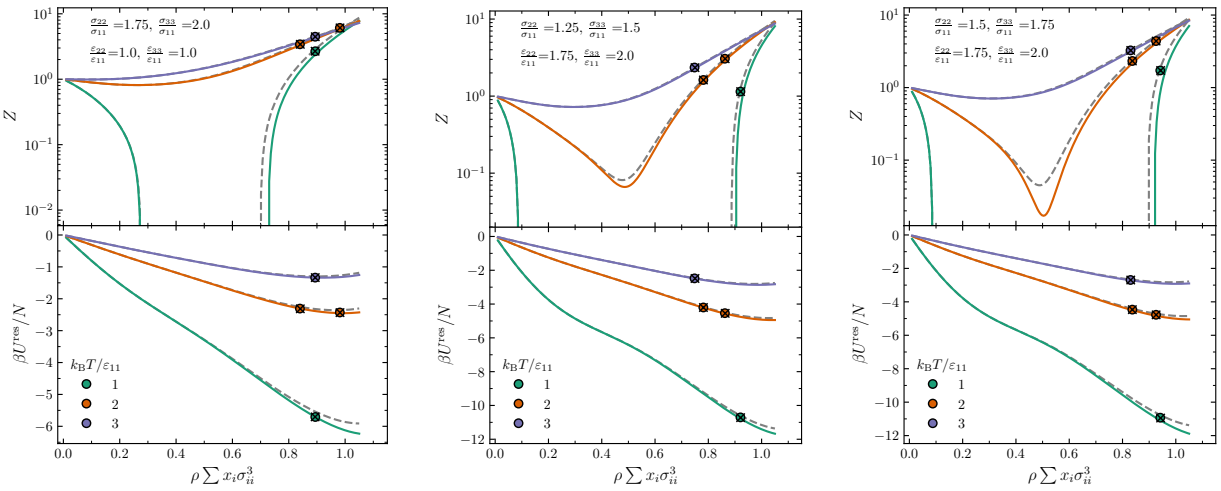


Figure C.8: Comparison of molecular simulation results of Fotouh and Shukla [13] (circles) and this work (crosses) for a ternary LJ mixture with *uv*-WCA-theory predictions calculated with the original theory (dashed lines) and including the correction factor (solid lines).

