Combinatorial and dispersion activity coefficient models for molecular solutions

Gerard J.P. Krooshof

Combinatorial and dispersion activity coefficient models for molecular solutions

PROEFSCHRIFT

door

Gerardus Johannes Paulus Krooshof

geboren te Utrecht

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt

voorzitter:	prof.dr.ir. E.J.M. Hensen
1e promotor:	prof.dr.ir. R. Tuinier
2e promotor	prof.dr. G. de With
leden:	prof.dr. C. Storm
	prof.dr.ir. J.A.M. Kuipers
	prof.dr.ir. C.F.J den Doelder PDEng
	prof.dr.ir. T.J.H. Vlugt (Technische Universiteit Delft)
	prof.dr. D.N. Theodorou (National Technical University of Athens)
reserve:	prof.dr. J. Meuldijk

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Eindhoven, 17 December 2019

Summary

The objective of this thesis is to improve the robustness of activity coefficient models (ACMs), so that the prediction of the phase behavior of molecules in multicomponent molecular liquids and related physical properties become more accurate and reliable. This addresses the industrial need to predict phase behavior in a robust way and to make it for physical-chemists possible to design molecules with a specific property or for a particular purpose.

The current and frequently applied predictive ACMs, versions of UNIFAC and COSMO-RS, have three unsolved or questionable issues:

- the use of an ad-hoc adjusted combinatorial contribution,

- the neglect of a contribution for interaction by dispersive forces, and

- the loss of molecular structural information by cutting molecules into atom groups (UNIFAC) or by collecting surface patches into histograms (COSMO-RS).

In this thesis these issues are examined. New activity coefficient equations are developed, explored and tested. It starts with the theory on the combinatorial contribution of the activity coefficient models, where it is shown that also an off-lattice combinatorial activity coefficient can be obtained from Guggenheim's theory. This off-lattice equation in combination with the residual term of the COSMOSPACE model gives better results for the description of phase equilibria behavior of alcohol-alkane mixtures, than previous models did.

Subsequently, it is shown that all cubic equations of state (cEoS) yield a van Laar type of ACM and that these models always overpredict the experimental activity coefficients. This systematic error is a characteristic for all cEoS-based ACMs. It is caused by the assumption introduced by van der Waals that all molecules are spherical, while most molecules are polyatomic, and thereby non-spherical, which reduces the covolume and intermolecular interaction.

To account for this reduced interaction by molecular form and structure, the perturbed hard-sphere chain model PC-SAFT equation of state is taken as starting point. However, the 1st perturbation integral of this model is not parameterized in the limit of random close-packed entities. Therefore, in order to define an effective number of nearest neighbors, topology theory is applied. This approach leads to a topology number that quantifies the dispersion interaction and to an expression for the dispersion activity coefficient.

The combination of the generalized combinatorial contribution and this perturbed and topology based dispersion contribution gives a very good correlation for the limiting activity coefficients of alkane mixtures between model and experiment. This method is extended to a group contribution activity coefficient model, which is advantageous to earlier group contribution methods, because it keeps the structural information of the molecules intact. The current model set-up is envisaged as a sound basis to construct a more robust version of COSMO-RS and UNIFAC.

Preface

This thesis is a result of having the opportunity to investigate more deeply the fundamentals of predictive activity coefficient models (ACMs). For many years I applied these models to estimate phase equilibria, in particular the solubility of solutes in solvents and polymers in the development of new products and processes at DSM. The motivation to improve the basis of ACMs lies in the fact that the group-contribution method UNIFAC and the a priori method of COSMO-RS [116] are failing in some cases. The cause for these failures, are:

- the use of a modified combinatorial term with no physical basis,

- the omission of an explicit dispersion contribution, and

- the loss of 3D features and information by splitting molecules into atom groups (UNIFAC) or surface patches (COSMO-RS).

The initial steps to improve the aforementioned weaknesses, prior to this thesis, were made in an ISPT project in the period 2011-2015. It was a collaboration between TU Delft (prof. Thijs Vlugt, Erin McGarrity, Juan Guttierez-Sevilla), University of Utrecht (prof. Jan van der Eerden) and the companies Albemarle (Peter Daudey), MSD (Rob Geertman), SCM (Stan van Gisbergen) and DSM (Gerard Kwant and me). It resulted in a paper [82], on which chapter 8 is based.

On the other hand, I had many discussions with prof. Jürgen Gmehling, prof. Jürgen Rarey, Dr. Dana Constantinescu and members of the UNIFAC consortium, on how to improve UNIFAC and how to combine the COSMO-RS model with the UNIFAC model to fill the gaps of the interaction matrix. In 2011 I proposed 10 ways to improve UNIFAC [128]. For the fruitful discussions and collaboration, I would like to thank the ISPT team and the UNIFAC consortium members for these interactions, which planted seeds, ideas, in my mind that led to the decision to carrying out a 3 year project with this thesis as result.

I am infinite grateful to Johan Hoorn, who gave me the opportunity to carry out

Contents

List of Figures

List of Tables

Detailed summary

What can you expect to read in this thesis?

This thesis is organized as follows. In chapter 1 we start with an overview of activity coefficient models. We explain the criteria to choose an activity model. An historical overview highlights developments in the last century with respect to phase equilibria description and prediction. An elaborate example is given on how an activity model fits into the modeling of self-assembling molecules. Subsequently, we give in chapter 2 the most important equations related to activity coefficient equations and discuss the different contributions that have been introduced to match the activity coefficient model with experimental results. The first two chapters can be form the basis for the chapters that follow. Chapters 3 and 4 are closely connected and discuss the way the number of nearest neighbors are computed and how this can be used to obtain a generalized expression for the combinatorial activity coefficient. This result is applied in the UNIQUAC and COSMOSPACE activity coefficient models. Chapters 5 - 7 demonstrate how to derive the dispersive activity coefficient from an equation of state and how this can be improved further. Chapter 8 shows how 3D information can be used to improve the COSMO-RS family of coefficient models. Finally, in chapter 9, we summarize the main conclusions made in this thesis and give an outlook to further research.

Outlook of chapters

Chapter 1

Here we give a general overview of activity coefficient models, explain the criteria to choose an activity model, and show how these models are related to phase equilibria description. In an overview we touch upon the key concepts that were introduced to improve the models in time. We end this chapter by showing which physical property methods other than phase equilibria are closely linked to a particular activity model.

Chapter 2

In this chapter we give the basic equations that are used in the field of phase equilibria descriptions. We explain the different contributions in activity coefficient models. The combinatorial and the free-volume contribution of activity models are discussed in detail. There is a variety of expressions for these available in literature, though never well scrutinized on consistency and interpreted from a physics point of view. We show in some cases the errors that where made in the derivation. This chapter also contains a number of basic equations that are frequently used in the subsequent chapters of this thesis.

Chapter 3

This chapter concerns the proof that the Guggenheim model can be improved and leads to an off-lattice model. Guggenheim proposed a theoretical expression for the combinatorial entropy of mixing of unequal sized and linear and branched molecules to improve the Flory-Huggins model. Later the combinatorial activity coefficient equation, which was derived from Guggenheim's model, was applied in the UNI-QUAC, UNIFAC, and COSMOSAC models. Here we derive from Guggenheim's entropy expression a new function for the number of nearest neighbors of a compound in a multicomponent mixture for which the knowledge of the coordination number and a reference area are not needed. The obtained relation requires only the mole, volume and surface fraction of the compounds in the mixture. The benefit of the new relation is that both the combinatorial and the residual term in the aforementioned models can be made lattice-independent. We demonstrate that the proposed relation simplifies the Staverman-Guggenheim combinatorial model.

Chapter 4

We show that the generalized expression for the Guggenheim-Staverman model can be applied with success to the UNIQUAC and COSMOSPACE model in the description of vapor-liquid phase equilibria and excess enthalpy. We also demonstrate that the expression for the number of nearest neighbors, derived in Chapter 3, can be used to replace the relative surface area and the number of surface patches in the residual part of the UNIQUAC and the COSMOSPACE model, respectively. As a result a more rigorous version of the UNIQUAC and the COSMOSPACE model is obtained. This could serve as a better basis for predictive models like UNIFAC, COSMO-RS and COSMOSAC.

Chapter 5

In this chapter an explicit expression for dispersion in activity coefficient models is derived from cubic equations of state (cEoSs). It is demonstrated that all the cEoSs deliver a van Laar type of equation. The difference between these equations can be characterized by a single parameter K, which can be computed directly from the cEoS characteristic parameters. Because the theoretical value for K is always higher than 9, the activity models derived from a cEoS yield always an activity coefficient value, which is higher than the experimental value. It is shown that that mixtures of linear and branched alkanes require $K = 4.13$ and $K = 3.04$, respectively. This mismatch in parameter K is caused by the assumptions, which are made in the derivation of the van der Waals equation of state and which remain present in later developed cEoSs. One of these is that all molecules are spherical, which leads to the inconsistency that the ratio of the covolume and the van der Waals volume is always 4, while this ratio for linear alkanes decreases rapidly to nearly 2 with increasing chain length. Another assumption is that all molecules experience the same number of external interactions, which neglects the fact that polyatomic molecules have less intermolecular interactions per spherical segment due to the presence of covalent bonds and the occurrence of intramolecular interaction. Therefore, the van Laar type of activity coefficient equations are limited in their use as predictive model for dispersion.

Chapter 6

Based on the mismatch in K revealed in chapter 5, which is the result of assuming all molecules in a mixture to be spherical, we derive in this chapter an activity coefficient model from the perturbed hard-sphere chain (PHSC) equation of state. We apply the method of Huron-Vidal, which implies that the system is set to infinite pressure, which place the molecules into close contact. We show that at this condition the second perturbation integral of the PHSC models vanishes and only the first perturbation integral survives. Because the power series expression of the first perturbation integral of the PC-SAFT model fails at random close-packed (RCP) condition, we derive, using the theoretical framework of PC-SAFT, another expression for the perturbation integral at RCP. Topology theory is applied to get realistic values for this integral. The obtained equation in combination with a generalized expression for the combinatorial activity coefficient of chapter 4 gives an excellent description of activity coefficients of alkanes mixtures. The accuracy is of the level of the UNIFAC(Do) model, which is regarded as the best predictive model for phase equilibria description. These results also demonstrate that the systematic deviations of the cubic equation of state based activity model are a result of neglecting the

shape and polyatomic character of molecules.

Chapter 7

In this chapter we explore the model as was proposed in chapter 6. The PC-SAFT model, and thereby the derived expression of the dispersion contribution of the activity model, applies a first order approach to quantify the total interaction energy of the molecules. As is known from the COSMOSPACE model this approach ignores that strong interaction are more persistent than weaker ones. Therefore, by self-consistency, interactions require a Boltzmann weight, as has been done in the COSMOSPACE model. We apply this model to quantify the interactions between different methyl groups. It leads to a further, though small, improvement. The obtained relative interaction energies between different alkyl groups are in qualitative agreement with the expected theoretical results. It indicates that there is still room for further improvement.

Chapter 8

This chapter concerns the improvement of the COSMO-SAC model. We improve the COSMO-SAC model for the calculation of activity coefficients by incorporating 3-dimensional geometric information of molecules. This information is added to the model by means of the effective contact area of the molecules. We define a procedure to compute this contact area by using a probing sphere. The probing sphere rolls around the COSMO surfaces of two contacting atoms to define a spherical cap on each atom. The segments of the COSMO surfaces within the spherical caps are marked and the summation of their areas defines the contact area for a pair of contacting molecules. The effective contact areas are used to compute the σ-profiles and the Onsager screening energies in the calculation of the surface activity coefficients, allowing us to remove one parameter from the COSMO-SAC model. The σ-profiles and effective areas need to be calculated only once and can be stored in a database. We show that the new model, named COSMO-3D because some 3 dimensional information (the position of surface segments and atom centers in 3D-space) is taken from the COSMO-output files, agrees better with experimental data than the COSMO-SAC implementation for non-polar and polar mixtures, even though it uses one parameter less. We feel that the new concept of system dependent effective contact area is a suitable starting point to improve COSMO-SAC further.

Chapter 9

Here we give a summary of the main conclusions of this thesis and put some ideas on paper for future research.

Chapter 1

Activity Coefficient Models

1.1   Choice of an Activity coefficient model

For the description and prediction of phase equilibria scientists and engineers can choose from a variety of equations of state (EoSs) and activity coefficient models (ACMs). The choice for a particular ACM or EoS model to describe the chemical potential of a component in the liquid phase depends mainly on the density of the liquid state, which is related to the temperature and pressure of the system, and the type of molecules in the mixture. Table 1.1 explains schematically when to choose for an activity model to describe the chemical potential of a liquid.

Liquid phase	Regular molecules	Complex molecules
High density	ACM	GC-ACM
Low density	EoS + classical mixing rules	EoS + ACM mixing rule

Table 1.1: Model choice to calculate the chemical potential of compounds in liquid phase.

In general, when the temperature of the system is below the critical temperatures of each of the pure components, the density of the mixture is high. The system is in the condensed state and not expanded by a supercritical compound. In this case an activity coefficient model would be a logical choice, because an ACM interpolates between the liquid states of the pure compounds, while an EoS extrapolates from the ideal gas state to the liquid state. Under such conditions an ACM provides a more accurate prediction of the chemical potential than an EoS model. An example of a frequently applied descriptive ACM is the Non-Random Two-Liquid (NRTL) model of Renon and Prausnitz [186]. It is mainly used for regular molecules. However, for 'complex' molecules, i.e. molecules containing one or more functional atom groups, a predictive group contribution (GC) based ACM is frequently applied, because

experimental data is in most cases lacking. Examples are the UNIFAC [59] and COSMO-RS [114] models.

Oppositely, when the liquid phase contains a major compound, which has a critical temperature that lies below the system temperature, then this compound expands the liquid phase thereby reducing its density. In that case an equation of state would be a better choice, because the interaction between the molecules is reduced by the increase in free-volume. This free-volume effect can be incorporated in an ACM, but this is not preferable, since volumetric changes are better described by an EoS than by an ACM. For systems with regular molecules or molecules of one class it is sufficient to use the classical mixing rule in the EoS. The Peng-Robinson [177] or the Soave-Redlich-Kwong [206] cubic EoS are used frequently for this purpose. Today the use of the PC-SAFT [78] EoS is more common. For liquid mixtures with a variety of complex molecules, the classical mixing rules are not sufficient to describe a phase equilibrium. In that case the classical mixing rules is replaced by activity coefficient based mixing rule. The predictive Soave-Redlich-Kwong (PSRK) model is an example of such an approach [91]. We mention that in the modeling of gasses dissolved in a condensed liquid, where the concentration of the gas is below 1 mol%, which implies that the liquid is still at high density, it is custom to use an ACM for the condensables and a Henry equation to quantify the dissolved gas by means of the partial pressure of the gas.

For the gas phase the chemical potential is either described by the ideal gas (IG) equation, a virial expansion equation, or an equation of state (EoS). The choice to use an EoS is made on basis of the system pressure and the presence of atom groups that cause association in the gas phase. As a rule of thumb the choice between an equation of state and ideal gas is on whether the pressure is above or below 10 bar, respectively. More precisely, the choice for an EoS is made, when the fugacity coefficient of one of the components in the gas phase is deviating more that 10% from unity, which happens to be around 10 bar. Virial expansion equations are also applied, but with the introduction of advanced EoS models, such as PSRK and PC-SAFT, the use of it has diminished. In fact, current chemical engineering flow-sheet programs, such as AspenPlus, do not have this option in the basis set of models. The use of a dimerization or strong association model for the gas phase is based on the presence of molecules that have one or more acid groups. For low pressure the fugacity of the gas phase is mainly calculated by chemical theory described by the model of Marek [202]. At high pressure this is done by extending the EoS with the 1A association scheme of Wertheim [232].

Gas phase	No dimerization	Dimerization
$P < 10$ bar	IG	IG with Chemical theory
$P > 10$ bar	EoS	EoS with 1A association scheme

Table 1.2: Model choice to calculate the chemical potential of compounds in gas phase.

1.2 Chemical potential, fugacity, and activity

The choice of an equation of state or activity coefficient model determines the way the chemical potential of a compound is described in a fluid phase. At equilibrium the chemical potential of a compound is the same in all phases. For instance, at fixed pressure and temperature, the chemical potential of compound j in a multicomponent mixture in a vapor-liquid-liquid equilibrium (VLLE) can be expressed as:

$$\mu_j^{\text{gas}} = \mu_j^{\text{liq}_1} = \mu_j^{\text{liq}_2}, \tag{1.1}$$

where the superscript indicates the phase state (gas phase, liquid phase 1, liquid phase 2).

The chemical potential of compound j in a mixture can be written in terms of the fugacity, f_j by:

$$\mu_j^{\text{mix}}(T) = \mu_j^0(T) + RT \ln(f_j(T)), \tag{1.2}$$

where R is the gas constant, T the absolute temperature, and μ_j^0 the chemical potential of the compound in pure liquid or saturated gas state.

Substitution into Eq. 1.1 yields:

$$\mu_j^0 + RT \ln(f_j^{\text{gas}}) = \mu_j^0 + RT \ln(f_j^{\text{liq}_1}) = \mu_j^0 + RT \ln(f_j^{\text{liq}_2}), \tag{1.3}$$

so

$$f_j^{\text{gas}} = f_j^{\text{liq}_1} = f_j^{\text{liq}_2}. \tag{1.4}$$

It follows that the criterion of isofugacity is equivalent to the definition of an equilibrium in terms of the chemical potential.

The fugacity is commonly calculated in two ways, sometimes called the gamma-phi and phi-phi route. The first refers to an equation using activity coefficients (γ), the second to an equation using fugacity coefficients (ϕ) only. These routes can be expressed by the following two equations for compound j between a liquid phase and a gas phase:

$$\text{gamma-phi route}: x_j \gamma_j f_j^0 = y_i \phi_j^{\text{gas}} P, \tag{1.5}$$

and

$$\text{phi-phi route}: x_j \phi_j^{\text{liq}} P = y_i \phi_j^{\text{gas}} P \Leftrightarrow x_j \phi_j^{\text{liq}} = y_i \phi_j^{\text{gas}}, \tag{1.6}$$

where x_j and y_j are the mole fractions of compound j in the liquid and gas phase, f_j^0 is the fugacity of pure compound j at system temperature T, ϕ_j^{liq} and ϕ_j^{gas} are the fugacity coefficients of compound j in the liquid and gas phase, respectively, and γ_j is the activity coefficient of the compound in the liquid at temperature T and pressure P. For pressures over 10 bar the calculation of the fugacity coefficient always requires an equation of state or a virial coefficient equation. For lower pressures both equations approximate the ideal gas equation of state, which implies that the fugacity coefficient of the gas phase is set to unity and that any deviation of the gas phase is lumped into the activity coefficient of the liquid phase. Since the fugacity of a pure compound is at low pressures equal to its saturation pressure, P_j^{sat}, it is custom to approximate the gamma-phi route for a vapor-liquid equilibrium (VLE) at low pressures by:

$$x_j \gamma_j P_j^{\text{sat}} = y_i P \tag{1.7}$$

For a liquid-liquid equilibrium (LLE), regardless of the pressure, we get:

$$x_j^{\text{liq}_1} \gamma_j^{\text{liq}_1} f_j^0 = x_j^{\text{liq}_2} \gamma_j^{\text{liq}_2} f_j^0, \tag{1.8}$$

which leads to the equation:

$$x_j^{\text{liq}_1} \gamma_j^{\text{liq}_1} = x_j^{\text{liq}_2} \gamma_j^{\text{liq}_2} \tag{1.9}$$

This equation shows that for LLE isoactivity rules the equilibrium.

We mention that the interaction parameters of an activity model, which are obtained from fitting experimental VLE data, are not applicable to describe LLE behavior. The reason is that in the fitting of experimental VLE data one uses the approximation that the gas fugacity coefficient is unity (Eq. 1.7), while this is not the case for LLE. In other words, small non-ideal behavior in the gas phase is transferred by the fitting procedure into the interaction parameters of the activity coefficient model. Since isoactivity of LLE is very sensitive to small differences, the resulting equilibrium concentrations between two liquid phases can differ a great deal. From a mathematical point of view it is better to start by fitting the parameters using experimental LLE data only, and subsequently fit experimental VLE data by using an equation of state to account for small non-ideal behavior in the gas phase. This

applies especially when the interaction between two compounds is strongly repulsive or attractive.

The activity coefficient γ_j of Eqs. 1.7 and 1.9 is calculated by an activity coefficient model (ACM). In section 1.3 we give a short overview of the most important ACMs. Commonly the activity coefficient is sequestered into an entropic (S) and an enthalpic (H) contribution. There are of course exceptions, such as the Non-Random Two-Liquid (NRTL) model [186], where such a differentiation is not made. The entropy and enthalpy parts are frequently denoted as combinatorial (comb) and residual (res) contributions. The combinatorial part of the activity coefficeint and is obtained from the ACM equation by setting the temperature to infinite value. This means that the athermal part of the ACM is seen as the combinatorial part. The free-volume contribution (fv), when included in the ACM, is an enthalpy contribution, because it changes with temperature. Other temperature dependent contributions are the dispersion (disp), permanent dipole-dipole or quadrupole-dipole interaction (pol), association and hydrogen bonding, denoted as acid-base interaction (AB). This gives the following forms for the total activity coefficient of non-electrolyte systems:

$$
\begin{aligned}
\ln \gamma_{\text{tot}} &= \ln \gamma_{\text{S}} + \ln \gamma_{\text{H}} \\
&= \ln \gamma_{\text{comb}} + \ln \gamma_{\text{res}} \\
&= \ln \gamma_{\text{comb}} + \ln \gamma_{\text{fv}} + \ln \gamma_{\text{disp}} + \ln \gamma_{\text{pol}} + \ln \gamma_{\text{AB}},
\end{aligned}
\tag{1.10}
$$

which can also been written as:

$$
\gamma_{\text{tot}} = \gamma_{\text{comb}} \gamma_{\text{fv}} \gamma_{\text{disp}} \gamma_{\text{pol}} \gamma_{\text{AB}}
\tag{1.11}
$$

In the next section we give a short overview of available models and the developments that have been made over the years.

1.3 Overview of activity coefficient models

There is a variety of descriptive and predictive activity coefficient models (ACMs). Table 1.3 shows a chronological list of the most common ACMs with their key feature that was introduced to improve the ACM in comparison to early models. The model of Margules is noted as the first activity model, though Margules is not the inventor of the concept of fugacity and activity. These concepts were introduced by Lewis in 1901 [137] and 1907 [138]. The Margules model is mentioned here, because it frequently gives a better description of activity coefficients than other models. Evidence for this can be found in the DECHEMA [68] data series on vapor-liquid equilibria, where model parameters for the Margules and other ACMs are given, together with

the absolute average deviation between model and experimental data.

First author	ACM name	Key feature	Year	Ref.
Margules		first known 'activity' model	1895	[160]
van Laar	-	derived from van der Waals' EoS	1910	[223]
Fowler	-	size effect of dimers in a lattice	1937	[57]
Huggins	-	solvent-polymer Lattice model	1941	[95]
Flory	-	solvent-polymer Lattice model	1942	[53]
Hildebrand	Regular solution	Solubility parameter	1947	[87]
Staverman	-	modification of combinatorial term	1950	[209]
Guggenheim	-	modification of combinatorial term	1952	[81]
Guggenheim	-	(binary) quasi-chemical theory	1952	[81]
Wilson	-	local composition for internal energy	1964	[234]
Renon [a]	NRTL	local composition for Gibbs energy	1968	[186]
Hansen	-	set of 3 solubility parameters	1967	[84]
Derr	ASOG	group contribution model	1971	[39]
Donohue[a]	-	fractal volume term	1975	[43]
Abrams[a]	UNIQUAC	(simplified) quasi-chemical theory	1975	[2]
Fredenslund [a]	UNIFAC	group contribution form of UNIQUAC	1975	[59]
Thomas	MOSCED	split of interaction types	1984	[214]
Huyskens	-	mobile order theory	1985	[98]
Aranovich	AD	local surface fraction composition	1996	[5]
Klamt	COSMO-RS	ab initio model	1995	[114]
Lin	COSMO-SAC	COSMO-RS like but other averaging	2002	[144]
Klamt	COSMOSPACE	descriptive self-consistent field model	2002	[117]
Bronneberg	MOQUAC	orientation effect in GEQUAC	2013	[22]
Krooshof	COSMO-3D	orientation effect in COSMO-RS	2015	[82]

[a] Co-author Prausnitz.