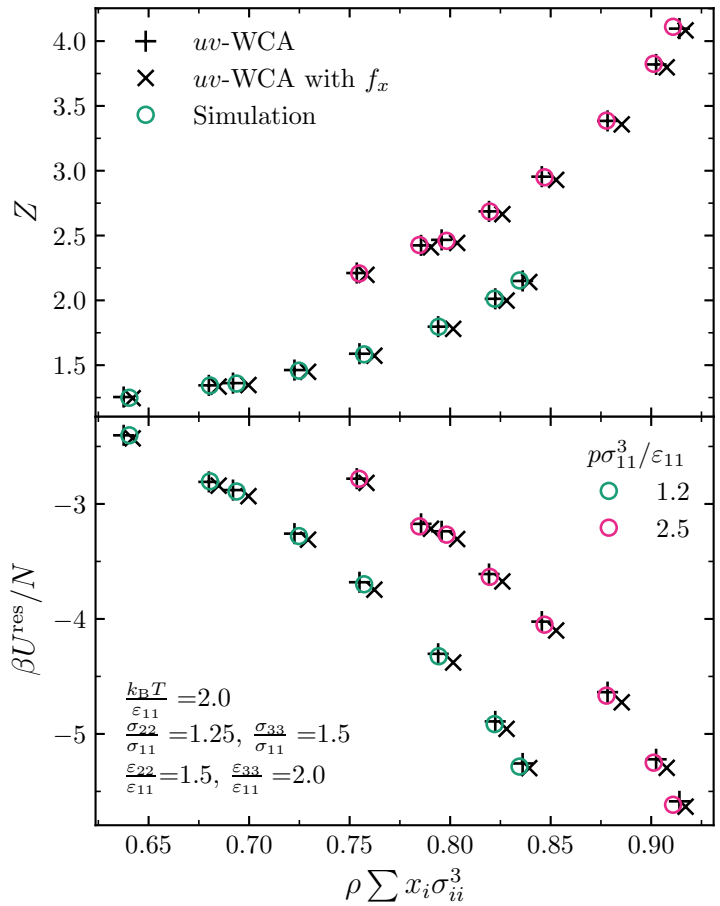


Figure C.9: Comparison of molecular simulation results of Fotouh and Shukla [13] (circles) with *uv*-WCA-theory predictions calculated with the original theory (plus symbols) and including the correction factor (crosses).

References

[1] T. Boublík. Hard-sphere equation of state. *J. Chem. Phys.*, **53**(1):471–472, 1970.

[2] G. Mansoori, N. F. Carnahan, K. Starling, and T. Leland Jr. Equilibrium thermodynamic properties of the mixture of hard spheres. *J. Chem. Phys.*, **54**(4):1523–1525, 1971.

[3] T. Boublík. Background correlation functions in the hard sphere systems. *Mol. Phys.*, **59**(4):775–793, 1986.

[4] T. Boublík. Hard-sphere radial distribution function from the residual chemical potential. *Mol. Phys.*, **104**(22-24):3425–3433, 2006.

[5] M. Cranmer. Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. 2023.

[6] B. Akhouri and J. Solana. Square-well mixtures revisited: computer simulation, mixing rules and one-fluid theory. *Mol. Simul.*, **46**(2):102–110, 2020.

[7] T. van Westen and J. Gross. Accurate thermodynamics of simple fluids and chain fluids based on first-order perturbation theory and second virial coefficients: uv-theory. *J. Chem. Phys.*, **155**(24), 2021.

[8] J. Gross and G. Sadowski. Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules. *Ind. Eng. Chem. Res.*, **40**(4):1244–1260, 2001.

[9] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, **271**:108171, 2022. doi:10.1016/j.cpc.2021.108171.

[10] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, **100**(1):191–198, 2002. doi:10.1080/00268970110089108.

[11] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, **31**(3):1695–1697, 1985. doi:10.1103/physreva.31.1695.

[12] J. Ely. A test of the mean density approximation for Lennard-Jones mixtures with large size ratios. *Int. J. Thermophys.*, **7**(2):381–393, 1986.

[13] K. Fotouh and K. Shukla. Excess properties of ternary fluid mixtures from simulation, perturbation theory and van der Waals one-fluid theory: size and energy parameter effects. *Chem. Eng. Sci.*, **52**(14):2369–2382, 1997.

[14] A. Hemmen, A. Z. Panagiotopoulos, and J. Gross. Grand Canonical Monte Carlo Simulations

Guided by an Analytic Equation of State—Transferable Anisotropic Mie Potentials for Ethers. *J. Phys. Chem. B*, **119**(23):7087–7099, 2015.

[15] V. K. Shen and J. R. Errington. Determination of fluid-phase behavior using transition-matrix Monte Carlo: Binary Lennard-Jones mixtures. *J. Chem. Phys.*, **122**(6):064508, 2005. doi:10.1063/1.1844372.

[16] J. R. Errington and V. K. Shen. Direct evaluation of multicomponent phase equilibria using flat-histogram methods. *J. Chem. Phys.*, **123**(16):164103, 2005. doi:10.1063/1.2064628.

[17] A. M. Ferrenberg and R. H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, **61**(23):2635–2638, 1988. doi:10.1103/physrevlett.61.2635.

[18] A. Z. Panagiotopoulos, V. Wong, and M. A. Floriano. Phase Equilibria of Lattice Polymers from Histogram Reweighting Monte Carlo Simulations. *Macromolecules*, **31**(3):912–918, 1998. doi:10.1021/ma971108a.

[19] J. Vrabec, A. Lotfi, and J. Fischer. Vapour liquid equilibria of Lennard-Jones model mixtures from the NpT plus test particle method. *Fluid Phase Equilib.*, **112**(2):173–197, 1995.

[20] V. Harismiadis, N. Koutras, D. Tassios, and A. Z. Panagiotopoulos. How good is conformal solutions theory for phase equilibrium predictions?: Gibbs ensemble simulations of binary Lennard-Jones mixtures. *Fluid Phase Equilib.*, **65**:1–18, 1991.

D Supporting Information: Primitive and Non-Primitive Model Electrolytes: Comparing Ion-Related Helmholtz Energies Using Molecular Simulations

The content of this chapter is a literal quote of the appendix of the publication: A. Reimer, I. Reisch and J. Gross. Primitive and non-primitive model electrolytes: Comparing ion-related Helmholtz energies using molecular simulations. *J. Chem. Phys.*, **162**, 124503 (2025), doi: 10.1063/5.0257401

D.1 Molecular Simulation Results

Table D.1: Simulation results for systems with temperature $T^* = 4.0$ and density $\rho^* = 0.2$. The uncertainty in the last digit, according to the 95% confidence interval, is shown in parentheses.

μ^{*2}	q^{*2}	x^{ions}	a^{dd}	$a^{\text{cc+cd}}$	ε_{r}
0.50	15	0.005	-0.04771(8)	-0.02460(6)	1.47(1)
0.50	15	0.010	-0.04729(8)	-0.0520(1)	1.46(1)
0.50	15	0.030	-0.04542(8)	-0.1722(2)	1.46(1)
0.50	15	0.050	-0.04367(8)	-0.3011(2)	1.45(1)
1.00	15	0.005	-0.1861(2)	-0.03006(6)	2.07(3)
1.00	15	0.010	-0.1844(2)	-0.06210(9)	2.04(2)
1.00	15	0.030	-0.1771(2)	-0.1972(1)	2.04(2)
1.00	15	0.050	-0.1706(2)	-0.3386(2)	2.01(2)
1.00	25	0.005	-0.1861(2)	-0.0574(1)	2.09(3)
1.00	25	0.010	-0.1844(2)	-0.1179(2)	2.05(2)
1.00	25	0.030	-0.1771(2)	-0.3718(3)	2.02(2)
1.00	25	0.050	-0.1706(2)	-0.6365(4)	1.98(2)
1.25	15	0.005	-0.2891(2)	-0.03275(6)	2.45(4)

Continued on next page

Continued from previous page

μ^{*2}	q^{*2}	x^{ions}	a^{dd}	$a^{\text{cc+cd}}$	ε_{r}
1.25	15	0.010	-0.2864(2)	-0.06696(9)	2.43(3)
1.25	15	0.030	-0.2759(2)	-0.2096(1)	2.37(3)
1.25	15	0.050	-0.2655(2)	-0.3570(2)	2.33(3)
1.25	25	0.005	-0.2891(2)	-0.0601(1)	2.43(3)
1.25	25	0.010	-0.2864(2)	-0.1231(2)	2.39(3)
1.25	25	0.030	-0.2759(2)	-0.3854(2)	2.37(3)
1.25	25	0.050	-0.2655(2)	-0.6559(3)	2.32(3)

Table D.2: Simulation results for systems with temperature $T^* = 1.8$ and density $\rho^* = 0.8$. The uncertainty in the last digit, according to the 95% confidence interval, is shown in parentheses.