Table 1.3: Chronological overview of frequent applied activity models.

Deviations of entropy with respect to the ideal entropy, were first investigated by Fowler and Rushbrooke [57]. They investigated the size effect of molecules by calculating the entropy of mixtures of monomers and dimers. Huggins [95], and later Flory [53] extended this approach to polymers, which leads to the conclusion that instead of mole, volume fractions are needed to quantify entropy differences, as was already indicated by Hildebrand [87]. The Flory-Huggins (FH) combinatorial expression, however, turned out to give too strong negative deviation from ideal solubility for the mixtures of alkanes. Staverman and Guggenheim (SG) tried to solve this by correcting the FH equation with a term that accounts for the probability to place subsequent segments of one molecule in a lattice. It was an adjustment in the right

direction, but not sufficient for alkane mixtures. In Chapter 3 we will take a closer look at the FH-SG model. Hansen [84] modified the Flory-Huggins model by replacing the the empirical Flory-Huggins χ-parameters by an expression involving the so-called solubility parameters, which are derived from pure component properties, i.e. the enthalpy of evaporation and liquid density. In some cases these parameters are tuned further to provide an accurate phase equilibrium prediction. The handbook of solubility parameters [10] lists these parameters for a wide range of molecules.

Meanwhile, Wilson [234], Renon [186], Abrams [2] and Aranovich [5] considered that, due to interaction, the molecules in a mixture are not randomly distributed. Molecules of compound 1 cluster differently around a molecule of type 2, than molecules of compound 2 cluster around molecule 1. This approach yields a certain degree of non-randomness. Guggenheim used a quasi-chemical bond to quantify the non-randomness and to arrive at expressions for the Gibbs energy of mixing for binary systems [81]. However, Guggenheim's equations become complicated for multicomponent mixtures, and require extensive computation when there are different types of quasi-bonds between molecules. Therefore, Wilson and followers worked on less complicated Gibbs energy equations to quantify the non-randomness and derived simpler ACMs. Later the non-randomness concept was also applied in the development of the predictive models ASOG [39] and UNIFAC [59]. Wu *et al.* [236] summarize the differences between several local composition models. In short, one could conclude that that Guggenheim's quasi-chemical theory [81] and the generalization of it in the COSMOSPACE model are the most accurate models to describe non-randomness, followed by the Aranovich-Donohue model, the UNIQUAC, the NRTL and the Wilson model, respectively. Worthwhile to mention is that the Flory-Huggins model as well as the Hansen method assume mixtures where molecules are randomly distributed. This simplification makes the interaction parameter concentration dependent.

The ASOG [39] and UNIFAC [59] methodologies are predictive models that divide molecules in groups. These approaches are more versatile than the method of Hansen, because it enables scientists and engineers to predict the phase behavior of molecules for which no physical data is available.

The original UNIFAC group contribution method gives large deviations from Raoult's law for asymmetric alkane mixtures, due to the absence of a dispersion contribution. In the UNIFAC methodologies the interaction energy between CH_3, CH_2, CH and C is assumed to be zero. Instead of assigning interaction energies between these subgroups to repair this omission, the developers of the modified UNIFAC models applied the ad-hoc proposal of Donohue and Prausnitz [43]. In here the volume fraction is raised to a power between 0 and 1 (see 2.3.3). By doing this they placed an enthalpic contribution, i.e. dispersion, into an entropic contribution,

i.e. the combinatorial part. This approach has lead to a series of modified versions of UNIFAC, of which the Dortmund [231] and Lyngby [133] approaches survived, while others [112, 47] are not used anymore.

The omission or neglect of a dispersion term, as done in the UNIFAC models, is also a problem in the COSMO-RS and COSMO-SAC models [114, 94, 142, 141, 229, 239, 82]. Improvements were made, but still the modified UNIFAC(Do) model gives a better prediction than the COSMO-RS family of models does. This is because a systematic deviations arising from the combinatorial term can be repaired in the UNIFAC model by the adjustment of the interaction energy parameters or the subgroup area parameter. This option of tuning the model is not as such available in QM-based models. The COSMO-RS family of models has only a limited number of universal parameters (<10), while that for the UNIFAC is large (>100). A way to overcome this problem, is to define in the COSMO-RS type of models specific classes of molecules, for which the universal parameters are optimized. But such an approach is not only elaborate, it also deteriorates the robustness of the COSMO-RS models. In order to improve UNIFAC and COSMO-RS type of models it is important to investigate what the most accurate model is for the combinatorial activity coefficient and how the dispersion can be formulated in an activity coefficient contribution.

1.4 The importance of activity models in applications

Activity models are mainly applied in the description and prediction of phase equilibria like VLE, LLE, SLE (solid-liquid equilibria), and combinations of these equilibria. In chemical processes it is crucial to quantify the phase behavior and stability of it, because these help the chemical engineer to understand how to optimize a process step and can alert him on show-stoppers that might happen in production. An example of a show-stopper is the oiling-out of pharmaceuticals, where one of the demixing liquid phases forms an organic phase prior to the crystallizing of an active pharmaceutical ingredient. The organic phase will form an oily layer on the crystals, which should be avoided. The use of a heterogeneous azeotrope in distillation is a way by which the separation of a low volatile from a solution or substrate can be improved. For instance, in the low volatility of essential oil in rose petals is enhanced by adding water (i.e. steam distillation). The oversaturation of a dissolved compound in a (polymeric) matrix is another example that can occur during product development. In this case a dissolved additive will diffuse out of the polymer matrix when the activity of the solute is higher than the equilibrium activity at room temperature. Such a phenomenon happens when producers dissolve an additive at process temperature, which has a higher concentration than matrix can contain at

room temperature. As a result, a softener in a plastic will eventually migrate to the surface and make the product more tacky in time.

Besides the description and prediction of phase equilibria, the ACMs are also applied in the calculation of other physical properties. Table 1.4 gives an overview of ACMs and binary interaction parameters (BIPs), which are applied into another physical property model or method.

Property/Theory	ACM	BIP	Lit.
Flash point of mixtures	UNIFAC	-	[69]
Viscosity of mixtures	UNIQUAC	-	[25]
Surface tension	NRTL	-	[208]
Thermal conductivity of mixtures	NRTL	-	[191]
Solvent diffusion in polymers	-	χ-parameter	[244]
Self-consistent field theory	-	χ-parameter	[196]
Dissipative particle dynamics	-	χ-parameter	[75]

Table 1.4: Applications of ACMs and BIPs in other physical property calculation and theory

Chapter 2

Gibbs energy of mixing and activity coefficient contributions

2.1 Introduction

In this chapter we start with basic equations related to the Gibbs energy and activity coefficients. Subsequently, we review the entropy part of activity coefficients. We derive expressions for the combinatorial and free-volume activity coefficient models and the modifications which have been proposed. Next, we discuss the enthalpy part, but restrict ourselves to the excess enthalpy caused by dispersive interactions.

2.2 Excess Gibbs energy

Non-ideal solvation of compounds is characterized by the excess Gibbs energy, ΔG^{E}, which is the difference between the Gibbs energy of the mixture and that of the pure compounds in reference to an ideal mixture. The excess Gibbs energy is related to the excess enthalpy and entropy, according to the thermodynamic relation:

$$\Delta G^{\mathrm{E}} = \Delta H^{\mathrm{E}} - T\Delta S^{\mathrm{E}}, \tag{2.1}$$

where ΔH^{E} and ΔS^{E} are the excess enthalpy and entropy, respectively. Mixtures with $\Delta G^{\mathrm{E}} = 0$ are called ideal mixtures. Mixtures that have $\Delta H^{\mathrm{E}} = 0$ are called athermal systems, because it shows neither exothermal ($\Delta H^{\mathrm{E}} < 0$), nor endothermal ($\Delta H^{\mathrm{E}} > 0$) behavior. Mixtures that have $\Delta S^{\mathrm{E}} = 0$ are called regular mixtures.

2.2.1 Activity coefficient and the excess Gibbs energy

The changes in these thermodynamic properties are measured in different ways. The change in excess enthalpy is often quantified by calorimetry, while the change in Gibbs free energy can be determined by measuring phase equilibria. This can be performed in various ways, for instance, by measuring the pressure and vapor composition as a function of liquid composition in VLE systems, the mutual solubility (LLE) of demixing liquids, or the solubility of solids in a liquid (SLE). From the obtained mixing enthalpy and Gibbs energy Eq. 2.1 yields the mixing entropy.

The relation between excess Gibbs energy, G^{E}, and the activity coefficient, γ_k, of compound k in a mixtures is given by the thermodynamic equation:

$$
\ln \gamma_k = \left(\frac{\partial \frac{\Delta G^E}{k_{\mathrm{B}}T}}{\partial N_k} \right)_{T,P,N_j \neq N_k}
\tag{2.2}
$$

The right-hand side of the equation can be split into a residual enthalpic part, γ_k^{res}, and a combinatorial entropic part, γ_k^{comb}:

$$
\ln \gamma_k^{\mathrm{res}} = \frac{1}{k_{\mathrm{B}}T} \left(\frac{\Delta H^E}{\partial N_k} \right)_{T,P,N_j \neq N_k}
\tag{2.3}
$$

$$
\ln \gamma_k^{\mathrm{comb}} = -\frac{1}{k_{\mathrm{B}}} \left(\frac{\Delta S^E}{\partial N_k} \right)_{T,P,N_j \neq N_k}
\tag{2.4}
$$

2.2.2 Gibbs-Duhem and consistency of activity coefficient models

The thermodynamic consistency of activity models can always be double checked by using the Gibbs-Duhem relation, which under isothermal and isobaric conditions reduces to:

$$
\sum_j N_j \mathrm{d}\mu_j = -SdT + VdP = 0,
\tag{2.5}
$$

where μ_j is the chemical potential of compound j, and S and V the entropy and volume of the system. The Gibbs-Duhem equation relates the chemical potential of one of the components in a mixture to the other ones. In terms of activity coefficients we have:

$$
\sum_{j=1}^{M} x_j \mathrm{d}\ln\gamma_j = 0.
\tag{2.6}
$$

This equation holds for any contribution of the activity coefficient, whether it is the combinatorial, the residual, or a part of the residual contribution. There are two ways to make use of this expression to test whether experimental data or a model is consistent with the Gibbs-Duhem criterion. These are the differential and the integral test.

Differential test

Differentiation of Eq. 2.6 with respect to x_1 yields for a binary system:

$$x_1 \left(\frac{\partial \ln \gamma_1}{\partial x_1} \right)_{T,P} + x_2 \left(\frac{\partial \ln \gamma_2}{\partial x_1} \right)_{T,P} = 0. \tag{2.7}$$

The equation can be used to check any activity coefficient equation numerically. For simple algebraic equations an analytic proof is possible. As an example we take the symmetrical activity coefficient, known as the Porter model, expression which has the form:

$$\begin{cases} \ln \gamma_1 = A x_2^2 \\ \ln \gamma_2 = A x_1^2 \end{cases} \tag{2.8}$$

Differentiation with respect to x_1 and substitution into Eq. 2.7 yields:

$$x_1 \left(-2A x_2 \right) + x_2 \left(2A x_1 \right) = 0, \tag{2.9}$$

which proofs that the symmetrical activity coefficient equation is thermodynamically consistent. In practice, one notices an inconsistency by plotting the activity coefficient as function of the mole fraction. In a binary system the slopes of the activity coefficients should be of opposite sign. When the slope of γ_1 zero then that of γ_2 has also be zero.

Integral test

The integral form of the Gibbs-Duhem relation for a binary system is given by the equation:

$$\int_0^1 \ln \left[\frac{\gamma_1}{\gamma_2} \right] dx_1 = 0, \tag{2.10}$$

where γ_1 and γ_2 are the activity coefficient of any contribution of components 1 and 2, respectively. Fig. 2.1 depicts the integral form of the Gibbs-Duhem test.

In this particular case it depicts the consistency check for the free-volume contribution of the PC-SAFT based hard-sphere chain model for the system n-hexane - eicosane. Since the areas A and B have equal size, the free-volume model is thermodynamically consistent.
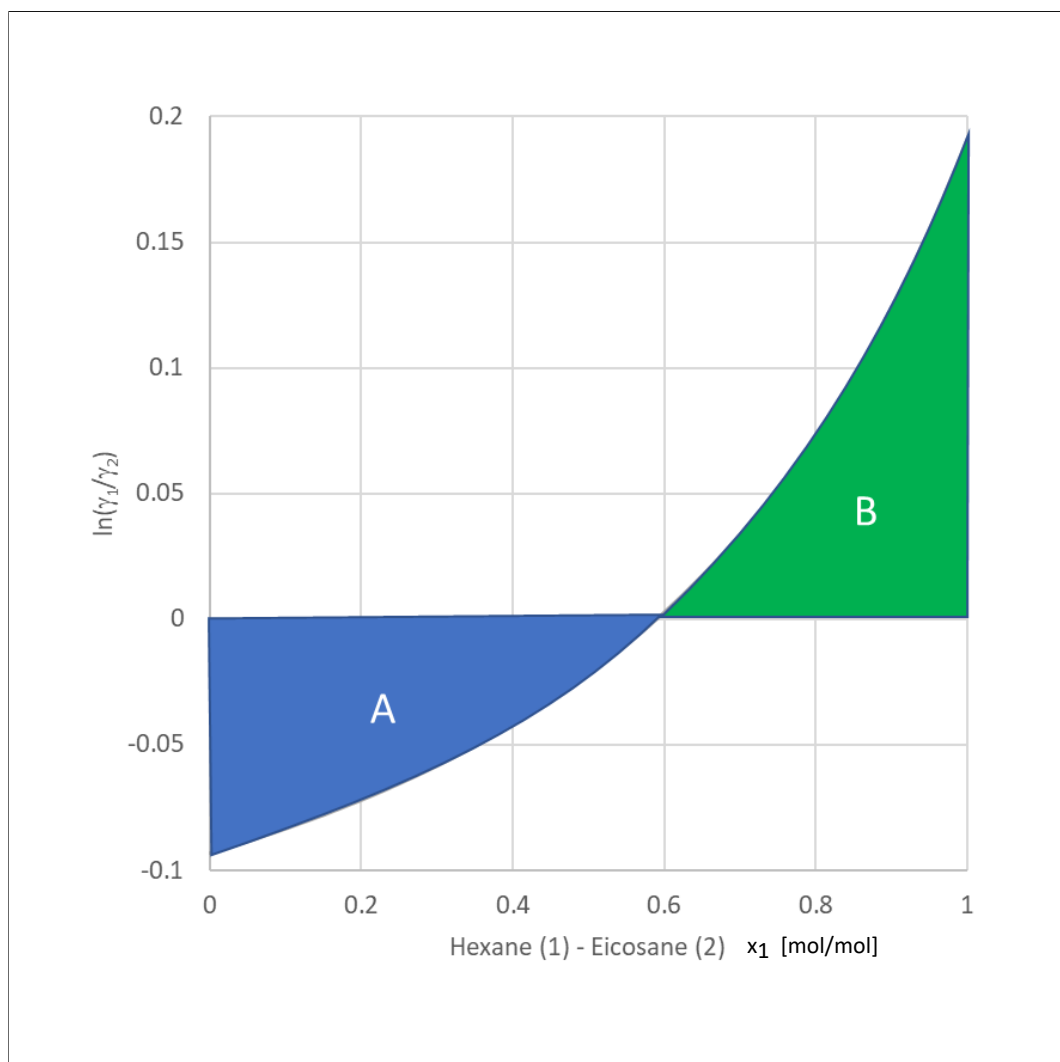
Figure 2.1: Gibbs-Duhem test (Eq. 2.10) applied to the system n-hexane-eicosane. A/B-1<0.5%.

It is worthwhile to mention that the early version of COSMO-RS was not thermodynamically consistent. An ad-hoc expression for the combinatorial term was applied. I informed Klamt about this inconsistency in 2001. In later versions of COSMO-RS this inconsistency was removed [115].

2.3 Entropy and the combinatorial activity coefficient

2.3.1 The entropy of ideal fluids

The entropy of a monoatomic ideal fluid consisting of N_j molecules of type j in an enclosed volume V_j is given by the Sackur-Tetrode [192] equation:

$$\frac{S_j}{k_{\mathrm{B}}} = N_j \ln(V_j) - \ln(N_j!) + \tfrac{5}{2} N_j \ln(2\pi e m_j k_{\mathrm{B}} T) - 3 N_j \ln(h), \qquad (2.11)$$

with k_{B} is the Boltzmann constant, T the absolute temperature, m_j the molar mass of molecules of type j, and h the Planck constant. The interpretation of Eq. 2.11 is that the number of microstates is related to several different contributions related to the position of the molecules, the differences between these molecules in size, shape, translational momentum, and the quantum mechanical state. Ben-Naim [12] explains it in terms of uncertainty. In this view, the first term reflects the uncertainty in position, the second the uncertainty to distinguish differences between atoms of the same type, the third the uncertainty in translational momentum, and the last the quantum mechanical uncertainty principle, where the de Broglie wavelength:

$$\Lambda_j = \sqrt{\frac{k_{\mathrm{B}} T h^2}{2\pi m_j}}, \qquad (2.12)$$

marks the transition from the quantum mechanical to the classical regime. For $\Lambda_j >> V_j / N_j$ the wave functions overlap and the individual character of particles is lost.

When N_1 molecules of type 1, which are initially in volume V_1, are set into communication with another ideal fluid of N_2 particles of type 2 in volume V_2, the final entropy of the mixed fluid is given by:

$$\frac{S_{1+2}}{k_{\mathrm{B}}} = N_1 \ln\left(\frac{V_1 + V_2}{N_1 \Lambda_1^3}\right) + N_2 \ln\left(\frac{V_1 + V_2}{N_2 \Lambda_2^3}\right) + \frac{5}{2}(N_1 + N_2). \qquad (2.13)$$

The difference in entropy with the pure state is given by:

$$\frac{\Delta S^{\mathrm{mix}}}{N_t k_{\mathrm{B}}} = \frac{S_{1+2} - S_1 - S_2}{N_t k_{\mathrm{B}}} = x_1 \ln\left(\frac{V_1 + V_2}{V_1}\right) + x_2 \ln\left(\frac{V_1 + V_2}{V_2}\right). \qquad (2.14)$$

In here N_t is the total number and x_j is the mole fraction of molecules of type j. Generalization to a mixture of M different compounds gives:

$$\frac{\Delta S^{\mathrm{mix}}_{\mathrm{ideal}}}{N_t k_{\mathrm{B}}} = -\sum_{j=1}^{M} x_j \ln \frac{V_j}{\sum\limits_{j=1}^{M} V_j}. \qquad (2.15)$$

Assuming that there exists no difference in the translational and rotational momentum of a particle in the pure and mixed state Eq. 2.15 can be extended to polyatomic molecules. Eq. 2.15 points out that the entropy change upon mixing pure ideal fluids is actually the entropy change of expansion of fluids [12].

Only when the volume is completely described by the ideal gas law, the partial volume can be replaced by mole fractions, since: $V_j = N_j k_B T / P$ or in general V_j/N_j = constant. In that particular case Eq. 2.15 reduces to:

$$\frac{\Delta S_{\text{ideal}}^{\text{mix}}}{N_t k_B} = -\sum_{j=1}^{M} x_j \ln x_j. \tag{2.16}$$

The ideal entropy of mixing, as it is frequently named, is thus actually the entropy of expansion and the resulting equation depends on how this expansion is equated. In the following paragraphs different expressions for the volume and thereby different forms of entropy of expansion will be discussed.

2.3.2 The original Flory-Huggins combinatorial model

The Flory-Huggins (FH) model is one of the most frequently used theoretical concepts in the area of multicomponent liquid mixtures containing polymers. It is a relative simple model in use. The total model consists on an entropy part and an enthalpy part. We focus on the entropy part. The concept of FH theory is based upon describing a fluid as a lattice, in which the space is completely filled by solvent(s) and one or more polymers. This concept of a lattice is, however, not necessary as was shown in the work of Longuet-Higgins [152]. The volume occupied by compound j is defined by the product of a hard-core reference volume or a rigid lattice cell volume, v_0, the number of units, r_j, and the number, N_j, of molecules of type j by:

$$V_j = N_j r_j v_0. \tag{2.17}$$

Substitution of Eq. 2.17 into Eq. 2.15 yields the FH entropy:

$$\frac{\Delta S_{\text{FH}}^{\text{mix}}}{N_t k_B} = -\sum_{j=1}^{M} x_j \ln \phi_j, \tag{2.18}$$

where ϕ_j is the volume fraction of the molecules of type j:

$$\phi_j = \frac{N_j r_j v_0}{\sum\limits_{j=1}^{M} N_j r_j v_0}. \tag{2.19}$$

The excess entropy change for the FH fluid follows from $S^E = S^{mix} - S^{id}$, which Eqs.2.18 and 2.15 yields:

$$\frac{\Delta S_{FH}^{E}}{k_B} = -\sum_{j=1}^{M} N_j \ln \frac{\phi_j}{x_j}. \tag{2.20}$$

For athermal mixtures, where the enthalpy of mixing is zero, $\Delta H^E = 0$, the Gibbs energy of mixing, ΔG^E, reduces to $\Delta G^E = -T\Delta S^E$. Substitution of Eq. 2.20 into Eq. 2.4 then gives the activity coefficient of component k:

$$\ln \gamma_k^{FH} = \ln \frac{\phi_k}{x_k} + 1 - \frac{\phi_k}{x_k} \tag{2.21}$$

As was noted by Nicolas *et al.* [170], and in line with statistical theory of Longuet-Higgins [152], the volume fraction is insensitive to the definition of a reference volume v_0. Key is the ratio in (rigid) volumes between the molecules, which is denoted by the dimensionless parameter r_j in equation 2.17. The assumption is that the mixing of the pure compounds involves no expansion or contractions. From an historical perspective the FH-equation gave an improvement for the entropy of mixtures consisting of molecules of different size, but it was soon recognized that the model gave large deviations from the ideal mixing of pure components. For example, the calculated activity coefficients of compounds in alkane mixtures were much lower than was found by experiment. For that reason scientists started to modify the Flory-Huggins concept.

2.3.3   Modification of the Flory-Huggins model

An ad-hoc modification of the combinatorial contribution of the FH model is the approach of Donohue and Prausnitz [43]. They defined the entropy of mixing as:

$$\frac{\Delta S_{FHD}^{E}}{k_B} = -\sum_{j=1}^{M} N_j \ln \frac{\psi_j}{x_j}, \tag{2.22}$$

where ψ_j is the fractal volume fraction of the molecules of type j defined by:

$$\psi_j = \frac{N_j r_j{}^{a_j}}{\sum\limits_{j=1}^{M} N_j r_j{}^{a_j}}. \tag{2.23}$$

Here the exponent a_j is a parameter that tunes the equation from the ideal equation

($a_j = 0$) to the FH equation ($a_j = 1$). This implies that not the volume but a fractal of the volume is used to define the combinatorial entropy. A value of 2/3 would indicate that the key property is the surface of the molecule, 1/3 the length, and 0 that only the position (e.g. center of mass of the molecule) is important. The last will lead to mole fractions instead of volume fractions, which implies that the mixture behaves like an ideal solution. The expression for the Donohue modified Flory-Huggins (mFH) combinatorial activity coefficient is:

$$\ln \gamma_k^{\mathrm{mFH}} = \ln \frac{\psi_k}{x_k} + 1 - \frac{\psi_k}{x_k}. \tag{2.24}$$

We note that the concept of Donohue and co-workers is in agreement with the Gibbs entropy expression, which is defined as:

$$S = -k_{\mathrm{B}} \sum_j P_j \ln(P_j), \tag{2.25}$$

where the index j denotes all the microstates with energy E_j, and P_j the probability of finding a particular configuration associated with E_j. Garces [63] showed that this can be written as:

$$S = -k_{\mathrm{B}} \sum_j N_j \ln(p_j), \tag{2.26}$$

where N_j is the number of molecules of type j and p_j denotes the probability of finding molecules j in a particular configuration. In terms of the excess entropy we have:

$$\Delta S_{\mathrm{Gibbs}}^{\mathrm{E}} = -k_{\mathrm{B}} \sum_{j=1}^{M} N_j \ln(\frac{p_j}{x_j}). \tag{2.27}$$

This equation shows that for the FH model $p_j = \phi_j$, for the ideal solution model $p_j = x_j$, and for the mFH model $p_j = \psi_j$.

The excess entropy equation, which was introduced by Huyskens [98] is:

$$\frac{\Delta S_{\mathrm{Huy}}^{\mathrm{E}}}{k_{\mathrm{B}}} = -\sum_{j=1}^{M} N_j \ln(\phi_j{}^{\alpha} x_j{}^{-\alpha}), \tag{2.28}$$

with α is the mobile order parameter. Initially, Huyskens set $\alpha = \frac{1}{2}$, because he assumed that this parameter reflects the average path length of a molecule to diffuse from the center to the border of a lattice cell, the place where it can interchange with another molecule. In fact, his view introduces a free-volume concept to the lattice model, since there is space for the molecules to move before they interchange from

lattice cell position. Later, Huyskens [99] proposed that α can be any value between 0 and 1, which implies that he rejected the concept of mobile order and used an engineering approach. However, in view of the aforementioned Gibbs entropy, i.e. Eq. 2.27, the Huyskens model uses a probability, defined by:

$$p_j = \phi_j^\alpha x_j^{1-\alpha} = \phi_j \left(\frac{x_j}{\phi_j} \right)^{1-\alpha}. \tag{2.29}$$

The format at the right-hand side shows that the probability can be seen as a conditional probability. In Bayesian statistics ϕ_j would be the prior probability function, while the term with the exponent is the likelihood function. When we normalize this probability we get the expression

$$p_j = \frac{\phi_j^\alpha x_j^{1-\alpha}}{\sum\limits_{k=1}^{M} \phi_k^\alpha x_k^{1-\alpha}} = \frac{(x_j r_j)^\alpha x_j^{1-\alpha}}{\sum\limits_{k=1}^{M} (x_k r_k)^\alpha x_k^{1-\alpha}} = \frac{x_j (r_j)^\alpha}{\sum\limits_{k=1}^{M} x_k (r_k)^\alpha} = \frac{\psi_j}{\sum\limits_{k=1}^{M} \psi_k} = \psi_j, \tag{2.30}$$

and observe that the normalized Huyskens equation is a particular form of the Donohue equation, in which the fractal $a_j = \alpha$ is the same for all molecules. The activity coefficient of the Huyskens model is found by substitution of Eq. 2.28 into Eq. 2.4, which yields:

$$\ln \gamma_k^{\text{Huy}} = \alpha \left[\ln \frac{\phi_k}{x_k} + 1 - \frac{\phi_k}{x_k} \right]. \tag{2.31}$$

This equation with $\alpha = 0.5$ was applied in a comparative study of Bronneberg *et al.* [22] and in the work of Guttierez *et al.* [82]. These studies demonstrate that Huyskens model with $\alpha = \frac{1}{2}$ gives good results for alkane mixtures. However, both studies neglect the effect of a dispersion contribution. Therefore, we can conclude that the Huyskens model is physically inconsistent.

Another example of a conditional probability is given by the Staverman and Guggenheim (SG) entropy concept, where molecules are placed on a lattice. We will discuss it in the next section.

2.3.4 Probability and segment connectivity

In the FH-model each segment of the long molecule can be placed in the lattice without taking into account the connectivity of segments. Staverman [209] and Guggenheim (SG) [81] independently showed that the probability of placing subsequent segments of a polymer chain into the Flory-Huggins lattice increases with chain length, because the number of nearest neighbors, zq_j, decreases with the number of segments, r_j:

$$zq_j = zr_j - 2(r_j - 1). \tag{2.32}$$

Here z is the coordination number of the lattice and q_j the relative area of molecule of type j. In the SG concept the connectivity changes the entropy, and thereby the excess entropy change upon mixing:

$$\frac{\Delta S_{\mathrm{SG}}^{\mathrm{E}}}{k_{\mathrm{B}}} = -\sum_{j=1}^{M} N_j \left[\ln \frac{\phi_j}{x_j} + \tfrac{z}{2} q_j \ln \frac{\theta_j}{\phi_j} \right], \tag{2.33}$$

where the surface fraction of compound j in the mixture is defined by:

$$\theta_j = \frac{N_j q_j}{\sum N_j q_j}. \tag{2.34}$$

It is worthwhile to mention that the limit of $z \to \infty$, sets all $q_j = r_j$ and thereby $\theta_j = \phi_j$. This a mathematical way to approximate the system as a random ensemble of segments. It reduces the Staverman-Guggenheim equation to the Flory-Huggins equation. In the SG-model the first term, the FH-term, is related to the probability for placing the first segment of a chain in the lattice, while the second term accounts for the conditional probability for placing the subsequent segments of the chain, i.e. the connectivity. If we compare Eq. 2.33 with the Gibbs' entropy equation, Eq. 2.26, than the probability would be:

$$p_j = \phi_j \left[\frac{\theta_j}{\phi_j} \right]^{\frac{z}{2} q_j}. \tag{2.35}$$

We recognize in Eq. 2.26 the Bayes' theorem, where ϕ_j is the prior function and the term with the exponent the reduced likelihood function. The exponent is half the number of nearest neighbors of compound j (see Eq. 4.2). In Bayesian statistics the question is: What is the probability that a segment of compound j is covered with segments of molecules of type j? In chapter 3 we discuss this in detail.

The combinatorial activity coefficient equation for the Staverman-Guggenheim model is:

$$\ln \gamma_k^{\mathrm{comb}} = \ln \gamma_k^{\mathrm{FH}} + \ln \gamma_k^{\mathrm{SG}}, \tag{2.36}$$

of which we denote the last term as the Staverman-Guggenheim (SG) contribution. This term reads:

$$\ln \gamma_k^{\mathrm{SG}} = -\frac{z}{2} q_k \left[\ln \left(\frac{\phi_k}{\theta_k} \right) + 1 - \frac{\phi_k}{\theta_k} \right] \tag{2.37}$$

In the UNIQUAC model [2] and in different versions of the UNIFAC model the

coordination number, z, is usually set to the value 10, which can be seen as an average value between a cubic ($z = 6$) and a hexagonal closed packed structure ($z = 12$), but also as the value of an infinite chain in a hexagonal closed packed structure, where 2 of the 12 surrounding sites of a central segment are already occupied by 2 of its adjacent segments. However, when transforming Eq. 2.32 into:

$$z = 2(r_k - 1)/(r_k - q_k), \qquad (2.38)$$

and using the reported values for the relative volume and area of compounds in the original UNIFAC model and the Dortmund modification of it, we obtain remarkable results for the coordination number of these molecules as listed in Table 2.1. The

Model	UNIFAC(Orig.)			UNIFAC(Do.)		
Parameter	r_k	q_k	z	r_k	q_k	z
Ethane	1.802	1.696	15.1	1.265	2.122	-0.7
Propane	2.476	2.236	12.3	1.898	2.830	-2.0
n-Butane	3.151	2.776	11.5	2.530	3.538	-3.1
n-Hexane	4.499	3.856	10.9	3.795	4.954	-4.9
$(CH_2)_\infty$	∞	∞	10	∞	∞	-16.6

Table 2.1: Check on the coordination number (z) using the original and modified UNIFAC volume (r_k) and area (q_k) parameters and Eq. 2.38.

most salient finding is that UNIFAC(Do) produces negative coordination numbers, because the relative area is always larger than the relative volume for all alkanes. The original UNIFAC model gives, unphysically high coordination number values for ethane and propane, but converges to the end value of 10 for the series of n-alkanes, which is in agreement with the model. The inconsistency in coordination number shows that developers of the UNIFAC models have loosened the connectivity condition, eq. 6.8, in order to get better agreement with experimental results. Such an approach deteriorates the robustness of the model.

Besides this inconsistency there are four issues in the development of UNIFAC and COSMO-RS type of models that remained unsolved:

- Is the 'Donohue' exponent to define volume fractions correct, and what value is the best?

- What is the size of the reference area?

- Does the SG-model contrast with the Gibbs probability definition for entropy?

The 'Donohue' exponent:

It is known that the SG-term improves the combinatorial activity coefficient of the FH-model. It increases the activity coefficient by a factor 1.05-1.10. This improvement is too small to get agreement with experiments. Therefore, developers tried several modifications of the combinatorial contribution to bring the results of the UNIFAC and COSMO-RS type of models in agreement with experimental phase equilibrium data. For example in the Dortmund and Lyngby versions of UNIFAC, the volume fraction for the Flory-Huggins part of the SG-model is modified with a 'Donohue' exponent; $p = \frac{3}{4}$ and $p = \frac{2}{3}$ respectively. On the other hand Kikic *et al.* [112] abandoned the SG model and used for the FH-term $p = \frac{1}{3}$. Soares [205] found the infinite dilution activity coefficients of binary systems that Eq. 2.23 in the COSMO-SAC model requires a volume fraction defined by the exponent $p_j = 0.637$ on basis of analysis of experimental infinite dilution activity coefficients. In a subsequent study Soares [51] showed that after introduction of a dispersive energy contribution in the COSMO-SAC model the SG-correction is obsolete, and that it is sufficient to use the FH-model with $p_j = 0.75$.

Size of the surface area:

While the reference volume has no effect on the outcome of the activity coefficient, the reference surface has influence on the result. Originally, the area size, A_{ref}, was set equal to the van der Waals area of a repeating methylene group (CH_2): $A_{\text{ref}} = 41.5$ Å^2. This corresponds to a sphere with radius $R_{\text{ref}} = 1.82$ Å. Bronneberg [22] questioned this choice and used $R_{\text{ref}} = 0.505$ Å, a value, according to him, at which molecules "feel" their geometric details. This might be an appropriate value for the calculation of the interaction energy in the residual part of the UNIQUAC activity coefficient, but this value is smaller than the Bohr radius, i.e. 0.529 Å. In other words, this choice breaks down the lattice concept, because one can not split an atom over cells. Therefore, Bronneberg's solution has to be regarded as a mathematical result with no physical meaning. Against it, Soares [205] found by analysis of experimental data, in this case the infinite dilution activity coefficients of binary systems, that Eq. 2.23 that $A_{\text{ref}} = 124$ Å^2, which is about three times larger than the area of a methylene group.

Gibbs probability

It seems that the SG-term, where the probability of placing subsequent chain segments, i.e.:

$$p_j = \left(\frac{\phi_j}{\theta_j} \right)^{\frac{z}{2} q_j},$$ (2.39)

does not follow the concept of Gibbs entropy, since summation of it does not lead to unity. But this conclusion is not correct, because the SG-term is a conditional probability term that quantifies the probability of placing subsequent segments, after the first has been placed in the liquid volume.

In Chapter 3 we revisit Guggenheim's derivation to derive from the same equations, which Guggenheim used in 1950, an off-lattice expression, that does not require a reference area, and a coordination number. It provides a solution for the aforementioned issues.

2.3.5 The Free Volume contribution

The free-volume contribution in the activity coefficient is a temperature-dependent contribution for compounds that expand with temperature. This contribution can be combined with the FH-SG model. In the FH and SG model the change in free volume upon mixing liquids is neglected. Although this assumption holds for most liquid mixtures, free volume effects are not negligible for systems in which one of the components is less tightly packed in the pure and mixture state such as gas-solvent, polymer-solvent [172, 47], and some solvent-solvent solutions [58, 126, 47]. Free-volume corrections on the activity coefficient model are derived from the repulsion term of the equation of state. For that reason the free-volume contribution is also called the equation of state contribution [172].

Free-volume contribution from a cubic equations of state

In the cubic equations of state (cEoS) the compression factor is given by:

$$Z^{\text{cEoS}} = \frac{V}{V - b} = 1 + \frac{4\eta}{1 - 4\eta} = Z^{\text{IG}} + Z^{\text{FV}}, \tag{2.40}$$

where V and b are the molar and van der Waals volume respectively, and η the packing factor. The compression factor of an ideal gas $Z^{\text{IG}} \equiv 1$, which defines the free-volume (FV) contribution of the compression factor to:

$$Z^{\text{FV,cEoS}} = \frac{4\eta}{1 - 4\eta}. \tag{2.41}$$

The free-volume entropy departure function, $S^{EoS} - S^{IG}$, is found by integration:

$$-\frac{S^{\text{EoS}} - S^{\text{IG}}}{N_t k_{\text{B}}} = \int_0^\eta \frac{Z^{\text{EOS}} - Z^{\text{IG}}}{\eta} d\eta = \int_0^\eta \frac{Z^{\text{FV,EOS}}}{\eta} d\eta. \tag{2.42}$$

This gives for the cEoS the expression

$$-\frac{S^{\text{FV}}}{N_t k_{\text{B}}} = \int_0^\eta \frac{4}{1 - 4\eta} d\eta = -\ln(1 - 4\eta). \tag{2.43}$$

The entropy of mixing is:

$$-\frac{\Delta S_{\text{FV,cubic}}^{\text{E}}}{k_{\text{B}}} = \sum_{j=1}^{M} N_j \ln\left(\frac{1-4\eta_j}{1-4\bar{\eta}}\right), \tag{2.44}$$

where $\bar{\eta}$ is the average packing fraction of the mixture. Substitution of Eq. 2.44 into Eq. 2.4 gives the free-volume contribution to the activity coefficient equation:

$$\ln\gamma_k^{\text{FV,cEoS}} = \ln\left(\frac{1-4\eta_k}{1-4\bar{\eta}}\right) + \frac{\phi_k}{x_k}\left(\eta_k - \bar{\eta}\right)\frac{4}{1-4\bar{\eta}} \tag{2.45}$$

It is important to note here that this contribution vanishes when the average packing fraction is equal to the packing fraction of compound k. In that case the molecules of compound k have in the pure liquid and in the mixture the same free-volume. We also see that $\gamma_k \equiv 0$ when $\eta = \bar{\eta} = \eta_k = 0$, which is the ideal gas situation.

There is another interesting point to mention here. We can rewrite the sum of the combinatorial and free-volume activity coefficient equation as follows:

$$\ln\gamma_k^{\text{FH}} + \ln\gamma_k^{\text{FV,cEoS}} = \ln\frac{\phi_k}{x_k} + 1 - \frac{\phi_k}{x_k} + \ln\left(\frac{1-4\eta_k}{1-4\bar{\eta}}\right) + \frac{4\left(\eta_k - \bar{\eta}\right)}{1-4\bar{\eta}}\frac{\phi_k}{x_k}.$$

$$= \ln\left(\frac{\phi_k}{x_k}\frac{1-4\eta_k}{1-4\bar{\eta}}\right) + 1 - \frac{\phi_k}{x_k}\left(\frac{1-4\eta_k}{1-4\bar{\eta}}\right)$$

$$= \ln\left(\frac{W_k}{x_k}\right) + 1 - \frac{W_k}{x_k}, \tag{2.46}$$

where the free-volume fraction W_k is defined by:

$$W_k = \phi_k\frac{(1-4\eta_k)}{(1-4\bar{\eta})} = \frac{x_k v_k}{\sum x_k v_k}\frac{1-\frac{b_k}{v_k}}{1-\frac{\sum x_k b_k}{\sum x_k v_k}} = \frac{x_k\left(v_k - b_k\right)}{\sum x_k\left(v_k - b_k\right)}. \tag{2.47}$$

This result can also be derived from Eq. 2.15 by replacing the partial volume by the partial free-volume. It shows that a free-volume model derived from a cEoS can be merged into the Flory-Huggins equation. This equation was also derived in another way by Elbro *et al.* [47], who applied it to polymer solvents systems. Eq. 2.46 yields activity coefficients that are closer to unity than the FH activity coefficient equation, i.e. Eq. 2.21.

Another interesting point to mention is that Eq.2.46 is in line with Gibbs definition for entropy (i.e. $P_j = \Omega_j$). But this agreement is a result of the silent assumption that the excess volume is zero; i.e. the summation term in Eq. 2.44 does not contain excess volume terms. It also shows that the Flory-Huggins concept can only be aligned to Eq. 2.46, when $N_j r_j = V_j - N_j b_j$.

Free-volume contribution from a non-cubic equation of state

An example of a non-cubic equation of state is the Tonks-Flory equation of state [55], which has been used by Oishi and Prausnitz (OP) in the Free-Volume UNIFAC model [172]. The model has been developed for the description of solvent activity in polymer-solvent mixtures. Like in the cubic equations of state the activity coefficient is divided in a Flory-Huggins combinatorial and a free-volume contribution:

$$\ln \gamma_j = \ln \gamma_j^{\text{FH}} + \ln \gamma_j^{\text{OP}}. \tag{2.48}$$

According to Oishi *et al.* [172] the free-volume contribution in terms of the average packing factor η and the packing factor of compound j is:

$$\ln \gamma_j^{\text{OP}} = 3c_j \ln \left[\frac{\eta_j^{-1/3} - 1}{\bar{\eta}^{-1/3} - 1} \right] + c_j \left[\frac{\frac{1 - \bar{\eta}}{\eta_j}}{1 - \eta_j^{1/3}} \right]. \tag{2.49}$$

A liquid in which molecules can only translate in 3 directions would have $c_1 = 1$, while a liquid in which the molecules can also rotate on three different axis would have $c_1 = 2$. Frequently developers set $c_j = 1.1$ to obtain the best agreement with experiments. However, Fried *et al.* [62] showed that with $c_1 = 2$ also a good description of polymer-solvent equilibria is obtained. We might question the choice for $c_j = 1.1$. Could it be that this choice is biased by omitting the implementation of a dispersion term?

Although the OP model has proven to be successful for solvent solubility calculations, the equation contains an inconsistency. We can already notice it from the singularity of the second term at the r.h.s. when $\eta = \eta_j = 0$, where the expression should recover the ideal mixture case. A more appropriate expression (see Appendix A.2) is:

$$\ln \gamma_k^{\text{FV,T}} = 3c_k \ln \left(\frac{1 - \tilde{\eta}_k^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}} \right) + \bar{c} \frac{\tilde{\eta}^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}} \left(\frac{\eta_k}{\tilde{\eta}} - 1 \right) \frac{\phi_k}{x_k} \tag{2.50}$$

But besides this inconsistency, it is essential to understand that Eq. 2.49 is based on the work of Tonks [218], who obtained this equation of state for spherical molecules caged in dodecahedronal cells in a 3-dimensional lattice. The Tonks (T) equation of state is in terms of the compression factor, Z, the packing factor η, and maximum packing factor η_∞:

$$Z^{\text{T}} = \frac{1}{1 - \left(\frac{\eta}{\eta_\infty} \right)^{\frac{1}{3}}} = \frac{1}{1 - \tilde{\eta}^{\frac{1}{3}}}. \tag{2.51}$$

The Tonks equation is not accurate enough as shown in Fig. 2.3. Therefore free-volume models based on this equation are expected to be less robust. The Carnahan-Starling and the SAFT equations of state are better options to apply. We will discuss these equations of state in the next sections.

Free-volume contribution derived from a hard-spheres EoS

Carnahan and Starling [26] introduced a simple algebraic expression for the compression factor of hard-sphere fluids. They obtained this from the Padé approximation of Ree and Hoover [185] by setting the six constants in this equation equal to integer values. This was done in such a way that it yielded a geometric series, which could be transformed into a closed-form equation. Later, Mansoori *et al.* [159] showed that the CS approximation can also be derived from the exact solutions of the Percus-Yevick integral equation for hard spheres. The two solutions, which are obtained via the compressibility and the pressure route, are summed with a weight factor 2/3 and 1/3, respectively. The derivation steps that Carnahan and Starling followed are given below

$$
\begin{aligned}
Z^{\mathrm{HS}} &\approx 1 + 4\eta + 10\eta^2 + 18.3648\eta^3 + 28.2243\eta^4 + 39.8153\eta^5 + 53.3418\eta^6 + 68.526\eta^7 \\
&\quad + 85.83\eta^8 + 105.6\eta^9 + 126.5\eta^{10} + 158.3\eta^{11} + \dots \\
&\approx 1 + 4\eta + 10\eta^2 + 18\eta^3 + 28\eta^4 + 40\eta^5 + 54\eta^6 + 70\eta^7 + 88\eta^8 + 108\eta^9 + \dots \\
&\approx 1 + \sum_{n=1}^{\infty} \left(n^2 + 3n \right) \eta^n \\
&\approx \frac{1 + \eta + \eta^2 - \eta^3}{(1 - \eta)^3}.
\end{aligned}
\tag{2.52}
$$

The first line of Eq. 2.52 shows additional constants for the Padé approximation. These were computed years after the publication of the Carnahan and Starling paper [101], when the CPU speed and the memory of computers increased. The differences between the accurate computed constants and the constants of the CS equation are relatively small, making the CS equation still a versatile expression for equations of states. There are better descriptions for the compression factor available, see e.g. the review of Gow [73], and recently the work of Hansen-Goos [85]. However, for liquids with a packing factor in the range of 0.4 to 0.5 the CS-equation is sufficiently accurate and easy to understand from a mathematical point of view.

For a mixture the average packing factor of the hard-sphere fluid is defined by:

$$
\bar{\eta} = \frac{\pi}{6} \rho \sum_{i,j=1}^{M} x_i x_j d_{ij}^3,
\tag{2.53}
$$

where d_{ij} is the average hard-core diameter between two spheres i and j, v_j the molar liquid volume and x_j the mole fraction. In the case that $d_{ij}^3 = (d_i^3 + dj^3)/2$, Eq. 2.53 reduces to:

$$\bar{\eta} = \frac{\pi}{6}\rho\sum_{i=1}^{M}x_i d_i^3, \tag{2.54}$$

Substitution of the compression factor into Eq. 2.43 gives the entropy departure and subsequently the excess entropy of mixing:

$$-\frac{\Delta S_{\text{CS}}^{\text{E}}}{k_B} = \sum_{j=1}^{M}N_j\left[\frac{4\bar{\eta} - 3\bar{\eta}^2}{(1-\bar{\eta})^2} - \frac{4\eta_j - 3\eta_j^2}{(1-\eta_j)^2}\right], \tag{2.55}$$

where $\bar{\eta}$ and η_j denote the packing factor of the mixture and the pure compound j, respectively. We mention that Dodd and Sandler [42] uses the Boublik-Mansoori equation for the mixture entropy. This expression is more precise, but it complicates the expression for the excess entropy and makes it more difficult to understand what the magnitude of the free-volume contribution is. The activity coefficient contribution from the Carnahan-Starling EoS is obtained with Eq. 2.4:

$$\ln\gamma_k^{\text{FV,CS}} = \left[\frac{4\bar{\eta} - 3\bar{\eta}^2}{(1-\bar{\eta})^2} - \frac{4\eta_k - 3\eta_k^2}{(1-\eta_k)^2}\right] + \frac{\phi_k}{x_k}(\eta_k - \bar{\eta})\frac{4 - 2\bar{\eta}}{(1-\bar{\eta})^3} \tag{2.56}$$

We see that this activity coefficient contribution becomes unity when there is no difference between the packing fractions of the mixture and the pure compounds. The contribution vanishes also when both η and η_k go to zero (ideal gas). A property that we also observed in the van der Waals equation of state (Eq. 2.45). The use of this equation is limited to spherical molecules, such as the noble gases. This equation also holds for methane and neopentyl, due their globular structure. However, in the area of material and life sciences most molecules are polyatomic and non-globular. Flexible and elongated polyatomic molecules can be represented by a hard-sphere chain EoS, which we will discuss in the next section.