μ^{*2}	q^{*2}	x^{ions}	a^{dd}	$a^{\text{cc+cd}}$	ε_{r}
1.25	15	0.005	-0.9183(3)	-0.0504(2)	11.9(3)
1.25	15	0.010	-0.9107(3)	-0.1005(7)	11.7(4)
1.25	15	0.015	-0.9036(3)	-0.1512(6)	11.6(3)
1.25	15	0.020	-0.8963(3)	-0.2019(8)	11.3(3)
1.25	15	0.030	-0.8817(3)	-0.3029(7)	10.8(3)
1.25	15	0.050	-0.8526(3)	-0.506(1)	10.4(2)
1.25	30	0.005	-0.9183(3)	-0.1010(7)	11.6(4)
1.25	30	0.010	-0.9107(3)	-0.203(1)	11.3(4)
1.25	30	0.015	-0.9036(3)	-0.306(2)	11.3(3)
1.25	30	0.020	-0.8963(3)	-0.406(1)	11.2(3)
1.25	30	0.030	-0.8817(3)	-0.612(2)	10.7(3)
1.25	30	0.050	-0.8526(3)	-1.021(3)	9.80(23)
1.25	50	0.010	-0.9107(3)	-0.341(2)	11.6(3)
1.25	50	0.020	-0.8963(3)	-0.687(2)	11.1(3)
1.25	50	0.030	-0.8817(3)	-1.034(2)	10.6(3)
1.25	50	0.050	-0.8526(3)	-1.734(4)	10.1(3)
2.00	15	0.005	-1.9189(4)	-0.0537(2)	26(2)
2.00	15	0.010	-1.9041(5)	-0.1072(6)	27(2)
2.00	15	0.015	-1.8904(5)	-0.1616(5)	28(2)

Continued on next page

Continued from previous page

μ^{*2}	q^{*2}	x^{ions}	a^{dd}	$a^{\text{cc+cd}}$	ε_{r}
2.00	15	0.020	-1.8757(5)	-0.2148(8)	27(2)
2.00	15	0.030	-1.8472(5)	-0.3232(7)	24(2)
2.00	15	0.050	-1.7908(5)	-0.537(1)	22(1)
2.00	30	0.005	-1.9189(4)	-0.1085(6)	29(2)
2.00	30	0.010	-1.9041(5)	-0.2162(8)	26(2)
2.00	30	0.015	-1.8904(5)	-0.325(1)	25(2)
2.00	30	0.020	-1.8757(5)	-0.433(2)	25(2)
2.00	30	0.030	-1.8472(5)	-0.651(1)	25(2)
2.00	30	0.050	-1.7908(5)	-1.082(2)	21(2)
2.00	50	0.005	-1.9189(4)	-0.181(1)	26(2)
2.00	50	0.010	-1.9041(5)	-0.364(1)	26(2)
2.00	50	0.020	-1.8757(5)	-0.723(2)	24(2)
2.00	50	0.030	-1.8472(5)	-1.092(2)	24(2)
2.00	50	0.050	-1.7908(5)	-1.820(3)	21(2)
3.00	50	0.005	-3.5251(6)	-0.190(1)	75(6)
3.00	50	0.010	-3.5008(6)	-0.378(2)	72(11)
3.00	50	0.020	-3.4522(6)	-0.756(3)	60(7)
3.00	50	0.030	-3.4027(7)	-1.136(3)	56(7)
3.00	50	0.050	-3.3056(6)	-1.887(5)	49(5)
3.00	100	0.005	-3.5251(6)	-0.389(2)	78(14)
3.00	100	0.015	-3.4766(6)	-1.173(3)	59(7)
3.00	100	0.020	-3.4522(6)	-1.561(4)	59(6)
3.00	100	0.030	-3.4027(7)	-2.346(7)	52(5)
3.00	100	0.050	-3.3056(6)	-3.912(7)	46(6)

Table D.3: Simulation results for systems with temperature $T^* = 4.0$ and density $\rho^* = 0.8$. The uncertainty in the last digit, according to the 95% confidence interval, is shown in parentheses.