Free-volume contribution derived from SAFT

An improvement over the hard-sphere model is the SAFT [29, 27] and the PC-SAFT [78] EoS, where these spheres are connected to define chains of various lengths. We consider the case that all the chains have equal sized hard spheres. Compounds differ in the chain length, i.e. the number of spheres. In this situation the compression

factor of the SAFT and the PC-SAFT EoS is given by:

$$Z^{\text{HC}} = mZ^{\text{HS}} - (m-1)\eta \left(\frac{\partial \ln g_{\text{HS}}(\sigma)}{\partial \eta} \right), \tag{2.57}$$

where

$$g_{\text{HS}} = \frac{Z_{\text{HS}} - 1}{4\eta} = \frac{(2-\eta)}{2(1-\eta)^3}, \tag{2.58}$$

is the value of the radial distribution function for hard-spheres at contact according to the Carnahan-Starling model. Substitution of the radial function gives for the compression factor of a pure liquid of hard-sphere chains yields:

$$Z^{\text{HC}}_{\text{pure}} - Z^{\text{ig}}_{\text{pure}} = Z^{\text{HC}}_{\text{pure}} - 1 = m\frac{2\eta(2-\eta)}{(1-\eta)^3} - (m-1) \left(\frac{\eta(5-2\eta)}{(1-\eta)(2-\eta)} \right). \tag{2.59}$$

The first term on the r.h.s. reflects the addition of m independent spheres to a volume, which increases the compression factor, while the last term is the contribution of linking these spheres, which happens $(m-1)$ times. This reduces the compression factor.

Applying Eq. 2.43 we find for the entropy departure function:

$$-\frac{S^{\text{HC}} - S^{\text{ig}}}{N_t k_B} = m\frac{4\eta - 3\eta^2}{(1-\eta)^2} - (m-1)\ln\frac{2-\eta}{(1-\eta)^3}. \tag{2.60}$$

With the entropy departure function of the mixture and the pure components the excess entropy change upon mixing becomes:

$$-\frac{\Delta S^{\text{E}}_{\text{HC}}}{k_B} = \sum_{j=1}^{M} N_j m_j \left[\frac{(4\bar{\eta} - 3\bar{\eta}^2)}{(1-\bar{\eta})^2} - \frac{(4\eta_j - 3\eta_j^2)}{(1-\eta_j)^2} \right]$$
$$- N_j (m_j - 1)\ln\frac{(2-\bar{\eta})(1-\eta_j)^3}{(2-\eta_j)(1-\bar{\eta})^3}. \tag{2.61}$$

Subsequently, with Eq. 2.4 we find the free-volume contribution of the activity coefficient of hard-sphere chains:

$$\ln\gamma_k^{\text{FH,HC}} = m_k \left[\frac{4\bar{\eta} - 3\bar{\eta}^2}{(1-\bar{\eta})^2} - \frac{4\eta_k - 3\eta_k^2}{(1-\eta_k)^2} \right]$$
$$- (m_k - 1)\ln\frac{(2-\bar{\eta})(1-\eta_k)^3}{(2-\eta_k)(1-\bar{\eta})^3}$$
$$+ (\eta_k - \bar{\eta})\frac{\phi_k}{x_k} \left\{ m\left[\frac{4-2\bar{\eta}}{(1-\bar{\eta})^3} \right] - (m-1)\frac{5-2\bar{\eta}}{(1-\bar{\eta})(2-\bar{\eta})} \right\} \tag{2.62}$$

The first term between the square brackets is the contribution from the single sphere equation, while the second term is the contribution of the bonds between the spheres. The last term is the effect of packing fraction differences. Eq. 2.62 reduces to the Carnahan-Starling hard-sphere model (i.e. eq. 2.56), when the mixture consists of spheres only.

According to Vahid [1] the sPC-SAFT of Tihic *et al.* [215, 216] and the Elliott-Suresh-Donohue (ESD) model [48], are equations of state that have no free-volume 'entropy' of mixing. But this conclusion can only be made when the packing fraction is the same for all compounds; $\eta_k = \eta$. Fig. 2.2 depicts the packing fraction of normal alkanes at 25 °C between ethylene and eicosane, using the parameters of the original PC-SAFT model. It shows that the assumption that the constant packing assumption is only valid for mixtures of molecules larger than pentadecane. We
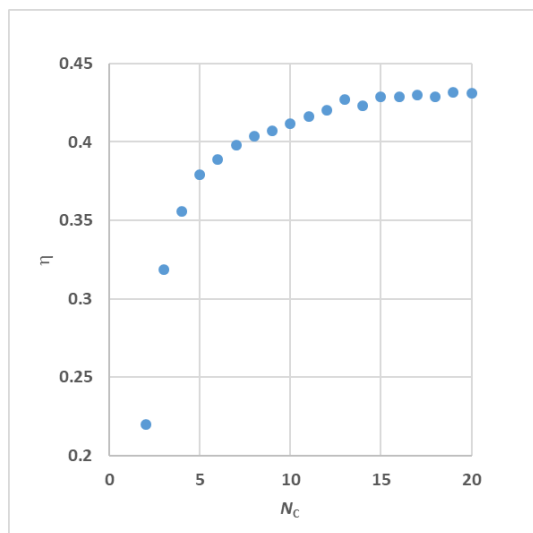
Figure 2.2: Packing fraction η of n-alkanes at T=25 °C as a function of the carbon number, N_C. Calculation performed with PC-SAFT using the original parameters [78] and the liquid density from DIPPR [23].

observe in Fig. 2.2 that above tetradecane the packing fraction is nearly constant, implying that here the free-volume contribution is negligible. Between hexane and hexadecane the packing fraction increases nearly linear with carbon number. In the next section we compare the free-volume models.

Comparison of the Free-Volume ACMs

Fig. 2.3 illustrates the compression factors of the the aforementioned free-volume activity coefficient models. We see in the left panel that the reciprocal compression factor of a fluid of spheres is best represented by the Carnahan-Starling EoS (CS). Especially in the liquid region, where $0.3 < \eta < 0.55$. The van der Waals EoS (vdW) is only applicable up to a packing factor of 0.05, which implies a gas phase, while Tonks equation deviates $+10\%$ in the liquid range where $0.4 < \eta < 0.5$. The right panel tells that the hard-sphere chain equation, Eq. 2.59, also known as Wertheim TPT-1 model [105], lies close to the by Molecular Dynamics (MD) determined values for fluids consisting of linear chains containing $m = 16$ and $m = 32$ spheres. Therefore, the hard-sphere chain free-volume model is expected to give better results for the free-volume contribution in activity coefficient models than the other models.
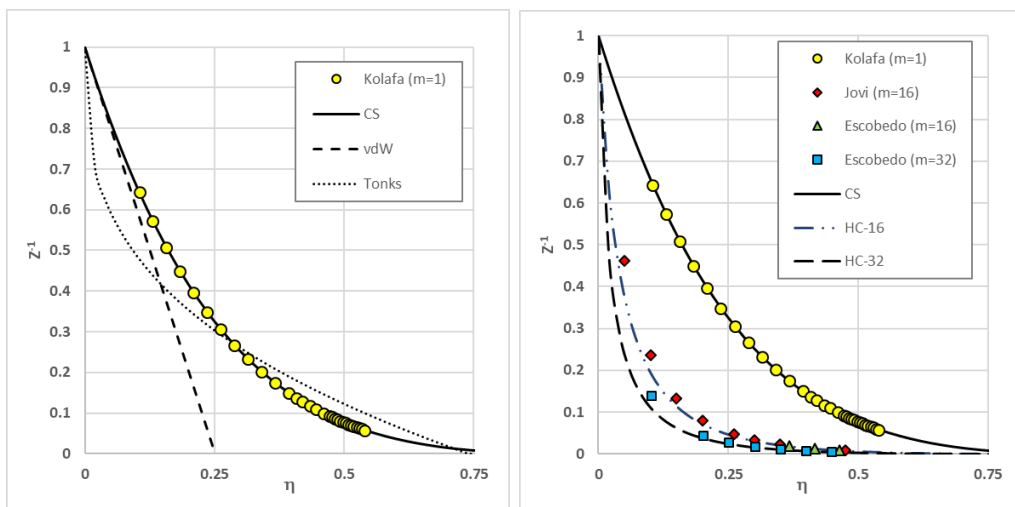
Figure 2.3: Reciprocal compression factor of hard-sphere (left panel) and linear hard-sphere chain fluids (right panel) with 16 and 32 spheres. Curves are calculated by Eqs. 2.40, A.7 and 2.59. Symbols represent MD results of Kolafa [121] (yellow circles), Escobedo (green triangles and squares) [50] and Jover (blue diamonds) [105]. Uncertainty range of the MD-results is smaller than the symbol width.

The maximum contribution of the free-volume is for the solute at infinite dilution in the solvent. In this limit we have:

$$\lim_{x_k \to 0} \frac{\phi_k}{x_k} = \frac{v_{\text{solute}}}{v_{\text{solvent}}} = \frac{\rho_{\text{solvent}}}{\rho_{\text{solute}}}. \tag{2.63}$$

The example given by Oishi and Prausnitz [172] gives us an idea about the mag-

nitudes of the FV contribution and the results are shown in Table 2.2.

ACM FV Model	η(solvent)	$\bar{\eta}$	$\ln \gamma_{benzene}$	$\gamma_{benzene}$
Oishi-Prausnitz (Eq. 2.49)	0.69	0.84	0.55	1.67
Oishi-Prausnitz (Rq. 2.50)	0.69	0.84	3.24	14
Tonks	0.69	0.84	8.11	26
Carnahan-Starling	0.421	0.424	-0.047	0.95
Hard-sphere chain	0.421	0.424	-0.096	0.91

Table 2.2: Example of free-volume contribution for benzene in polyisobutylene for various free-volume models with η defined by the models.

The packing factors of 0.69 and 0.84 originate from the publication of Oishi and Prausnitz. The liquid densities of benzene and polyisobutylene are 0.873 and 0.917 g/cm^3, respectively (see appendix A.2). The packing factor values for the hard-sphere and the hard-sphere chain length were taken from the PC-SAFT model [78, 79]. We see that the early models give large deviation from unity, while the hard-sphere and hard-sphere models are close to ideal behavior.

In this thesis we focus on condensed systems, where volume contraction/expansion of the liquid is small, so that free-volume changes can be ignored. We will not investigate the free-volume effect of the activity model. In Chapter 3 we show that the Guggenheim-Staverman correction of the Flory-Huggins term can be generalized. Subsequently, in Chapter 4, we will show that this correction can be applied in models like UNIQUAC and COSMOSPACE. It implies that the free-volume effect is lumped into the residual part of the activity coefficient model.

2.4 Enthalpy and the activity coefficient of dispersion

The enthalpy of mixing, more commonly known as the excess enthalpy, is observed by an endothermal ($\Delta H > 0$) or exothermal ($\Delta H < 0$) change, when two or more fluids are mixed. In general, an observer will notice endothermal behavior when associating bonds break or polar interactions weaken. Opposite to this exothermal behavior is observed when bonds are formed or polar interactions have become stronger. These processes are dependent on system and concentration.

Fig. 2.4 shows in the left and right panel the excess enthalpy of mixing hexane in dodecane and hexadecane, respectively at 25 °C. In both cases heat is added to maintain the temperature. Both systems show over the whole concentration range endothermal behavior, implying that the net result is the breaking of bonds and weakening of interactions. It demonstrates, as was also pointed out by Kontegeorgis [125], that mixtures of alkanes are not athermal, as is assumed in the UNIFAC and COSMO-RS models. Due to the differences in the packing factor of these alkanes, n-dodecane and n-hexadecane have $\eta = 0.42$ and $\eta = 0.43$, respectively, while n-hexane has $\eta = 0.39$, the overall distance between the molecules increases, thereby weakening the interaction, which implies an endothermal process.
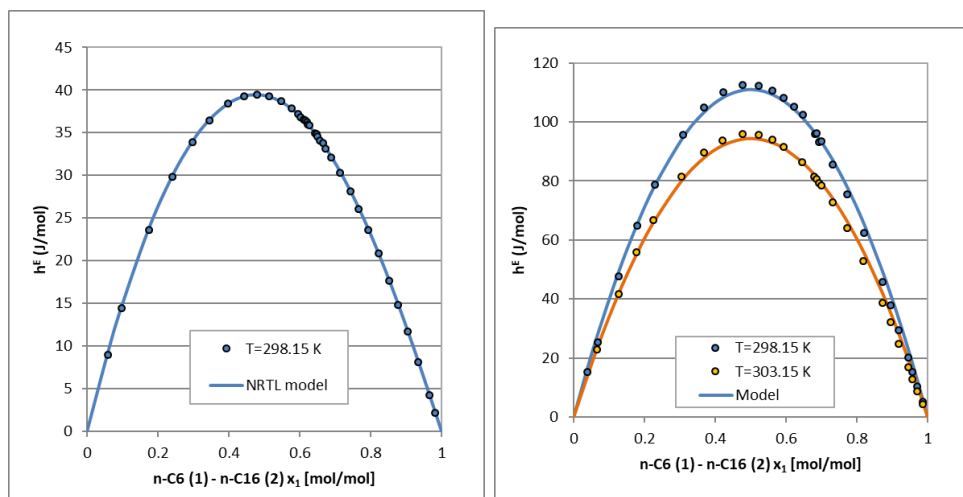
Figure 2.4: Excess enthalpy of the systems n-hexane(1)-n-dodecane(2) at 25°C (left panel) and the system n-hexane(1)-n-hexadecane(2) at 25 and 35 °C (right panel). Data from [174, 167, 162]. Curves depict the NRTL model [186].

The mixing enthalpy of alkanes is much smaller than, for instance, the case of toluene and phenol, where hydrogen bonds are broken (see Fig. 2.5). Table 2.3 demonstrates the magnitude of the excess enthalpy for typical systems.

binary system	h^E_{extr}/RT	comment	ref.
n-hexane - n-dodecane	0.016	dispersion	[174]
n-hexane - n-hexadecane	0.046	dispersion	[167]
n-pentane - acetone	0.86	dispersion/polar	[67]
acetone - water	-0.31 and $+0.08$	hydrogen bonds	[67]
toluene - phenol	0.28	hydrogen bonds	[171]

Table 2.3: Excess enthalpy extrema in binary systems.

2.4.1 Basic equations for the excess enthalpy

The excess enthalpy at constant pressure is the sum of changes in dispersion energy (ΔU_{disp}), dipole-dipole interaction energy (ΔU_{pol}), association energy (ΔU_{AB}) and free-volume (ΔV_{FV}). In mathematical form we have:

$$\Delta H^{\mathrm{E}} = \Delta U_{\mathrm{disp}} + \Delta U_{\mathrm{pol}} + \Delta U_{\mathrm{AB}} + P\left(\Delta V_{FV}\right). \tag{2.64}$$

The molar excess enthalpy, h^{E} can be written as:

$$h^{\mathrm{E}} = \sum_{j=1}^{M} x_j \bar{h}_j\left(x_j\right) = -RT \sum_{j=1}^{M} x_j \ln \gamma_j^{\mathrm{res}}, \tag{2.65}$$

where $\bar{h}_j\left(x_j\right)$ is the partial molar enthalpy of component j, and γ_j^{res} the residual activity coefficient. When the mixture contains molecules, which have no permanent multipole nor association sides to form a hydrogen bonds, then the molecules will interact by dispersion. This situation is present for all alkanes, see Fig. 2.4. For these molecules the residual activity coefficient reduces to the dispersion activity coefficient. We can write for the dispersion contribution to the excess enthalpy:

$$h^{\mathrm{E,disp}} = RT \sum_{j=1}^{M} x_j \ln \gamma_j^{\mathrm{disp}}, \tag{2.66}$$

The dispersion activity coefficient is obtained by:

$$\ln \gamma_k^{\mathrm{disp}} = \frac{1}{RT}\left(\frac{h^{\mathrm{E,disp}}}{\partial x_k}\right)_{\mathrm{T,P,x_j \neq x_k}} \tag{2.67}$$

2.4.2 Temperature derivative of the activity coefficient

The temperature derivative of the activity coefficient of compound j at constant pressure is related to the partial molar excess enthalpy of compound j by the relation:

$$\left(\frac{\partial \ln \gamma_j \left(x_j \right)}{\partial T} \right)_P = -\frac{\bar{h}_j^E \left(x_j \right)}{RT^2}, \tag{2.68}$$

where $\bar{h}_j^E \left(x_j \right)$ is the partial molar excess enthalpy of compound j at concentration x_j. Eq. 2.68 tells that an endothermal mixing process implies an activity coefficient drop with increasing temperature. For small temperature differences, we can approximate Eq. 2.68 by the relation:

$$\frac{\ln \gamma_j \left(x_j, T_2 \right) - \ln \gamma_j \left(x_j, T_1 \right)}{T_2 - T_1} = -\frac{\bar{h}_j^E \left(x_j, T_1 \right)}{RT_1^2} \Leftrightarrow$$

$$\gamma_j \left(x_j, T_2 \right) = \gamma_j \left(x_j, T_1 \right) \exp \left[- \left(T_2 - T_1 \right) \frac{\bar{h}_j^E \left(x_j, T_1 \right)}{RT_1^2} \right]. \tag{2.69}$$

In the case of toluene-phenol the tangent line located at $x_1 = 0.4$, where $h^E = 0.79$ kJ/mol, has a slope of 1.23 kJ/mol. This gives for the partial excess enthalpy of toluene and phenol $\bar{h}_1^E = 1.58$ kJ/mol and $\bar{h}_2^E = 0.35$ kJ/mol, respectively. The largest change in activity coefficient is for this case at the limiting concentrations, since there the slope of the excess enthalpy is the largest.

Table 2.4 summarizes the excessenthalpy and activity coefficient results of the toluene phenol binary at the given finite and the two limiting liquid concentrations.

	$T = 414$ K		$T = 414$ K		$T = 424$ K (Eq. 2.69)	
x_1	γ_1	γ_2	$\bar{h}_1^E$ (kJ/mol)	$\bar{h}_2^E$ (kJ/mol)	γ_1	γ_2
0	2.03	1	2.85	0	1.99	1
0.4	1.36	1.11	1.59	0.35	1.34	1.11
1	1	2.52	0	5.47	1	2.43

Table 2.4: Activity coefficient change of the binary mixture Toluene-Phenol.

The table demonstrates that an endotherm mixture yields activity coefficients, which drop with temperature.
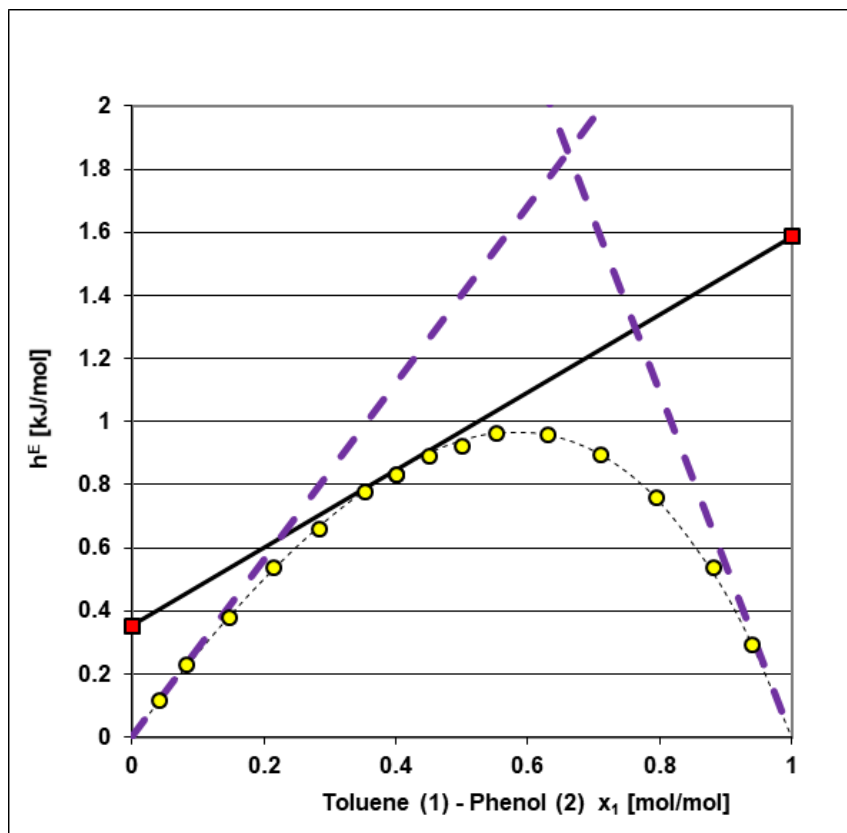
Figure 2.5: Excess enthalpy of toluene(1)-phenol(2) at 140 °C. Exp. data from [171]. Dashed curve depicts NRTL model [186]. The tangent lines at $x =0$, 0.4 and 1 are required to determine the partial molar enthalpy of the compounds as explained in text.

2.5 Concluding remarks

This introduction chapter supports the work discussed in the next chapters. We showed how expressions for the combinatorial activity coefficient can be obtained. We look into this matter in more detail in chapters 3 and 4.

The paragraph on the free-volume contribution showed that this contribution does influence the total activity coefficient, and can not be neglected. When molecules are treated as chains of spheres, the differences in diameters cause a positive deviation from the law of Raoult, as we showed by the CS-based free-volume activity coefficient, while the formation of chains causes a negative contribution. We pointed out that the free-volume term of the Oishi-Prausnitz model, has to be treated with caution. A more consistent version is derived, but we do not recommend to use this

corrected equation for future research, since the hard-sphere chain model, including sphere diameter differences, offers a better alternative. In conclusion, the free-volume contribution mediates the activity coefficient, when packing fractions differ significantly. The effect of this in parametrization of a dispersion contribution is out of the scope of this thesis. Including free-volume implies that the activity model requires information on the packing fraction of the compounds. This can be supplied by a Rackett equation or an equation of state.

Instead of incorporating a free-volume contribution, we investigated the dispersion contribution by deriving expressions from the van der Waals and the PC-SAFT EoS. The reason for this choice is made at the time when this work was not conclusive on free-volume models. The dispersion contribution is discussed in chapters 5 and 6, respectively. In chapter 6 we derive an elegant equation that incorporates the effect of chain architecture, which is also applied in chapter 7 when we investigate the improvement of the interaction energy by applying the self-consistent field equation of the COSMOSPACE model [115]. In chapter 8 we will show that, due to the 3D structure of molecules, a part of the interactions have to be disregarded. This development that was undertaken before we investigated the combinatorial and dispersion contribution in detail. On purpose we omitted systems with large size differences in this work, in order to avoid large deviating effects from the combinatorial and free-volume term. Instead we used the Huyskens equation as a temporarily solution, which would bring deviation down to an estimated maximum error of 10%. We expect, however, that an update of the COSMO-3D electrostatic energy-parameters is required after the dispersion term (e.g. Eq. 6.16) has been implemented.

Chapter 3

Generalized combinatorial activity coefficient model[1]

3.1 Introduction

After Huggins [95] and Flory [52] introduced a lattice model to quantify the Gibbs energy of an athermal mixture, i.e. a mixture in which there is no energetic interaction between molecules, Guggenheim [80] refined it. While according to the Flory-Huggins model small molecules can be placed at random in a lattice, Guggenheim took into account that the placement of mono-segmented molecules in a lattice is influenced by the presence of multi-segmented molecules. Although Guggenheim's model has been applied in many activity coefficient models later, there are two important issues in the Guggenheim model to be aware of. First, Guggenheim considers that all molecules have the same coordination number, while it is known that the coordination number ranges from 6 to 12. Secondly, liquids do not posses a long range order, and therefore the concept of a lattice will limit the number of possible configurations that molecules can make in a liquid. We will touch upon these two issues by going through Guggenheim's derivation and propose a generalized expression for the Gibbs energy of mixing. Finally, we derive a new relation for the combinatorial activity coefficient, which can be applied to mixtures having molecules of arbitrary shape.

---

[1]Based on the publication: Generalization of Guggenheim's combinatorial activity coefficient equation, J. Mol. Liq. 266 (2018) 467-471, by Gerard J.P. Krooshof, Remco Tuinier, and Gijsbertus de With.

3.2 Theory

3.2.1 Definition of volume and surface area fractions

We start with a liquid mixture consisting of two components, A and B. When we consider that molecules A and B fit exactly into r_A and r_B lattice cells, respectively, and assume that there are no empty cells, then the total number of lattice cells is given by

$$r_{\text{tot}} = r_A N_A + r_B N_B, \tag{3.1}$$

where N_j is the number of molecules of type j. On the other hand, we can treat the system as a fluid, which has a total volume defined by the number and molecular volumes of the molecules V_j

$$V_{\text{tot}} = V_A N_A + V_B N_B. \tag{3.2}$$

In the lattice model the probability that a lattice cell is populated by a segment of molecule A or B is given by the volume fraction, defined as

$$\phi_A = \frac{N_A r_A}{N_A r_A + N_B r_B} \quad \text{and} \quad \phi_B = \frac{N_B r_B}{N_A r_A + N_B r_B}, \tag{3.3}$$

respectively, while in the fluid model the volume fractions are expressed by

$$\phi_A = \frac{N_A V_A}{N_A V_A + N_B V_B} \quad \text{and} \quad \phi_B = \frac{N_B V_B}{N_A r_A + N_B V_B}, \tag{3.4}$$

respectively. The definition of Eq. 3.3 is used in the Flory-Huggins model, while Eq. 3.4 is due to Hildebrand [87], who demonstrated by this that the equation of the Flory-Huggins model does not require a lattice definition.
Both methods assume that molecules are randomly positioned in the liquid space.

Guggenheim, Stavermann [209], and Tompa [217] pointed out, however, that the placement of segments can not be performed randomly, because the segments of chain molecules have a higher probability to be placed aside each other. To correct the volume-based probability, they introduced the surface area fraction to account for the probability to form a contact pair of lattice cells. The surface area fractions can be defined as

$$\theta_A = \frac{N_A z_A q_A}{N_A z_A q_A + N_B z_B q_B} \quad \text{and} \quad \theta_B = \frac{N_B z_B q_B}{N_A z_A q_A + N_B z_B q_B}. \tag{3.5}$$

In this generalized Guggenheim expression, z_j is the coordination number of the lattice of the pure component j and q_j the relative surface area of molecule j. Guggenheim used one lattice number for all molecules and put the relative surface area

of one of the molecule types to unity. We do not use this simplification, so that we can interchange the indices A and B in equations later on.

As we mentioned above, the volume fraction can be defined using the lattice approach (Flory-Huggins), as well as the fluid approach (Hildebrand). The same can be done for the surface fraction. In a fluid the external surface area of the molecules can be used to define the probability that the molecules make contact, therefore

$$\theta_A = \frac{N_A a_A}{N_A a_A + N_B a_B} \quad \text{and} \quad \theta_B = \frac{N_B a_B}{N_A a_A + N_B a_B}. \tag{3.6}$$

In here a_j is the external contact area of molecule j.

It is worthwhile to mention that the volume and surface area fractions are related to each other by the expressions

$$\theta_A = \frac{\phi_A}{\phi_A + f\phi_B} \quad \text{and} \quad \theta_B = \frac{f\phi_B}{\phi_A + f\phi_B}, \tag{3.7}$$

where the factor f can be calculated from

$$f_{\text{lat}} = \frac{z_B q_B r_A}{z_A q_A r_B} \quad \text{and} \quad f_{\text{fluid}} = \frac{a_B V_A}{a_A V_B}, \tag{3.8}$$

for the lattice and the fluid approach, respectively. These relations enable us to simplify the expressions above, which will be used later on. Further, we mention that in the original lattice model of Guggenheim $r_A = q_A = 1$, $r_B = r$, $q_B = q$, or $z_A = z_B = z$, which makes $f_{\text{lat}} = q/r$.

3.2.2 Generalized Gibbs energy of mixing expression: Lattice-based derivation

We follow the work of Guggenheim [80] and Tompa [217] to derive an expression for the Gibbs energy of mixing of an athermal system. Suppose that a lattice cell is occupied by a segment of molecule B, then the ratio of the probability that a given neighboring cell is occupied by a segment of another molecule B', $p_{BB'}$, or by a segment of molecule A, p_{BA}, is equal to the ratio in number of nearest neighbors, $z_B q_B N_B$ and $z_A q_A N_A$

$$\frac{p_{BB'}}{p_{BA}} = \frac{N_B z_B q_B}{N_A z_A q_A} = \frac{\theta_B}{\theta_A}. \tag{3.9}$$

Since the total number of nearest neighbors of molecule B is $z_B q_B$ and not $z_B r_B$ due to the connectivity of the segments, the sum of these probabilities, i.e. $p_{BB'} + p_{BA}$, is equal to the rescaled value of the volume fraction of molecule B:

$$p_{BB'} + p_{BA} = \phi_B \frac{z_B q_B}{z_B r_B} = \phi_B \frac{q_B}{r_B}. \tag{3.10}$$

Combining the two equations leads to

$$p_{\text{BB'}} = \phi_B \frac{q_B}{r_B} \theta_B \quad \text{and} \quad p_{\text{BA}} = \phi_B \frac{q_B}{r_B} \theta_A. \qquad (3.11)$$

We note that the probability that the neighboring cell is occupied by the same B molecule is calculated by

$$p_{\text{BB}} = \phi_B - p_{\text{BB'}} - p_{\text{BA}} = \phi_B \left(1 - \frac{q_B}{r_B} \right). \qquad (3.12)$$

Likewise, when a lattice cell is occupied by a segment of molecule A, the probability that a given neighboring site is occupied by molecule B or by another molecule A is

$$p_{\text{AB}} = \phi_A \frac{q_A}{r_A} \theta_B \quad \text{and} \quad p_{\text{AA'}} = \phi_A \frac{q_A}{r_A} \theta_A, \qquad (3.13)$$

This procedure can be extended to the situation that a lattice cell is surrounded by more neighboring sites. For the probability that segment A has $x - 1$ neighboring sites of type A we obtain

$$p_{\text{AxA}} = \phi_A \frac{q_A}{r_A} (\theta_A)^{x-1}. \qquad (3.14)$$

In the same way, the probability that segment B has $y - 1$ adjacent sites of type B we have

$$p_{\text{ByB}} = \phi_B \frac{q_B}{r_B} (\theta_B)^{y-1}. \qquad (3.15)$$

We mention that Guggenheim derived Eqs. 3.14 and 3.15 by the concept of exchange rate, where a segment of a molecule having r segments, here molecule B, is replaced by the mono-segmented molecule A. In the above equation this implies that $r_A = q_A = x = 1$, $r_B = r = y$, and $q_B = q$. We continue with Guggenheim's derivation, but use the term probability, instead of the term exchange rate, and we take a small volume element in the liquid, which has the size of the least common multiple of the number of segments of the molecules in the mixture. Suppose, for instance, a mixture of trimers and tetramers. This requires a partial liquid volume element equivalent to 12 lattice cells, in which 4 trimers or 3 tetramers fit. In general, this volume element is equivalent to $r_A \cdot r_B$ adjacent lattice cells. The ratio of the probability that this liquid volume contains a cell of type A and $r_B - 1$ adjacent cells of type A and the probability that the cell is of type B and has $r_A - 1$ adjacent cells of type B is

$$\alpha_{\text{liq}} = C \frac{\phi_B \frac{q_B}{r_B} (\theta_B)^{r_A-1}}{\phi_A \frac{q_A}{r_A} (\theta_A)^{r_B-1}} = K_{\text{liq}} \frac{\phi_B (\theta_B)^{r_A-1}}{\phi_A (\theta_A)^{r_B-1}}, \qquad (3.16)$$

where C is a proportionality constant, that is independent of concentration, and K_{liq} is a constant that includes the size parameters r_j and q_j of the components.

Guggenheim's formula is recovered when $r_A = 1$ and $r_B = r$. In order to connect this probability to the Gibbs energy, Guggenheim used the thermodynamic constraint that at equilibrium the Gibbs energy of the liquid and gas are equal. We now consider a gas volume element. The probability to fill this volume element by A or B molecules is proportional to the partial pressures, P_A and P_B. Therefore, the ratio of the probability to fill this gas volume element with r_A molecules of type B and the probability to fill it with r_B molecules of type A is

$$\alpha_{\text{gas}} = K_{\text{gas}} \frac{P_B^{r_A}}{P_A^{r_B}}. \tag{3.17}$$

with K_{gas} a constant that is concentration independent. The partial vapor pressure P_j of compound j can be expressed in terms of the chemical potential μ_j or the reduced chemical potential $\tilde{\mu}_A$ by

$$P_j = \exp\left(\frac{\mu_j}{RT}\right) = \exp\left(\tilde{\mu}_j\right). \tag{3.18}$$

Substitution of Eq. 3.18 into the logarithmic expression of Eq. 3.17 gives

$$\begin{aligned}
\ln \alpha &= \ln K_{\text{gas}} + r_A \ln P_B - r_B \ln P_A \\
&= \ln K_{\text{gas}} + r_A \tilde{\mu}_B - r_B \tilde{\mu}_A.
\end{aligned} \tag{3.19}$$

Differentiation yields

$$\mathrm{d} \ln \alpha = r_A \mathrm{d}\tilde{\mu}_B - r_B \mathrm{d}\tilde{\mu}_A. \tag{3.20}$$

By using the Gibbs-Duhem equation

$$\sum_j N_j \mathrm{d}\tilde{\mu}_j = 0, \tag{3.21}$$

the reduced chemical potential of compound B can be eliminated and we obtain

$$\tilde{\mu}_A = -\frac{N_B}{(N_A r_A + N_B r_B)}\mathrm{d}\ln\alpha = -\frac{\phi_B}{r_B}\mathrm{d}\ln\alpha = -\frac{\phi_B}{r_B\alpha}\mathrm{d}\alpha. \tag{3.22}$$

Integration of Eq. 3.22 from the pure state of compound A, $\phi_B = 0$, to a final concentration ϕ_B provides the change in Gibbs energy upon mixing of compound A with B to concentration ϕ_A

$$\int_{\text{pure A}}^{\text{mixture}} \mathrm{d}\tilde{\mu}_A = \Delta\tilde{\mu}_A = -\frac{1}{r_B}\int_0^{\phi_B} \frac{\phi_B}{\alpha}\left(\frac{\partial\alpha}{\partial\phi_B}\right)\mathrm{d}\phi_B. \tag{3.23}$$

Substitution of Eq. 3.16 into Eq. 3.23 gives for the chemical potential difference of compound A in the pure liquid state and the mixture

$$\Delta\tilde{\mu}_A = -\frac{1}{r_B} \int_0^{\phi_B} \frac{1}{\phi_A} + (r_A - 1)\frac{\phi_B}{\theta_B}\left(\frac{\partial\theta_B}{\partial\phi_B}\right) - (r_B - 1)\frac{\phi_B}{\theta_A}\left(\frac{\partial\theta_A}{\partial\phi_B}\right) d\phi_B. \quad (3.24)$$

After evaluating the partial derivatives the integral reduces to

$$\Delta\tilde{\mu}_A = -\frac{1}{r_B} \int_0^{\phi_B} \frac{r_A\theta_A + r_B\theta_B}{\phi_A} d\phi_B. \quad (3.25)$$

With Eqs. 3.7 and 3.8 the surface area fractions in the integral can be written in terms of ϕ_A, after which the integration yields

$$\begin{aligned} \Delta\tilde{\mu}_A &= -\frac{1}{r_B}\left[\frac{r_A\ln(\phi_A + f_{\text{lat}}\phi_B)}{f_{\text{lat}} - 1} - r_B\ln(1 - \phi_B) - \frac{r_B\ln(\phi_A + f_{\text{lat}}\phi_B)}{f_{\text{lat}} - 1}\right] \\ &= \ln(\phi_A) + \frac{1 - r_A/r_B}{f_{\text{lat}} - 1}\ln\left(\frac{\phi_A}{\theta_A}\right) \\ &= \ln(\phi_A) - \Gamma_A\ln\left(\frac{\phi_A}{\theta_A}\right). \end{aligned} \quad (3.26)$$

Interchanging the indices results for component B in

$$\Delta\tilde{\mu}_B = \ln(\phi_B) - \Gamma_B\ln\left(\frac{\phi_B}{\theta_B}\right). \quad (3.27)$$

In the above equations we defined the concentration independent factors:

$$\Gamma_A \equiv \frac{z_A q_A r_A - z_A q_A r_B}{z_B q_B r_A - z_A q_A r_B} \quad \text{and} \quad \Gamma_B \equiv \frac{z_B q_B r_B - z_B q_B r_A}{z_A q_A r_B - z_B q_B r_A}. \quad (3.28)$$

From Eq. 3.26 and Eq. 3.27 follows that the total reduced Gibbs energy change upon mixing of compounds A and B is given by

$$\Delta\tilde{G} = N_A\ln(\phi_A) + N_B\ln(\phi_B) - \left[N_A\Gamma_A\ln\left(\frac{\phi_A}{\theta_A}\right) + N_B\Gamma_B\ln\left(\frac{\phi_B}{\theta_B}\right)\right]. \quad (3.29)$$

Again, Guggenheim's original equation is recovered by setting $r_A = q_A = 1$, $r_B = r$, $q_B = q$, and $z_A = z_B = z$, which yields $\Gamma_A = (r - 1)/(r - q)$ and $\Gamma_B = q\Gamma_A$.

3.2.3 Generalized Gibbs energy of mixing expression: Off-lattice derivation

The above derivation was carried out in terms of lattice parameters, but the same results can be derived without the definition of a lattice. We start by considering a small volume element in the liquid with volume V_{box}. The number of molecules of type j that fit into this volume element, m_j, is determined by the molecular volume of molecule V_j given by

$$m_j = \frac{V_{\text{box}}}{V_j}. \tag{3.30}$$

The ratio of the probabilities that this volume element is filled with only A or B molecules from the liquid phase, and which are all adjacent to each other, is

$$\alpha_{\text{liq}} = K_{\text{liq}} \frac{\phi_{\text{B}} (\theta_{\text{B}})^{m_{\text{A}}-1}}{\phi_{\text{A}} (\theta_{\text{A}})^{m_{\text{B}}-1}}. \tag{3.31}$$

In the gas phase the ratio of the probabilities that a volume element is filled with only A or B molecules, is ruled by the partial vapor pressures

$$\alpha_{\text{gas}} = K_{\text{gas}} \frac{P_{\text{B}}^{m_{\text{A}}}}{P_{\text{A}}^{m_{\text{B}}}}. \tag{3.32}$$

By applying the same procedure as in section 3.2.2 we obtain again Eq. 3.29, but now the factors, Γ_j, are defined by the molecular volume and external surface area of the compounds

$$\Gamma_{\text{A}} = \frac{a_{\text{A}} V_{\text{A}} - a_{\text{A}} V_{\text{B}}}{a_{\text{B}} V_{\text{A}} - a_{\text{A}} V_{\text{B}}} \quad \text{and} \quad \Gamma_{\text{B}} = \frac{a_{\text{B}} V_{\text{B}} - a_{\text{B}} V_{\text{A}}}{a_{\text{A}} V_{\text{B}} - a_{\text{B}} V_{\text{A}}}. \tag{3.33}$$

3.2.4 Combinatorial activity coefficients

Differentiation of Eq. 3.29 with respect to N_{A} yields the combinatorial activity coefficient of component A

$$\ln(\gamma_{\text{A}}) = \left[\ln\left(\frac{\phi_{\text{A}}}{x_{\text{A}}}\right) + 1 - \frac{\phi_{\text{A}}}{x_{\text{A}}} \right] - \Gamma_{\text{A}} \left[\ln\left(\frac{\phi_{\text{A}}}{\theta_{\text{A}}}\right) + 1 - \frac{\phi_{\text{A}}}{\theta_{\text{A}}} \right]. \tag{3.34}$$

The factor Γ_{A} can be written in terms of x_{A}, ϕ_{A}, and θ_{A} by substitution Eqs. 3.3 and 3.5 into Eq. 3.28

$$\Gamma_{\text{A}} = \frac{1 - \frac{\phi_{\text{A}}}{x_{\text{A}}}}{1 - \frac{\phi_{\text{A}}}{\theta_{\text{A}}}} .. \tag{3.35}$$

The same result is obtained when Eqs. 3.4 and 3.6 are substituted into Eq. 3.33. The factor Γ_A as we will discuss in Chapter 4 is equal to half the number of nearest neighbors. Here we note that, regardless of whether a lattice- or fluid-based model is used, and regardless of the shape of the molecules, the general expression for the combinatorial activity coefficient of component j in a multicomponent mixture is given by

$$\ln\left(\gamma_j\right) = \ln\left(\frac{\phi_j}{x_j}\right) + \left(1 - \frac{\phi_j}{x_j}\right)\left[\frac{\ln\left(\frac{\phi_j}{\theta_j}\right)}{\frac{\phi_j}{\theta_j} - 1}\right]. \tag{3.36}$$

Here, the term in the square brackets defines the Guggenheim correction with respect to the Flory-Huggins model. For $\phi_j = \theta_j$ the correction term is unity. Fig. 3.1 shows the correction term as a function of the ratio ϕ_j/θ_j.
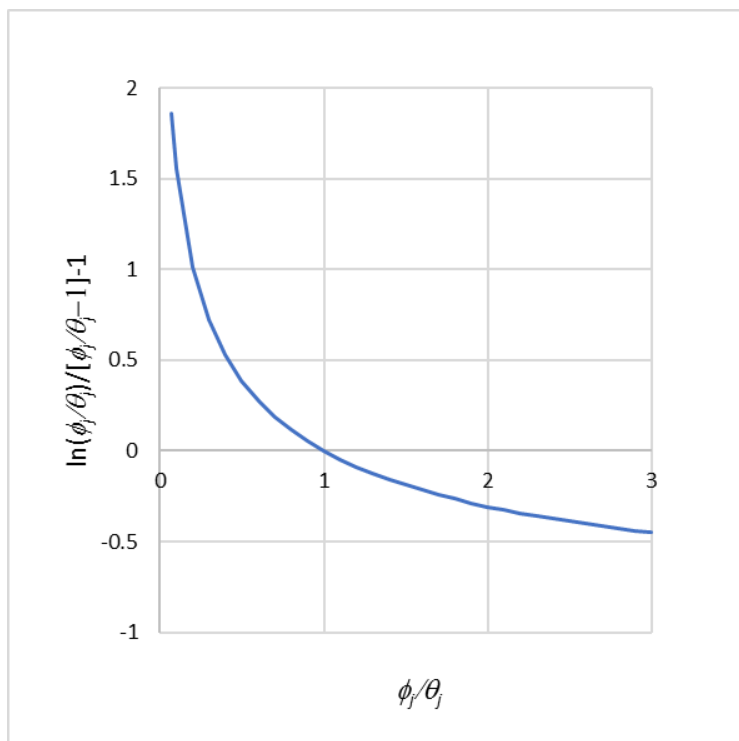
Figure 3.1: Correction term of Eq. 3.36 as a function of ratio of volume and surface fraction of compound j.

3.3 Discussion and concluding remarks

We recall that Guggenheim used a simplified form for Γ_j. He imposed $r_A = q_A = 1$, $r_B = r$, $q_B = q$, and $z_A = z_B = z$ and applied the lattice relation:

$$zq = zr - 2(r - 1), \tag{3.37}$$

which is valid for linear and branched molecules. As a result Guggenheim obtained $\Gamma_A = \frac{z}{2}$ and $\Gamma_B = q\frac{z}{2}$ for Eq. 3.29. Later, Vera *et al.* [224] used this equation to derive the following restricted form of the combinatorial activity coefficient

$$\ln(\gamma_j) = \ln\left(\frac{\phi_j}{x_j}\right) - \frac{zq_j}{2}\ln\left(\frac{\phi_j}{\theta_j}\right), \tag{3.38}$$

but this form was seldom applied in activity coefficient models. Instead, lattice activity coefficient models still contained the combinatorial term of the original UNIQUAC model [2]; Eq. 3.34 with $\Gamma_A = zq_A/2$. See, for instance, the development of the UNIFAC [59], the COSMOSPACE [117], the MOQUAC [22] and the COSMOSAC [144] model. Our result of Eq. 3.36 shows that there is no need to define:

- a coordination number (z), since the concept of a lattice is obsolete

- a relative volume (r_j) to calculate volume fractions (ϕ_j)

- a relative surface area (q_j) to calculate surface area fractions (θ_j)

- the number of nearest neighbors by the term $z.q_k/2$

Therefore, the original and restricted form of the combinatorial activity coefficient could be replaced by Eq.3.36, and supplemented with an activity coefficient model that accounts for the dispersive interaction between non-polar molecules. Such an approach has been applied in the MOSCED model [214], which is used to calculate the activity coefficients at infinite solution of binary mixtures. In here the dispersion is based on the Hansen solubility model [84], while the combinatorial activity coefficient at infinite dilution of solute A in solvent B is computed by

$$\ln(\gamma_A^\infty) = \ln\left(\frac{V_A}{V_B}\right)^{0.953} + 1 - \left(\frac{V_A}{V_B}\right)^{0.953}. \tag{3.39}$$

This is a modified form of the combinatorial activity coefficient of the Flory-Huggins model. The use of an exponent was proposed by Donahue and Prausnitz [43] to tweak the Flory-Huggins model, where the exponent is 1, towards the ideal solution model, where the exponent is zero. This modification, however, has no physical
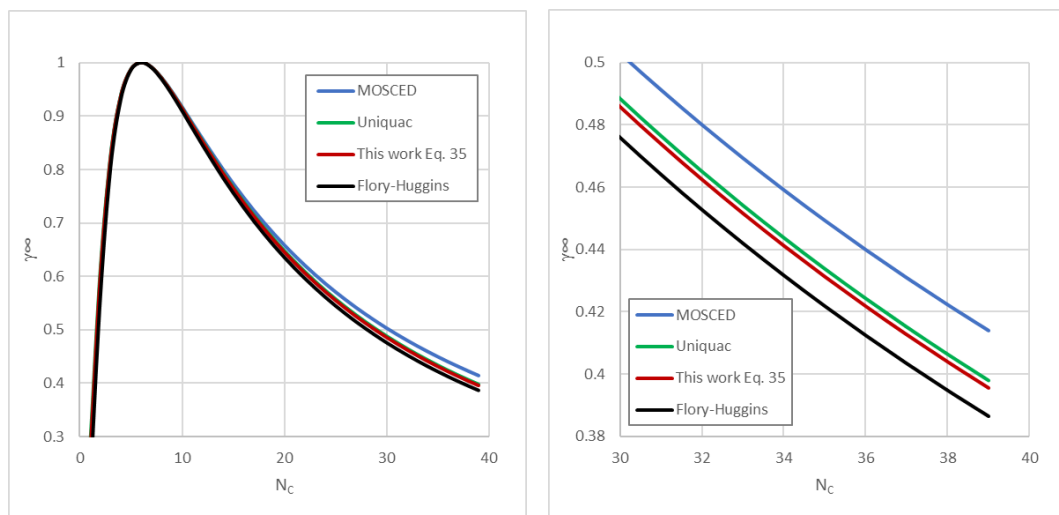
Figure 3.2: Comparison of models to predict the combinatorial activity coefficient of hexane in n-alkanes at infinite dilution as function of the carbon number. Right graph is a detail of the left graph to emphasize differences.

basis. Fig. 3.2 shows that this correction of the combinatorial term is larger than the Stavermann-Guggenheim correction used in the UNIQUAC model and also larger than the correction given by Eq. 3.36. Since Eq. 3.36 is a generalized form of the UNIQUAC model and has a physical meaningful basis rather than being an ad-hoc modification, it could replace the combinatorial term in the MOSCED model. However, this implies that the dispersion contribution needs to be made more repulsive in order to bring the MOSCED model in agreement with the experimental activity coefficients of alkane mixtures.