μ^{*2}	q^{*2}	x^{ions}	a^{dd}	$a^{\text{cc+cd}}$	ε_{r}
0.50	15	0.005	-0.2089(2)	-0.0412(2)	3.58(6)
0.50	15	0.010	-0.2072(2)	-0.0842(4)	3.59(6)
0.50	15	0.030	-0.1996(2)	-0.2607(5)	3.46(5)

Continued on next page

Continued from previous page

μ^{*2}	q^{*2}	x^{ions}	a^{dd}	$a^{\text{cc+cd}}$	ε_{r}
0.50	15	0.050	-0.1924(2)	-0.4415(8)	3.35(6)
0.50	25	0.005	-0.2089(2)	-0.0706(4)	3.57(6)
0.50	25	0.010	-0.2072(2)	-0.1446(5)	3.56(6)
0.50	25	0.030	-0.1996(2)	-0.453(1)	3.43(6)
0.50	25	0.050	-0.1924(2)	-0.769(1)	3.38(5)
1.00	25	0.005	-0.6943(3)	-0.0833(4)	8.20(16)
1.00	25	0.010	-0.6885(3)	-0.1665(5)	8.00(15)
1.00	25	0.030	-0.6655(3)	-0.505(1)	7.70(15)
1.00	25	0.050	-0.6430(3)	-0.847(1)	7.20(14)
1.25	25	0.005	-1.0031(3)	-0.0869(5)	11.5(2)
1.25	25	0.010	-0.9949(4)	-0.1735(6)	11.1(2)
1.25	25	0.030	-0.9625(3)	-0.524(1)	10.1(2)
1.25	25	0.050	-0.9310(3)	-0.876(1)	9.80(23)
1.25	50	0.005	-1.0031(3)	-0.175(1)	11.3(4)
1.25	50	0.010	-0.9949(4)	-0.355(2)	11.1(3)
1.25	50	0.030	-0.9625(3)	-1.074(3)	10.4(2)
1.25	50	0.050	-0.9310(3)	-1.802(2)	10.1(2)
2.00	25	0.005	-2.1177(5)	-0.0932(6)	25(2)
2.00	25	0.010	-2.1018(5)	-0.1863(8)	25(2)
2.00	25	0.030	-2.0391(5)	-0.557(1)	23(2)
2.00	25	0.050	-1.9771(5)	-0.928(2)	21(1)
2.00	50	0.001	-1.9304(5)	-0.0378(4)	26(2)
2.00	50	0.002	-1.9273(4)	-0.0747(8)	26(2)
2.00	50	0.003	-1.9247(4)	-0.1136(10)	26(3)
2.00	50	0.004	-1.9212(4)	-0.1513(9)	26(2)
2.00	50	0.005	-2.1177(5)	-0.1883(9)	25(2)
2.00	50	0.010	-2.1018(5)	-0.375(2)	24(2)
2.00	50	0.030	-2.0391(5)	-1.132(3)	21(1)
2.00	50	0.050	-1.9771(5)	-1.891(4)	20(2)
4.00	25	0.005	-6.181(1)	-0.0942(7)	140(50)
4.00	25	0.010	-6.144(1)	-0.1873(9)	89(36)
4.00	25	0.030	-5.9890(9)	-0.565(2)	148(33)
4.00	25	0.050	-5.835(1)	-0.938(3)	103(38)

Continued on next page

Continued from previous page

μ^{*2}	q^{*2}	x^{ions}	a^{dd}	$a^{\text{cc+cd}}$	ε_{r}
4.00	50	0.005	-6.181(1)	-0.196(2)	139(18)
4.00	50	0.010	-6.144(1)	-0.394(2)	174(47)
4.00	50	0.030	-5.9890(9)	-1.180(3)	105(22)
4.00	50	0.050	-5.835(1)	-1.967(5)	108(25)
4.00	100	0.005	-6.181(1)	-0.417(2)	140(56)
4.00	100	0.010	-6.144(1)	-0.839(3)	135(39)
4.00	100	0.030	-5.9890(9)	-2.510(5)	88(44)
4.00	100	0.050	-5.835(1)	-4.160(9)	66(19)