The essence of the derivation, which leads to the generalized expression for the combinatorial activity coefficient, is the introduction of a reference volume, in which only a limited number of molecules fit due to their size and shape. The probability to fill this volume with one type of molecules is the second essential step in the derivation. Because of the existence of a gas-liquid equilibrium, it is permitted to relate these probabilities to the probabilities to fill a reference volume in the vapor phase. This relation in combination with the Gibbs-Duhem law leads to a general expression of the Gibbs energy of mixing and subsequently a general expression for the combinatorial activity coefficient.

Chapter 4

On the Number of Nearest Neighbors[1]

4.1 Introduction

In this chapter we show how the generalized Guggenheim combinatorial activity coefficient equation, Eq. 3.36 and the formula for the number of nearest neighbors, Eq. 3.35 can be implemented in the UNIQUAC and the COSMOSPACE models. The new expressions are compared to the original form used in UNIQUAC and COSMOSPACE. In this comparison we will also consider the effect of the molecular size and shape defined by the Pauling bond lengths and the set of van der Waals radii, defined by Rowland and Taylor [190], respectively. Subsequently, the revised UNIQUAC and COSMOSPACE models are evaluated by comparing the description of vapor-liquid phase equilibria and excess enthalpy for alkane-alcohol binary systems.

4.2 Theory

4.2.1 The generalized and the Gugenheim Staverman combinatorial activity coefficient equations

In chapter 3 we showed how a generalized expression for the combinatorial activity coefficient can be derived, Eq. 3.36. The expression that was derived for the original

UNIQUAC model [158] is:

$$\ln(\gamma_k^{\text{comb}}) = \left[\ln(\frac{\phi_k}{x_k}) + 1 - \frac{\phi_k}{x_k}\right] - \frac{zq_k}{2}\left[\ln(\frac{\phi_k}{\theta_k}) + 1 - \frac{\phi_k}{\theta_k}\right], \tag{4.1}$$

where x_k, ϕ_k and θ_k are respectively the mole, volume and surface area fraction of compound k in the mixture and q_k is the relative surface area of this compound. The latter is obtained by defining a reference area. In Eq. 4.1 we recognize as first term the Flory-Huggins contribution and as second term the Staverman-Guggenheim correction [112, 126]. The coordination number in the UNIQUAC model z is set as a constant and has often the value 10. When we define the term Q_k, which is half the number of nearest neighbors of compound k, we can write:

$$Q_k \equiv \frac{zq_k}{2} = \frac{1 - r_k}{1 - \frac{r_k}{q_k}}. \tag{4.2}$$

As was proven in chapter 3, but also by comparison of Eq. 4.1 with Eq. 3.36, we have:

$$Q_k = \frac{1 - \frac{\phi_k}{x_k}}{1 - \frac{\phi_k}{\theta_k}}. \tag{4.3}$$

This tells us that Q_k can be calculated from the mole, area and volume fraction, and that zq_k is not a free parameter anymore, as it was considered in the original UNIQUAC model. In appendix B we show that Eq. 4.3 yields a constant, despite the fact that it contains functions of mole fraction. In order to apply Eq. 4.3 in the original UNIQUAC model it was required to define a reference area and a specific lattice coordination number z. In the past this Q_k was obtained by setting $z = 10$ and by calculating q_k using a reference area. We remark that the reference volume $V_{\text{ref}} = 25.17$ Å^3 was never changed after Abrams and Prausnitz [2] introduced it. The reference area, however, has been optimized several times in order to bring the activity models more in agreement with experimental results. Table 4.1 shows several values for the reference area for different models, in which the SG-combinatorial correction term was implemented. In here the lattice coordination number was always fixed on the value $z = 10$, as was proposed by Abrams and Prausnitz [2]. We remark that with smaller values of the reference area, the number of nearest neighbors of a compound increases. This has little impact on the combinatorial term, but has large effect on the residual activity coefficient. In fact, the binary interaction parameter needs to be adjusted to lower values in order to keep the description of the experimental phase equilibria data in agreement.

Author	$A_{\text{ref}}(\text{Å}^2)$	Model
Abrams and Prausnitz[2]	41.56	UNIQUAC
Wang *et al.*[229]	79.53	COSMOSAC
Soares [205]	124	COSMOSAC
Bronneberg and Pfennig [21, 22]	3.21	MOQUAC

Table 4.1: Values of the reference surface areas A_{ref} proposed by various groups.

4.2.2   Lattice independent UNIQUAC

In chapter 3 is was shown that choosing a reference area and a lattice coordination number for the Stavermann-Guggenheim expression for the combinatorial activity coefficient is no longer required, and that Eq. 4.3 defines the combinatorial part of the activity coefficient model. Substituting Eq. 4.3 into the original combinatorial activity coefficient, Eq. 4.1, gives

$$\ln(\gamma_{\text{k}}^{\text{comb}}) = \ln(\frac{\phi_{\text{k}}}{x_{\text{k}}}) + (1 - \frac{\phi_{\text{k}}}{x_{\text{k}}}) \left[\frac{\ln(\frac{\phi_{\text{k}}}{\theta_{\text{k}}})}{1 - \frac{\phi_{\text{k}}}{\theta_k}} \right] = \ln(\frac{\phi_{\text{k}}}{x_{\text{k}}}) - Q_k \ln(\frac{\phi_{\text{k}}}{\theta_{\text{k}}}). \qquad (4.4)$$

The latter equation is referred to Vera *et al.* [224] as the restricted form of the UNIQUAC equation. However, we point out that Vera *et al.* [224] apply Eq. 4.2, while we use Eq. 4.3, which requires no definition of the reference area A_{ref}. We mention that in the limit $\theta_{\text{k}} \to \phi_{\text{k}}$, the term between the square brackets becomes unity, and the Flory-Huggins equation is obtained. This also happens when we set $q_{\text{j}} = r_{\text{j}}$, which by Eq. 3.37 gives $r_{\text{j}} = q_{\text{j}} = 1$ or $z = \infty^2$, from which it follows that $\theta_{\text{j}} = \phi_{\text{j}}$. The derived equation for Q_{k} is also directly applicable in the residual part of the UNIQUAC model. It is noted that the residual part of the original UNIQUAC model contains an inconsistency. Instead of the frequently published equation, our analysis leads to

$$\ln(\gamma_{\text{k}}^{\text{res,UQ}}) = Q_k \left[1 - \ln(\sum_{\text{j}=1}^{M} \theta_{\text{j}}\tau_{\text{jk}}) - \sum_{\text{j}=1}^{M} \frac{\theta_{\text{j}}\tau_{\text{kj}}}{\sum\limits_{i=1}^{M} \theta_{\text{i}}\tau_{\text{ij}}} \right]. \qquad (4.5)$$

This expression contains a factor Q_{k} instead of q_{k}, which has been used in the original UNIQUAC model (see appendix C). In Eq. 4.5 τ_{ij} denotes the Boltzmann weighting factor given by the contact energy of a lattice cell of compound i and that

---

^{2}Note: $r_{\text{j}} = q_{\text{j}} = 1$ implies that all molecules have same shape. $z = \infty$ implies that molecules have no local molecular orientation and thus randomly distributed

of compound j

$$\tau_{ij} = \exp\left[-\frac{\Delta U_{ij}}{RT}\right]. \tag{4.6}$$

The total activity coefficient for the lattice independent UNIQUAC model can therefore be expressed as

$$\ln(\gamma_k^{UQ}) = \ln(\frac{\phi_k}{x_k}) - Q_k \ln(\frac{\phi_k}{\theta_k}) + Q_k \left[1 - \ln(\sum_{j=1}^{M}\theta_j\tau_{jk}) - \sum_{j=1}^{M}\frac{\theta_j\tau_{kj}}{\sum_{i=1}^{M}\theta_i\tau_{ij}}\right]. \tag{4.7}$$

The above principle holds also for the UNIFAC model. The difference between UNIQUAC and UNIFAC is that in the latter not molecules but molecule fragments are used as interacting entities in the residual part. Although UNIQUAC has been applied with success, we must not forget that the UNIQUAC model is based upon a first order approximation of the local concentration around a molecule in a mixture. This approach is accurate for systems having weak interactions, but for strongly interacting molecules the activity coefficient of the solute is underestimated. In that case the local activity instead of the local concentration of the solute should be used. This has been done in the COSMOSPACE [117] and GEQUAC [46] models.

4.2.3 Lattice independent COSMOSPACE

For a binary system, where each molecule can have two types of surfaces, the COSMOSPACE activity coefficient is expressed as [117]

$$\ln(\gamma_k^{res,CS}) = \frac{n_k}{2}\left(\zeta_k^A \ln\left[\frac{\gamma_{A,\underline{x}}}{\gamma_{A,k}}\right] + \zeta_k^B \ln\left[\frac{\gamma_{B,\underline{x}}}{\gamma_{B,k}}\right]\right), \tag{4.8}$$

where n_k is the total number of surface segments of molecule k, and ζ_k^J is the fraction of surface type J on molecule k. The $\underline{x} \equiv (x_1, x_2)$ is used to define the fractions of the two components in the mixture and the pure state. We mention that in the original publication the total number of surface patches is calculated by the relative area as defined in the UNIQUAC model: $n_k = 2q_k$. Since we consider a fully occupied lattice, i.e., there are no empty cells in the lattice, the total number of surface segments on a molecule has to be equal to the number of nearest neighbors. That is to say, we do not subdivide the contact area into smaller patches, as is done in COSMOSAC [145], because the patches of opposite segment sides make simultaneous contact, thereby

averaging the interaction energy of the two contacting surface segments. Therefore we consider

$$\frac{n_k}{2} = \frac{zq_k}{2} \equiv Q_k. \tag{4.9}$$

The variables $\gamma_{J,\underline{x}}$ and $\gamma_{J,k}$ in Eq. 4.8 denote the activity coefficient of surface type J in the mixture and the pure state, respectively. They can be calculated by

$$\ln(\gamma_{A,\underline{x}}) = \ln\left[\frac{1}{\theta_{A,\underline{x}}} + \frac{1 - \sqrt{(1 + 4\theta_{A,\underline{x}}\theta_{B,\underline{x}}\omega)}}{2\omega\theta_{A,\underline{x}}^2}\right], \tag{4.10}$$

and

$$\ln(\gamma_{B,\underline{x}}) = \ln\left[\frac{1}{\theta_{B,\underline{x}}} + \frac{1 - \sqrt{(1 + 4\theta_{A,\bar{x}}\theta_{B,\underline{x}}\omega)}}{2\omega\theta_{B,\underline{x}}^2}\right], \tag{4.11}$$

with $\omega = 1/\tau^2 - 1$ reflecting the interaction between the two different surface segments A and B, and where τ is the Boltzmann weighting factor of the COSMOSPACE model which is for weak interactions related to the interaction parameters of UNIQUAC; $\tau \approx \sqrt{\tau_{21}\tau_{12}}$. In the limit $\omega \to 0$ the surface activity coefficients become unity, and the binary system becomes an ideal solution. The surface fractions of the two components in the mixture and the pure state are calculated by

$$\theta_{A,\underline{x}} = \frac{x_1 q_1 \zeta_1^A + x_2 q_2 \zeta_2^A}{x_1 q_1 \zeta_1^A + x_2 q_2 \zeta_2^A + x_1 q_1 \zeta_1^B + x_2 q_2 \zeta_2^B}, \tag{4.12}$$

and

$$\theta_{B,\underline{x}} = \frac{x_1 q_1 \zeta_1^B + x_2 q_2 \zeta_2^B}{x_1 q_1 \zeta_1^A + x_2 q_2 \zeta_2^A + x_1 q_1 \zeta_1^B + x_2 q_2 \zeta_2^B}. \tag{4.13}$$

The total activity coefficient of the lattice independent COSMOSPACE model can be expressed by the sum of Eqs. 4.4 and 4.8

$$\ln(\gamma_k^{CS}) = \ln(\frac{\phi_k}{x_k}) - Q_k \ln(\frac{\phi_k}{\theta_k}) + Q_k\left[\zeta_k^A \ln\left[\frac{\gamma_{A,\underline{x}}}{\gamma_{A,k}}\right] + \zeta_k^B \ln\left[\frac{\gamma_{B,\underline{x}}}{\gamma_{B,k}}\right]\right] \tag{4.14}$$

4.3 Results and discussion

In this section we evaluate the accuracy of the renewed expression for half the number of neighbors, Eq. 4.3. For this purpose we chose binary mixtures of n-alkanes and 1-alcohols. Subsequently, we investigate how the UNIQUAC and COSMOSPACE models perform using the new definition for the number of nearest neighbors.

4.3.1 The number of nearest neighbors of n-alkanes

The calculation of the number of nearest neighbors by Eq. 4.3 for binary mixtures of alkanes between methane and nonacosane has been carried out for a series of 200 mole fractions between 0.001 and 0.999. Outside this range Eq. 4.3 suffers from computational inaccuracy due to the vanishing denominator, approaching zero. To calculate the volume and surface fractions in the binary mixture we used the van der Waals volume (V_{vdW}) and surface area (A_{vdW}) given by Bondi [18]. For n-alkanes Bondi gives the following relations

$$V_{vdW} = \frac{6.88 + 10.23N_C}{0.6022},$$
(4.15)

and

$$A_{vdW} = \frac{1.54 + 1.35N_C}{0.06022},$$
(4.16)

where we kept the original Bondi parameters and used the factors 0.6022 and 0.06022 to scale from molar size (cm^2/nmol, and cm^3/mol) to molecular area (Å^2) and volume (Å^3), respectively. As can be seen from the equidistant lines in Fig. 4.1, there is a clear linear relationship between zq_k and N_C. Methane, depicted at the bottom line, has a constant value for zq_k over the entire range from ethane to n-nonacosane. The same holds for the higher alkanes. The explanation for this is given in B. The open space in Fig. 4.1 represents the position of the pure liquid, to which Eq. 4.3 does not apply. However, since the values on the left and the right side of the open space are the same, it seems logical to assign this value as zq_k of the pure compound. This is an important observation, which we will use later. Based on the results depicted in Fig. 4.1, the relation between the number of nearest neighbors of the pure compound and the alkane number is given by

$$zq_k = 4.87 + 4.27N_C.$$
(4.17)

Sayegh and Vera [195, 224] proposed the following empirical relation for the number of nearest neighbors based on an analysis of 15 classes of compounds

$$zq_k = 0.4228V_{vdW} + 2(1 - l_k),$$
(4.18)

where l_k is the number of internal contacts in a molecule and the van der Waals volume is expressed in cm^3/mol. The empirical constant 0.4228 has the dimension mol/cm^3 and is equal to the quantity $(z - 2)/V_{ref}$. We see that the value for the reference volume is dependent on the value of z. E.g. if $z = 10$, then the reference volume is 18.92 cm^3/mol or 31.4 Å^3/molecule, which gives $A_{ref} = 48.2$ Å^2/molecule, which is 20% higher than applied in the original UNIQUAC model (see Table 4.1).
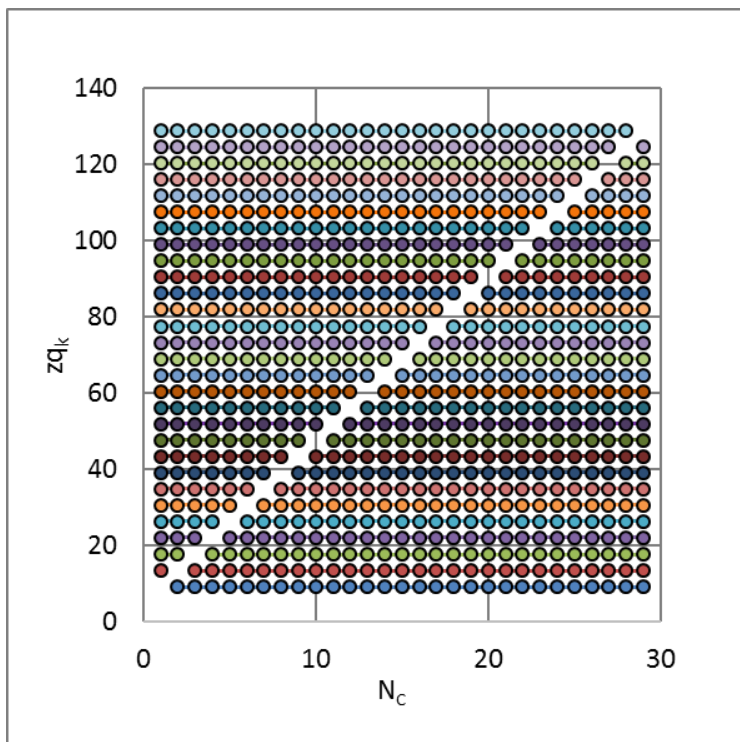
Figure 4.1: The number of nearest neighbors (zq_k) of two n-alkanes between methane and nonacosane calculated by Eq. 14 using the van der Waals volumes and areas of Bondi [18], as has been shown in appendix B. In the bottom row the zq of methane in a mixture of ethane ($N_C = 2$) to nonacosane ($N_C = 29$) is shown, continuing to the top row where the zq of nonacosane in mixture of methane ($N_C = 1$) to octacosane ($N_C = 28$) is given.

For n-alkanes, where $l_k = 0$, and using the van der Waals volumes defined by the van der Waals radii of Bondi, Eq. 4.18 becomes

$$zq_k = 4.91 + 4.33 N_C. \tag{4.19}$$

This relation is close to Eq. 4.17. Although the van der Waals radii as proposed by Bondi are often used in the calculation of molecule dimensions, they are still matter of debate. We refer to the review article of Batsanov [11], in which it is argued that the van der Waals radii of Rowland and Taylor [190] should be considered as more reliable, because a much larger experimental XRD-data set has been used. The van der Waals radii of Rowland and Taylor for hydrogen, carbon and oxygen are $R_H = 1.09$ Å, $R_C = 1.75$ Å, and $R_O = 1.56$ Å, respectively. With these values and the Pauling bond lengths [176] the van der Waals volume and surface area can be
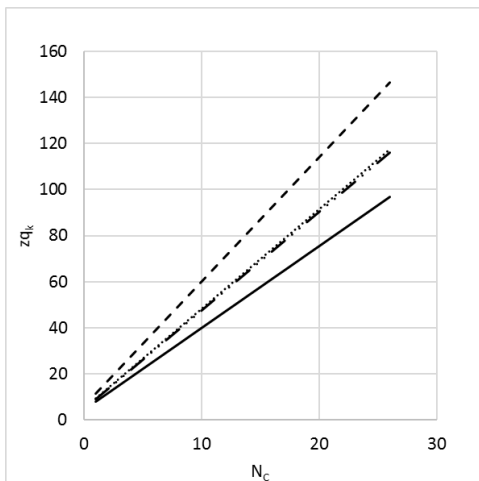
Figure 4.2: The number of nearest neighbors (zq_k) of n-alkanes calculated by Eq. 4.20 (solid line), Eq. 4.17 (dashed-dotted line), Eq. 4.18 (dotted line) and Eq. 4.21 (dashed line).
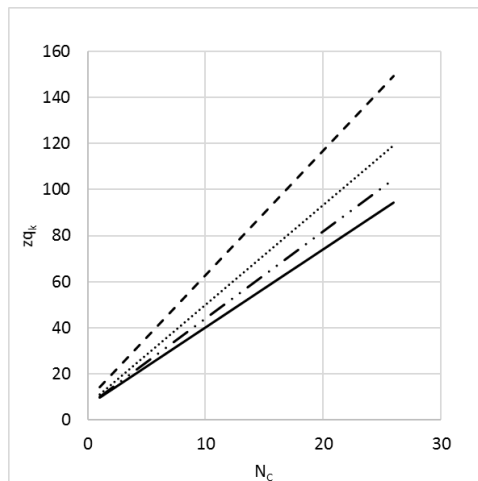
Figure 4.3: The number of nearest neighbors (zq_k) of 1-alcohols calculated by Eq. 4.23 (solid line), Eq. 4.22 (dashed-dotted line), Eq. 4.18 (dotted line) and Eq. 4.21 (dashed line).

calculated. Subsequently, using the same procedure as before, the number of nearest neighbors is obtained

$$zq_k = 4.37 + 3.55 N_C. \tag{4.20}$$

For comparison we recall that in the original UNIQUAC model the number of nearest neighbors is calculated by

$$zq_k = z \frac{A_{vdW}}{A_{ref}} = \frac{A_{vdW}}{4.15}, \tag{4.21}$$

where z is set to 10, $A_{ref} = 41.5$ Å^2 and A_{vdW} is calculated with the Bondi set of van der Waals radii (i.e. Eq. 4.16). Figure 4.2 shows a comparison of the aforementioned equations for zq_k of n-alkanes. We see that the set of van der Waals radii of Rowland and Taylor, i.e. Eq. 4.20, gives the lowest values for zq_k, while Eq. 4.21 gives significantly higher values, which might lead to systematic deviations in the UNIQUAC model.

4.3.2 The number of nearest neighbors of 1-alcohols

Another class of linear molecules are the 1-alcohols. The required van der Waals radii of Bondi [18] and those of Rowland are (in Å): $R_H = 1.20$, $R_C = 1.70$ and

$R_O = 1.52$, and $R_H = 1.09$, $R_C = 1.75$ and $R_O = 1.56$, respectively. Using the same method as described in the section on n-alkanes, this leads to

$$zq_k = 6.24 + 3.78N_C \tag{4.22}$$

for the Bondi parameter set of van der Waals radii. The set of van der Waals radii of Rowland and Taylor yields

$$zq_k = 6.14 + 3.39N_C \tag{4.23}$$

Results of the various equations for zq_k of the 1-alcohols are depicted in Fig. 4.3.

Like in the case of the n-alkanes, Eq. 4.3 gives for the alcohol mixtures lower values for the number of nearest neighbors than those calculated by the original UNIQUAC formula. Comparing Fig. 4.2 with Fig. 4.3 it follows that the relation based on the van der Waals radii of Rowland and Taylor are more in line with the expectation that the alcohols have more nearest neighbors than the alkanes with the same carbon number, as long as $N_C < 11$. This suggests that further optimization of the van der Waals radii is still needed. In fact, we expect that the slope of the number of nearest neighbors as function of the carbon number of the alcohols should be the same as that of the alkanes, because in both cases this only depend on the number of the CH_2 groups. Batsanov [11] indicated that the concept of spherical atoms in a molecule probably needs to be adjusted to ellipsoids, because the radius in the bond direction is smaller that the one perpendicular to it, which might also be important in a further refinement.

4.3.3 The number of nearest neighbors of mixtures of n-alkanes and 1-alcohols

While zq_k of an alkane is independent of the other alkanes in the mixture, it becomes N_C dependent when an 1-alcohol is added. Fig. 4.4 shows the case of n-hexane
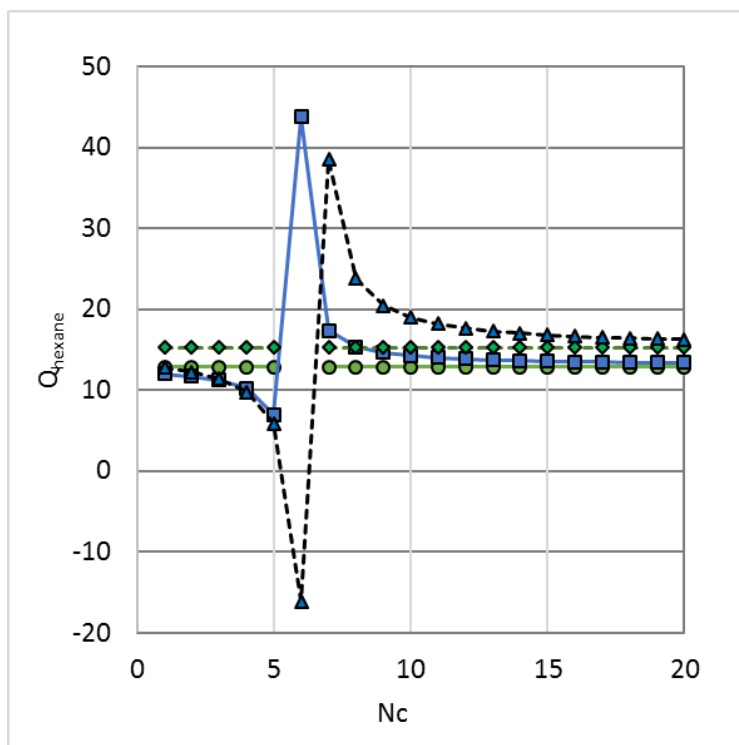
Figure 4.4: Results for Q_k of hexane dissolved in alkanes (green dots/diamonds) and alcohols (blue squares/triangles) with Eq. 4.3 using Rowland's set (solid curves) or Bondi's set (dashed curves) of van der Waals radii.

dissolved in an n-alkane and in a 1-alcohol. Although the number of nearest neighbors for n-hexane in a 1-alcohol does not vary with concentration, we observe that it becomes 1-alcohol dependent over a small range of N_C-values. When the carbon number of the alcohol is low, the number of nearest neighbors of n-hexane is slightly below the value of that in an alkane. In the case of Bondi's set of van der Waals radii it even gives negative values when the difference in carbon number becomes smaller, which is unphysical. This artefact is caused by small errors in the calculation of the van der Waals volume and area, which is magnified by the denumerator of Eq. 4.3. This artefact would also occur in the original UNIQUAC method, if the

developers had use the right part of Eq. 4.2, instead of the applied $zq/2$, for which an ad-hoc value of the coordination number ($z = 10$) and the reference area ($a_{\text{ref}} = 41.56$ Å^2) had to be defined. Above the carbon number of hexane the number of nearest neighbors approaches the pure component value. This makes sense, because the alcohol starts to increasingly resemble an alkane. The presence of an apparent singularity around molecules of equal size, however, demonstrates that the choice for a value of the number nearest neighbors is still a critical issue. In order to make the new expression for the UNIQUAC model applicable for all types of solvent mixtures, we take the value obtained from the calculation of the compound in a mixture of molecules of its own class. This is equivalent to the number of nearest neighbors of the compound in its pure state. This choice implies that the coordination of other molecules in the first shell around a central molecule is set equal to that of the pure state.

4.3.4   The Staverman-Guggenheim correction term

We will now compare the combinatorial activity coefficient using the different expressions for the number of nearest neighbors and apply these to the binary system hexane - hexadecane. Fig. 4.5 shows the comparison between the combinatorial activity coefficient of the different models and parameter sets for the binary system of n-hexane (1) and n-hexadecane (2). The activity coefficients of the Flory Huggins model, for which we used the original UNIQUAC parameters $r_1 = 4.5$ and $r_2 = 11.25$, are plotted as dashed curves, which end at the limiting activity coefficient values of $\gamma_1^{\text{comb},\infty} = 0.729$ and $\gamma_2^{\text{comb},\infty} = 0.558$. The different Staverman-Guggenheim models give curves which nearly coincide. The original UNIQUAC model, where $z = 10$, $q_1 = 3.86$ and $q_2 = 9.26$, gives curves that end at $\gamma_1^{\text{comb},\infty} = 0.7404$ and $\gamma_2^{\text{comb},\infty} = 0.5798$. The refined Staverman-Guggenheim model, i.e. Eq. 4.4 with Q_k defined by Eq. 4.3, gives using the van der Waals molecular volume and area based on the radii set of Bondi: $\gamma_1^{\text{comb},\infty} = 0.7380$ and $\gamma_2^{\text{comb},\infty} = 0.5752$, while the set of Rowland and Taylor gives: $\gamma_1^{\text{comb},\infty} = 0.7404$ and $\gamma_2^{\text{comb},\infty} = 0.5798$. These values demonstrate that the difference between the original and the refined UNIQUAC models is negligibly small; less than 0.1%. This observation shows that the new equation for the number of nearest neighbors (Eq. 4.3) gives nearly the same result as the equation used in the original UNIQUAC equation (Eq. 4.2) for which a reference surface area had to be defined. Our approach demonstrates that the lattice model of UNIQUAC can be transformed into a fluid model without the definition of a reference area. This transformation is also observed in the Flory-Huggins model. Initially the number of lattice cells for each molecule was defined, but in the end the activity coefficient equation contains only mole and volume fractions, for which this
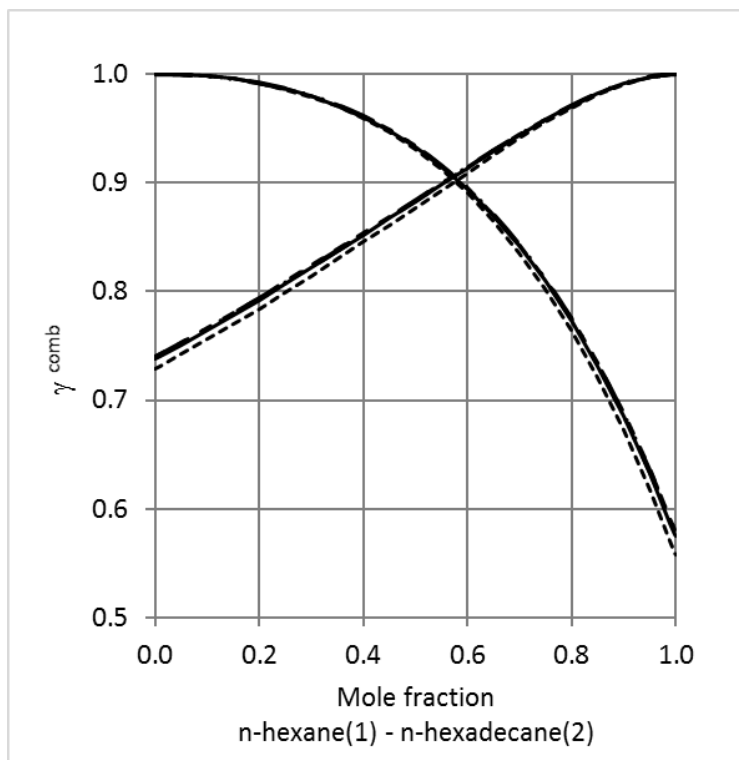
Figure 4.5: Combinatorial activity coefficient γ^{comb} of the binary n-hexane and n-hexadecane. Flory-Huggins model (small dashed curve, $\gamma_1^{\mathrm{comb},\infty} = 0.729$). The curves of the original UNIQUAC and the refined Staverman-Guggenheim models based on Bondi's and Rowland's set of van der Waals radii nearly coincide.

definition is not needed. Likewise, the Staverman-Guggenheim correction requires the definition of the number of sites of each molecular surface, i.e. $zq/2$, but this is later replaced with Eq. 4.3, and the activity becomes a function of mole, volume and surface fractions only. In other words, in both cases the concept of a lattice has served as a vehicle to reveal the dependency between activity coefficient and concentrations.

4.3.5   Activity coefficient models

UNIQUAC

In the theoretical section of this work (see appendix C) we show that the residual part of the original UNIQUAC model contains an error. The correct derivation

delivers a residual activity coefficient equation with Q_k and not q_k as molecular dependent variable. Fig. 4.6 depicts several experimental isothermal vapor-liquid equilibria data and the UNIQUAC descriptions in three different ways: the original UNIQUAC model (dashed curves), and the lattice independent UNIQUAC model as given in Eq. 4.7 using Eq. 4.3. The optimized UNIQUAC interaction parameters are given in Table 4.2. The key quantity Q_k is calculated by the volume and the

Table 4.2: UNIQUAC binary interaction parameters for n-hexane and a 1-alcohol using number of nearest neighbors of the pure compound as given in Figs. 4.2 and 4.3.

	Original		Bondi		Rowland	
1-Alcohol (T °C)	τ_{12}	τ_{21}	τ_{12}	τ_{21}	τ_{12}	τ_{21}
Methanol (25)	0.129	1.017	0.237	1.733	0.165	1.814
Ethanol (45)	0.161	1.289	0.282	1.785	0.261	1.787
Propanol (50)	0.395	1.898	0.432	1.568	0.397	1.578
Pentanol (30)	0.370	1.293	0.528	1.477	0.431	1.601

surface area fractions defined by the molecular volumes and areas from Bondi's (solid curves) and Rowland's (dashed-dotted curves) set of van der Waals radii. We observe in Fig. 4.6 that the three different forms of the UNIQUAC model give equivalent descriptions of the vapor-liquid equilibria. The new equation somewhat overpredicts the pressure in the case of ethanol, and gives lower pressures in the case of pentanol. These deviations are not caused by the new expression, but by the combination of parameters and the first order approximation of the local concentration. This is more clear in the description of the binary n-hexane and ethanol, where all three models yield a heterogeneous azeotrope, while the experimental data show a homogeneous azeotrope. The exaggeration in pressure exists for all UNIQUAC models, but is more present in case of the lattice independent versions. The left plot of fig. 4.7 shows that the lattice independent UNIQUAC models yield higher activity coefficients than the original UNIQUAC model; for hexane at mole fractions between 0.1 and 0.5 and for ethanol at mole fractions between 0.7 and 0.9. This difference between the two types of UNIQUAC models is caused by the small q_k in the original UNIQUAC model versus the applied correct factor Q_k in the lattice independent models; $q_{\text{Ethanol}} = 1.97$, $q_{\text{Hexane}} = 3.86$, and Rowland's data give $Q_{\text{Ethanol}} = 6.48$ and $Q_{\text{Hexane}} = 12.85$.

To compare the parameters of the different UNIQUAC models, it is convenient to take the mean interaction parameter (i.e. $\tau_{298} = \sqrt{\tau_{12}\tau_{21}}$) at 25 °C. Table 4.3 shows that with increasing alkane number of the alcohol the binary system becomes more ideal; the mean interaction parameter shifts to unity and the interaction energy, calculated by $\ln(\tau_{298})$, goes to zero.

The interaction parameters for the original UNIQUAC model are the smallest in

2ndCompound	τ_{298}			ΔE_{298}(kJ/mol)		
	UQ	Bondi	Rowl.	UQ	Bondi	Rowl.
Methanol	0.362	0.641	0.547	2.52	1.10	1.50
Ethanol	0.456	0.710	0.683	1.91	0.85	0.95
1-Propanol	0.584	0.810	0.776	1.33	0.52	0.63
1-Pentanol	0.687	0.881	0.828	0.93	0.31	0.47

Table 4.3: Mean UNIQUAC interaction parameter, τ_{298}, and interaction energy, ΔE, values for the binary pair n-hexane - alcohol at 298.15 K. UQ denotes original UNIQUAC model. Bondi denotes the lattice independent UNIQUAC model with Bondi's basis set of van der Waals radii. Rowl. denotes the lattice independent UNIQUAC model with Rowland's basis set of van der Waals radii.

value, followed by those of the lattice independent model based on Rowland's and Bondi's basis set of van der Waals radii. This is in line with the fact that the original UNIQUAC model uses the smallest value for the relative area, while in the refined models the van der Waals radii of Bondi give a larger molecular surface area than those of Rowland.

COSMOSPACE

In the article on the COSMOSPACE model [117] it was already made clear that poor results are obtained when the molecular surface of each compound is considered to be made of one type of surface. This so-called homogeneous surface approach fails, because in reality each interacting molecule, even an alkane, consists of at least two different types of surfaces. Therefore the most comprehensive form of the COSMOSPACE model is the so-called non-homogeneous double-binary set-up. We note that more surface types of interaction can be taken into account, but this requires more computation [117]. For this work, however, we can show that the non-homogeneous double-binary set-up is sufficient to demonstrate that Eq. 4.3 provides an accurate description for a series of vapor-liquid equilibria of hexane and an alcohol. For the description of the isothermal experimental vapor-liquid equilibria we have used two parameters: the number of nearest neighbors of the hydroxyl group (n_{OH}) in the alcohol and the alkyl-hydroxyl interaction parameter (τ_{298}) at the reference temperature of 298.15 K. An Arrhenius equation is used to define the interaction parameter τ at other temperatures

$$\ln(\tau) = -\frac{\Delta E_{298}}{RT} = \frac{298.15}{T}\ln(\tau_{298}), \qquad (4.24)$$

where ΔE_{298} is the interaction energy between the alkyl and the hydroxyl surfaces at 298.15 K. Simultaneous optimization of the isothermal vapor-liquid equilibria yields

COSMOSPACE Model Parameter	Original Bondi	New Bondi	New Rowland
$Q_{OH}(-)$	-	1.18	1.25
$n_{OH}(-)$	2.29	2.35	2.50
$\tau_{298}(-)$	0.0979	0.0409	0.0471
ΔE_{298}(kJ/mol)	5.76	7.92	7.58
$n_{OH}\Delta E_{298}$(kJ/mol)	13.4	18.6	18.9

Table 4.4: COSMOSPACE parameters for alkane-1-alcohol systems derived from fitting VLE data (table 4.5).

to a set of parameter values listed in Table 4.4. Figure 4.8 shows that the COSMOSPACE model gives an excellent description of the phase equilibria data. The two lattice independent COSMOSPACE models show hardly any difference between the results obtained with the van der Waals radii of Bondi and Rowland; the curves almost coincide. Statistical information in the form of the average absolute deviation indicates that Bondi's set of van der Waals radii gives slightly better results than Rowland's set (See table 4.5).

Binary VLE Alkane-Alcohol	T °C	NPS -	CS (Original) Bondi	CS (New) Bondi	CS (New) Rowland	lit.
Hexane-Methanol	25	4	18.7	3.5	5.1	[100]
Hexane-Ethanol	40	38	2.9	1.2	2.4	[173]
Heptane-Ethanol	30	50	6.4	3.3	4.1	[16]
Nonane-Ethanol	25	25	3.8	0.9	1.6	[16]
Hexane-1-Propanol	50	20	3.2	1.4	2.3	[40]
Hexane-1-Butanol	60	24	0.8	1.5	2.4	[15]
Hexane-1-Pentanol	40	13	0.9	1.7	2.2	[189]
Hexane-1-Hexanol	30	21	7.2	4.1	5.4	[235]
Average			4.3	2.1	3.1	

Table 4.5: Average absolute deviation in pressure: $100 \times |1 - P_{calc}/P_{exp}|$ of the various COSMOSPACE (CS) models. NPS = number of mixture data points in regression. Pure data points are used to scale pressure.

In comparison to the UNIQUAC approach the COSMOSPACE model clearly gives a better quantitative description of the experimental vapor-liquid phase equilibria, especially the transition from heterogeneous to homogeneous azeotrope is captured well. This is caused by the better activity coefficient description in the mid concentration range as can be observed by comparing the two plots in figure 4.7. The description of the experimental data is best performed by the lattice independent COSMOSPACE model with Bondi's set of van der Waals radii. With the number of contacts associated with hydrogen bonding, $n_{OH} = 2Q_{OH}$, the change in pair interac-

tion energy is calculated. The obtained values of 13.4, 18.6 and 18.9 kJ/mol indicate that the lattice independent model is closer to the energy of hydrogen bonding of alcohols (20 kJ/mol for methanol/ethanol [90]) than the original COSMOSPACE model.

In Fig. 4.9 we depict the prediction of the excess enthalpy of the binary systems n-hexane - ethanol at 40 °C [173] and n-nonane - ethanol at 25 °C [31], for which the parameters are taken from the vapor-liquid equilibria optimization. The prediction is qualitatively very good with respect to the location of the maximum. The average

Binary H^{E}	T	NPS	CS (Original)	CS (New)	CS(New)	
Alkane - Alcohol	°C	-	Bondi	Bondi	Rowland	lit.
Hexane - Ethanol	40	21	15	9	9	[173]
Heptane - Ethanol	30	26	14	14	15	[203]
Nonane - Ethanol	25	26	14	21	22	[31]
Average			14	15	15	

Table 4.6: Absolute average deviation in excess enthalpy: $100 \times |1 - H^{\mathrm{E}}_{\mathrm{calc}}/H^{\mathrm{E}}_{\mathrm{exp}}|$ of the various COSMOSPACE (CS) models.

absolute deviation of the predicted excess enthalpy is about 15%. This deviation is understandable from the fact that the induced dipole interaction between the alkyl parts of the alkane and the alkanol has not been taken into account in the calculations. It is expected that the inclusion of a dispersion term, as has been done in the DISQUAC [109, 110], will improve the prediction of the excess enthalpy. Figure 4.10 shows the fraction of the number of nearest neighbors, ζ^{OH}, that acts as hydroxyl surface according to the optimization of the COSMOSPACE parameter n_{OH} and the total number of nearest neighbors n_{k} given by Eq. 4.22 (dashed curve) and Eq. 4.23 (solid curve). For comparison also the ratio in the van der Waals surface area of the -OH group and the total molecule are plotted. The van der Waals surface areas of the hydroxyl group are 24.3 Å^2 and 23.7 Å^2, for the Bondi and Rowland sets of van der Waals radii, respectively. The total van der Waals surface areas of the 1-alcohols are given by the linear relations

$$A_{\mathrm{vdW}} = 37.03 + 22.41 N_{\mathrm{C}}, \qquad (4.25)$$

and

$$A_{\mathrm{vdW}} = 36.37 + 20.09 N_{\mathrm{C}}, \qquad (4.26)$$

for the Bondi and Rowland set of van der Waals radii, respectively. We observe that the COSMOSPACE model gives lower values for the fraction of hydrogen bonding area, than is predicted by the ratio of van der Waals surface areas. This is logical because only a part of the hydroxyl van der Waals surface area is involved in hydrogen

bond breaking. The original model overpredicts the surface area involved in hydrogen bonding, as a result of incorrect use of the q_k parameter, instead of Q_{k}. This also explains why the incrementation of the hydroxyl group in the UNIFAC method failed for methanol, which would consist of one methylene group and one hydroxyl group. In UNIFAC the surface area of the strongly interacting part of the hydroxyl group was set equal to the van der Waals surface of the hydroxyl group. Extrapolation towards methanol would make it a strongly hydrophilic compound. Therefore methanol was defined as a new UNIFAC group.
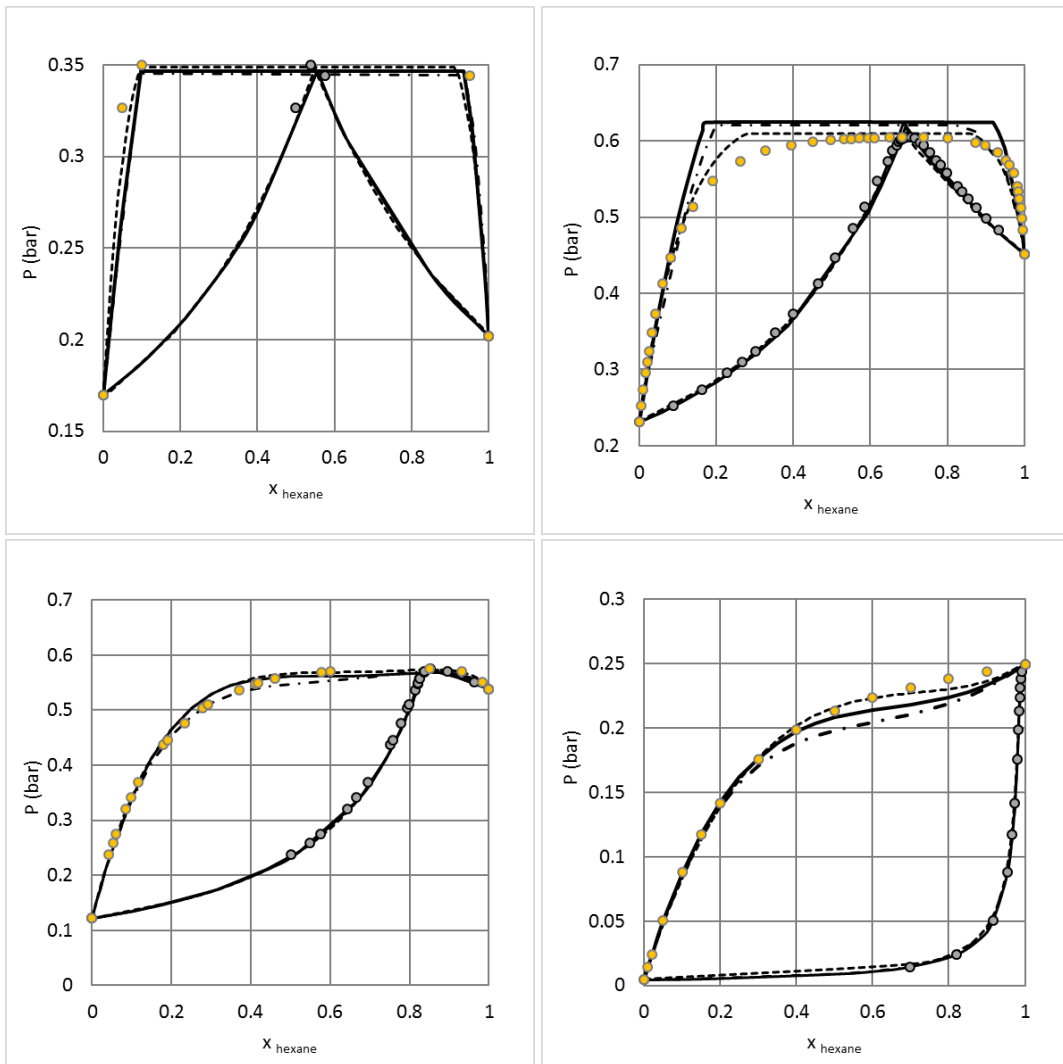
Figure 4.6: Experimental data and UNIQUAC description of binaries between n-hexane(1) and alcohol(2). From left to right, and from top to bottom respectively: methanol (25 °C, from [100]), ethanol(40 °C, from [173]), propanol(50 °C, from [40]) and pentanol(40 °C, from [189]). Dashed curves (Original UNIQUAC model), solid and dashed-dotted curve (Eq. 4.7) with Q_k based on Bondi's and Rowland's set of van der Waals radii, respectively.
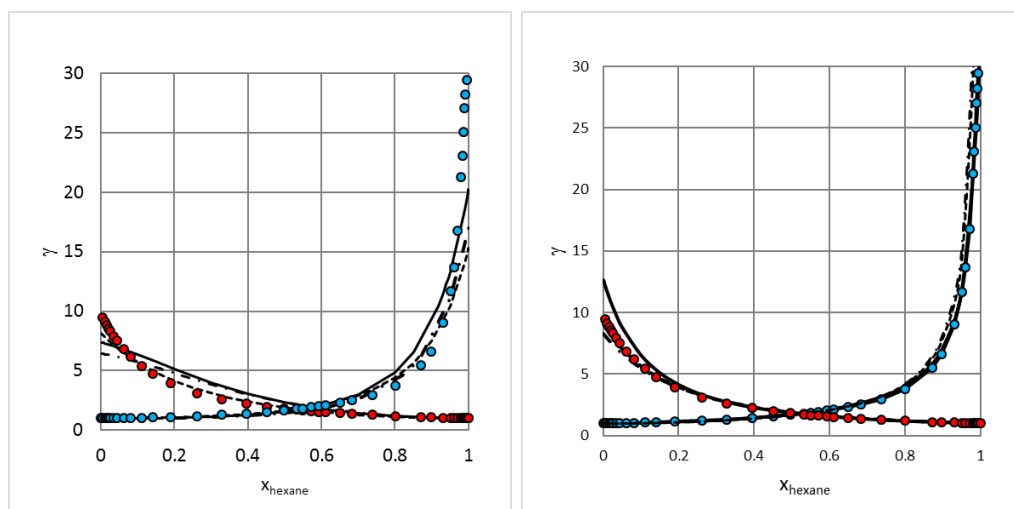
Figure 4.7: Activity coefficient of hexane-ethanol. Experimental data (symbols) [173]. Left plot: original Uniquac model (dashed curves), new UNIQUAC models with Bondi's (solid curves) and Rowland's (dashed dotted curves) set of van der Waals radii. Right plot: original COSMOSPACE model (solid curves), new COSMOSPACE models with Bondi's (dashed curves) and Rowland's (dashed dotted curves) set of van der Waals radii.
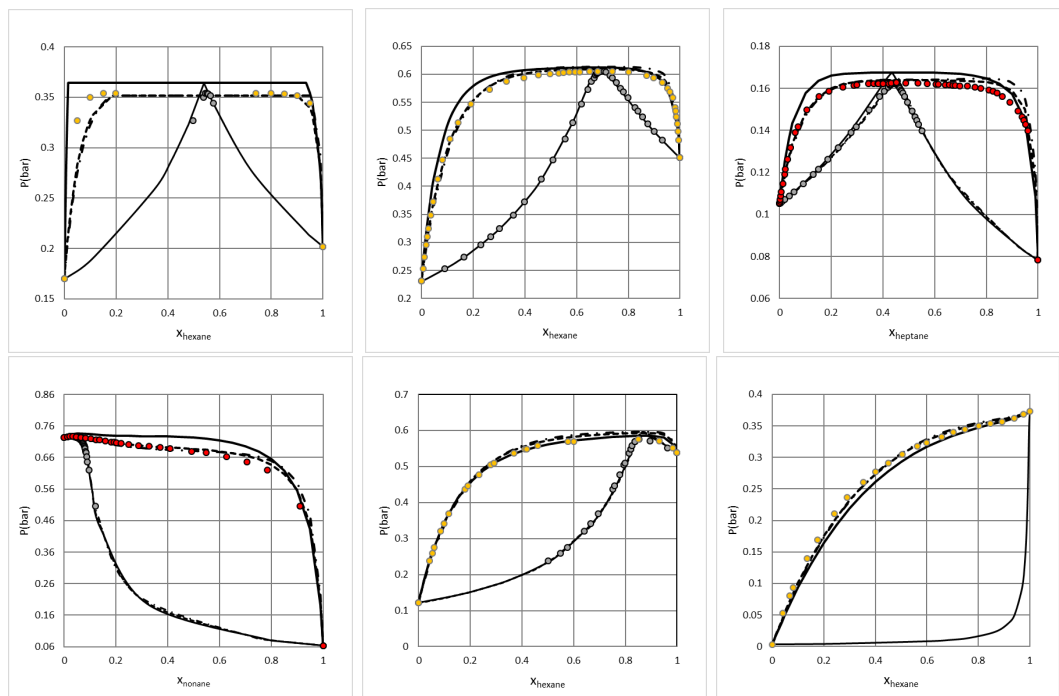
Figure 4.8: Experimental data (symbols) and COSMOSPACE description (curves) of vapor-liquid equilibria between alkane(1) and alcohol(2). From left to right and from top to bottom: hexane-methanol [100], hexane-ethanol [173], heptane-ethanol and nonane-ethanol [15]),hexane-propanol [40], and hexane-1-hexanol [235]. Solid curves represent the original COSMOSPACE model, the dashed and dashed-dotted curves are the lattice independent COSMOSPACE model based on Bondi's and Rowland's set of van der Waals radii, respectively.
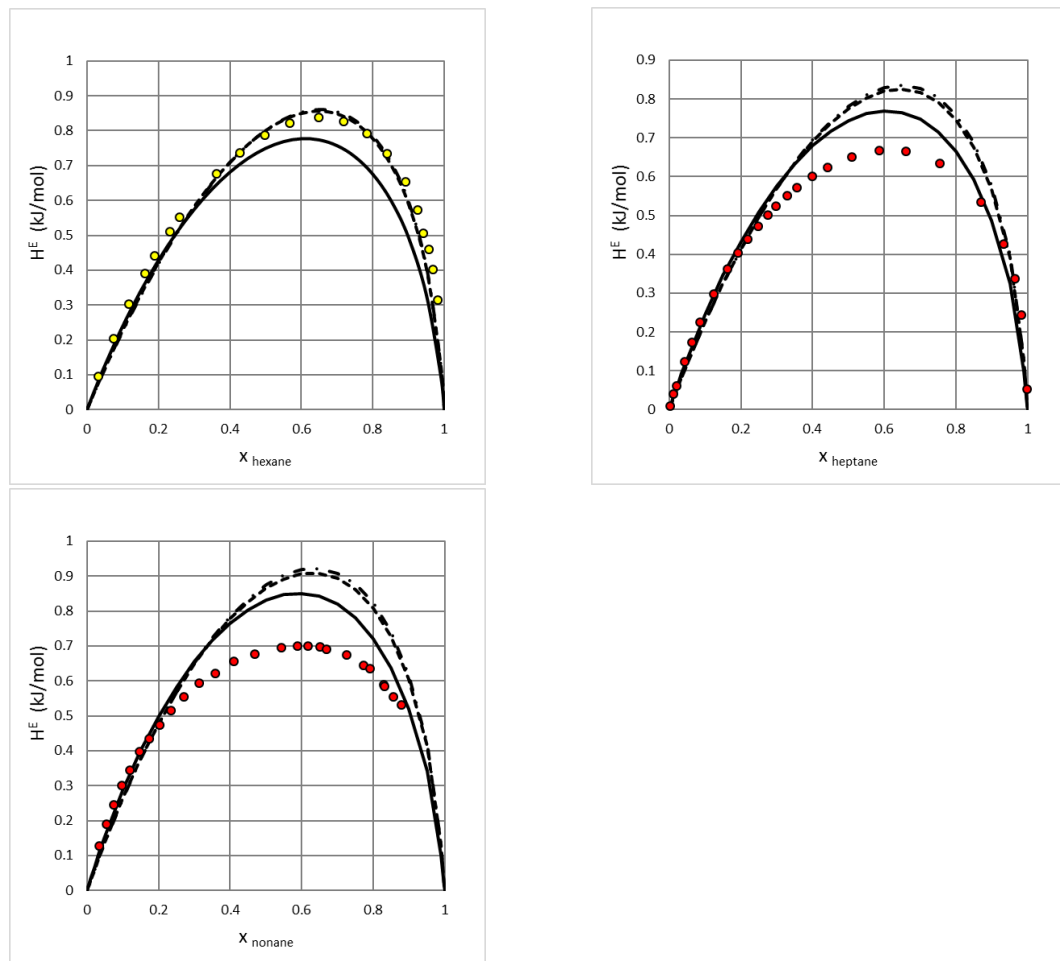
Figure 4.9: Predicted excess enthalpy of alkane-ethanol binaries with parameters from Table 4.4. Symbols represent experimental data. From top to bottom: hexane-ethanol [173], heptane-ethanol [203], and nonane-ethanol [31]. The solid curves represent the original COSMOSPACE model, the dashed and dashed dotted curves the lattice independent COSMOSPACE model based on Bondi's and Rowland's set of van der Waals radii, respectively.
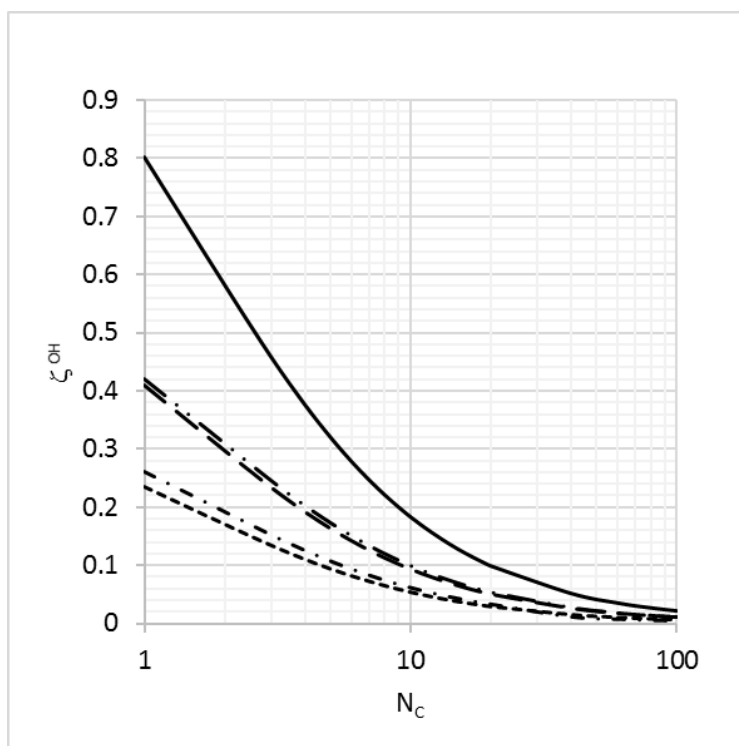
Figure 4.10: Fraction of hydroxyl surface to total surface area as function of alkane number. From bottom to top: Lattice-free COSMOSPACE model based on Bondi's (dashed curve) and Rowland's (dashed-dotted curve) set of van der Waals radii, geometric model using on Bondi's (long dashed curve) and Rowland's (dashed-double dotted curve) set of van der Waals radii, and original COSMOSPACE mocel (solid curve).

4.4 Concluding remarks

We have shown that two expressions for the combinatorial activity coefficient can be derived from Guggenheim's model for the entropy of mixing of linear molecules. These expressions yield a unique expression for the number of nearest neighbors of a molecule in a multicomponent mixture. This formula is a function of mole, volume and area fractions only, and makes a choice for the value for the lattice coordination number and the reference surface superfluous. In other words, in combination with the (revised) residual term as obtained for the COSMOSPACE model, it makes the total activity coefficient model lattice independent. We expect that the new formula also works for more complicated molecules, but this still needs to be verified.

We have demonstrated that the new expression for the number of nearest neighbors leads to lattice independent versions of the UNIQUAC and COSMOSPACE models. In order to put the various models in perspective, Table 4.7 shows a comparison indicating the lattice (in)dependency of the combinatorial and residual activity coefficients. In this table we use the label σ-profiles, by which we indicate the whole procedure starting with QM-calculations, the averaging of surface charge densities of molecules and the collection of surface area into histograms. Hence, the three-dimensional lattice information vanishes in these conversions prior to the calculation of the surface and residual activity coefficients.

For the UNIQUAC model we first had to remove an inconsistency from the residual term of the UNIQUAC model in order to show this possibility. The original and the lattice independent version of UNIQUAC describe the investigated vapor-liquid phase equilibria with the same quality and accuracy as before. A much better description of these phase equilibria is obtained with the lattice independent version of the COSMOSPACE model. For 1-alcohols mixed with n-hexane at moderate temperatures it captures the transition from a heterogeneous to a homogeneous azeotrope when changing from methanol to ethanol. The number of parameters is substantially lower than those for the original and the lattice independent version of the UNIQUAC model. Especially for strongly interacting systems the lattice independent COSMOSPACE model is recommended above the UNIQUAC model.

The new formula gives a constant value for the number of nearest neighbors of a compound dissolved in a mixture composed of molecules of the same class. This value stays constant over the entire concentration range, except at limiting concentrations, where computational accuracy plays a role. Because the value of the number of nearest neighbors of a compound is independent from the other molecules of the same class in the mixture, one can assign this value as the number of nearest neighbors of the pure compound. This has shown to be a useful choice, because sometimes unphysical values are obtained. This happens when one of the compounds

Model	Combinatorial term and parameters	Residual term and parameters	Lattice model?[a]
UNIQUAC	FH-SG[b], $Q = zq/2$ $z = 10$, A_{ref}, V_{ref}[g]	1st order solution of QCT[c], $Q = q$	C: Yes R: Yes
COSMOSPACE	FH-SG, $Q = zq/2$ UNIQUAC parameters	Self-consistent solution of QCT = CS-solution[d], $Q = n = 2q$	C: Yes R: No
COSMO-RS	Generic model with FH and SG factors λ_0, λ_1, λ_2[g]	CS-solution[e], $Q = n = A_{COSMO}/a_{eff}$ and σ-profiles $a_{eff} = 7.1\text{Å}^2$	C: Yes[f] R: No
COSMOSAC	FH-SG, $Q = zq/2$ $z = 10$, A_{ref}, V_{ref}[h]	CS-solution[e], $Q = n = A_{COSMO}/a_{eff}$ and σ-profiles $a_{eff} = 7.5\text{Å}^2$	C: Yes R: No
This paper	FH-SG $Q = \frac{(1-\phi/x)}{(1-\phi/\theta)}$ no parameters	revised UNIQUAC residual: $Q = (1 - \phi/x)/(1 - \phi/\theta)$ revised CS residual $n = (1 - \phi/x)/(1 - \phi/\theta)$	C: No R: No

[a] C = Combinatorial term, R = Residual term.
[b] FH-SG = Flory-Huggins with Staverman-Guggenheim correction.
[c] QCT = Guggenheim's Quasi-Chemical Theory.
[d] Homogeneous and non-homogeneous double binary solution.
[e] General CS-solution.
[f] Contains FH-SG terms, therefore debatable.
[g] Parameters for the combinatorial term in COSMOTherm C1.2.
[h] Reference area as given in Table 4.1, $V_{ref} = 25.17\text{Å}$.

Table 4.7: Comparison of the various models.

has nearly equal length as the compound of interest. Although this deviation is not significant for the combinatorial term, because the SG-term suppresses this by the ϕ_k/θ_k term, it is not acceptable for the residual term. For the time being the proposed procedure for defining the number of nearest neighbors works well as shown by the results in this work. It is hypothesized that the strong deviation is a result of Guggenheim's entropy model, where a single lattice coordination number was used. It seems that the Guggenheim model can be improved by introducing a molecule dependent lattice number, which is invariant for concentration, and to define the area fraction, θ_k, by the product $z_k q_k$ instead of q_k. This has been proven later by us in [129] and discussed in 3.

Chapter 5

Dispersion ACMs.
Part I: Cubic Equations of State[1]

5.1 Introduction

Activity coefficient and equation of state models enable us to describe or predict phase equilibria. Both approaches can be used to give an accurate phase equilibrium description by adjusting the pure component and binary interaction parameters. In the prediction of phase equilibria activity coefficient models are often preferred over equation of state methods, when the mixture contains no supercritical compounds and the liquid density is far from the critical density. One of the reasons for this is that the reference state of an activity model is the pure liquid state of the compounds, while for an equation of state this is the ideal gas. One could say that activity models interpolate between the pure liquid states of compounds, while equations of state extrapolate from the ideal gas to the condensed state using the critical point as a reference. This implies that in the case of predicting of phase equilibria, the predictive activity models are often more accurate than equation of state models, where the interaction parameters of the latter are set to zero and not predicted by an activity model, such as is done in the PSRK model [92].

Activity models such as UNIFAC [59], COSMO-RS [116] and COSMOSAC [145], often reasonably predict solid-liquid, liquid-liquid and vapor-liquid phase equilibria, especially for molecules with polar and associating groups. A weak point of these predictive models is that they do not contain a contribution that accounts specifically for the dispersive interaction between molecules, while in the family of

---

[1]Based on the publication: Dispersion activity coefficient models. Part 1: Cubic equations of state. Fluid Phase Equil. (2019), by Gerard J.P. Krooshof, Remco Tuinier, and Gijsbertus de With.

SAFT equations of state [29, 28, 78] this contribution is actually one of the corner stones. To repair the absence of an explicit contribution for the dispersive interaction, predictive activity models frequently use a modification of the Flory-Huggins (FH) combinatorial term [52], which goes back to the work of Donahue and Prausnitz [43]. The modification is done by fitting alkane mixtures. In other words the dispersion contribution eminent in alkane mixtures is lumped into the combinatorial part. However, this approach represents an ad-hoc modification that lacks a physical basis and introduces a systematic error in the interaction between molecules with a small alkyl fraction, such as water, methanol, acrylonitrile, which do not require a modified Flory-Huggins term. MOSCED [214] is one of the models that does explicitly take this dispersive contribution into account. However, MOSCED is developed for binary systems and yields only the activity coefficient at infinite dilution.

A way to obtain an activity coefficient expression for dispersion is to derive it from an equation of state. Van Laar [223] was the first who derived such an expression from the van der Waals (vdW) cubic equation of state [221]. His equation was criticized, because binary mixtures of compounds which have nearly the same critical pressure would yield ideal behavior, while a binary mixture of methanol ($P_\mathrm{c} = 79$ bar) and carbon disulfide ($P_\mathrm{c} = 81$ bar) was known to demix at 36 °C. However, this reasoning is not justified, because the critical pressures of these compounds have a different cause. Methanol interacts mainly by hydrogen bonds and only slightly by dispersion, while carbon disulfide interacts mainly via a strong permanent quadrupole-quadrupole interaction. Both the hydrogen bond and the quadrupole interaction are orientation-dependent. These are specific interactions which are not included as separate contributions in the van der Waals cEoS and its derivative, the van Laar activity coefficient equation. Besides the omission of specific interaction in the van der Waals cEoS, there are also two other important points to mention. First, the van de Waals equation is based on the assumption that all molecules are spherical [219], and second, that each molecule interacts with the same extent with its surroundings [220]. Therefore, the van der Waals and van Laar models would perform well as a predictive model for compound mixtures, in which the interactions are weak and without a preferred orientation, such as noble gases, and small hydrocarbons. However, the van Laar model is seldom applicable for these compounds, because most mixtures contain large molecules and the temperature is often above the critical temperature of these compounds. As a result, the van Laar model and derivatives of it use adjusted parameters; i.e. liquid volume instead of co-volume, solubility parameter instead of critical parameters. The work of Kontogeorgis and Folas [124] shows that the Peng-Robinson cubic EoS with classical mixing rules yields limiting activity coefficient values above unity, while an a/b mixing rule gives values below unity, which agree with experimental observation. However, such an approach is not

easy applicable in combination with the existing UNIFAC and COSMO-RS models, and has not been tested on branched alkanes, as we will do in this work.

Kontogeorgis and Coutsikos [127] derived expressions for the activity coefficient at infinite dilution of a solute in binary mixture from the Soave-Redlich-Kwong (SRK) [206, 184] and the Peng-Robinson (PR) [177] cEoS. This was done in a study on g^E-EoS models, which is a hybrid form of a cEoS, where an activity model is applied to adjust the mixing rule. A critical issue is the definition of a consistent reference state, for which the activity coefficient model and the equation of state should yield the same excess Gibbs (or Helmholtz) energy. Basically, there are two options. The Vidal and Huron method [225, 97] takes as reference state the mixture at infinite pressure. The Mollerup and Michelsen method [168, 166] uses the liquid at zero pressure as reference state. Other methods interpolate between these two approaches. For example, Tassios *et al.*, who introduced a linear combination of the Vidal and Michelsen (LCVM) model [19], or Ahlers and Gmehling [3], who introduced a volume translated PR cEoS with UNIFAC mixing rules at 1 atm. However, the developments of cEoS where the reference state is set at intermediate or zero pressure, should be seen in the context of trying to improve the description of phase equilibria near and above the critical point, where the system has a packing fraction of about 0.2. Since we are developing an activity coefficient model for the condensed state, or in general a system where the packing fraction is above 0.45, the Huron and Vidal method is the best option of the aforementioned methods.

In this work we start with the derivation of dispersion activity coefficient models using existing cEoS models. We compare two different size and energy combining rules with the classical combining rule for the binary energy and size parameters of the cEoS. We introduce a group contribution method to change the activity model from a descriptive to a predictive model. The models for the dispersive activity coefficient are combined with an off-lattice combinatorial contribution [129] to obtain the total activity coefficient. Results derived with the total activity models are compared to experimental data and discussed. We emphasize that our objective is not to define a model for hydrocarbons, but to show that the modified Flory-Huggins combinatorial term, can be replaced by a generalized Guggenheim combinatorial activity coefficient [129] in combination with an explicit dispersion activity coefficient, in order the arrive at a more robust basis for the predictive activity coefficient models like UNIFAC and COSMO-RS. In a follow-up paper we will develop a more refined dispersive activity coefficient model derived from the perturbed chain equation of state.

5.2 Theory

5.2.1 Cubic equations of state

There are two good reasons to look at cubic equations of state first, before we discuss the perturbed chain equations of state in a second paper. First, they are popular in flow-sheeting programs, because the cubic form is easy to solve. It makes flow-sheets, where the number of flash calculations can be huge, fast and robust. Second, it shows more clearly the key factors in activity models, as we will come across in this work. We start with a general form for the cubic equation of state (cEoS):

$$P = \frac{RT}{V - b} - \frac{a(T)}{(V + \lambda_1 b)(V + \lambda_2 b)}, \tag{5.1}$$

where P is the pressure, V the volume, b the excluded covolume of the molecules, T the absolute temperature, and $a(T)$ an energy term, which is temperature dependent in order to describe accurately the saturation pressure of pure liquids from their melting point up to their critical point. The parameters λ_1 and λ_2 are empirical constants, which have been introduced mainly to improve the description of the liquid density. Table 5.1 gives typical values for the parameters λ_1, λ_2 for the most common cubic equations of state. Cismondi and Mollerup (CM) [32] introduced

cEoS Name	Reference	λ_1	λ_2
van der Waals (vdW)	[221]	0	0
Soave-Redlich-Kwong (SRK)	[206, 184]	0	1
Peng-Robinson (PR)	[177]	$1 + \sqrt{2}$	$1 - \sqrt{2}$
Cismondi-Mollerup (CM)	[32]	$< 0, 1 + \sqrt{2} >$	Eq. 5.2

Table 5.1: Cubic equation of state parameters.

a three-parameter cEoS and showed that it can be made equally accurate as the PC-SAFT model when λ_2 is linked to λ_1 by:

$$\lambda_2 = \frac{1 - \lambda_1}{1 + \lambda_1}. \tag{5.2}$$

Basically, the CM model interpolates for each compound and mixture between the SRK and the PR cEoS. The λ_1 parameter is found by fitting the cEoS to pure component density and saturated vapor pressure data. In the computation of properties of mixtures, λ_1 is the molar average value of the pure liquids. The λ_2 parameter is still calculated by Eq. 5.2, but now using the average value of λ_1. In this paper we use the CM notation, Eq. 5.2, to define a class of two-parameter cubic equations of state.

Mixing and combining rules

For multicomponent mixtures Eq. 5.1 is often used in combination with mixing rules that quantify the parameters a and b for each fluid phase. The classical mixing rules are given by:

$$a = \sum_{j,k} x_j x_k a_{jk} \quad \text{and} \quad b = \sum_{j,k} x_j x_k b_{jk}, \tag{5.3}$$

where x_j is the mole fraction of compound j, and a_{jk} and b_{jk} are binary mixture parameters, which are often defined by the generalized Berthelot [17] and Lorentz [153] combining rules:

$$a_{jk} = \sqrt{a_j a_k}(1 - k_{jk}) \quad \text{and} \quad b_{jk} = \frac{(b_j + b_k)}{2}(1 - l_{jk}). \tag{5.4}$$

The interaction parameters k_{jk} and l_{jk} are used to adjust the energy and excluded volume mixing rule, respectively, in order to bring the results of the cEoS model in agreement with experimental phase equilibria data. In this work we use the term ideal classic (IC) to denote mixtures where these parameters are zero; i.e. Eq. 5.4 with $k_{jk} = l_{jk} = 0$.

In the molecular dynamics field molecular interactions are frequently described using Lennard-Jones (LJ) potentials. Then other combining rules are applied and here we investigate two of them. Kong [123] defined three energy criteria for the repulsive and attractive part of the LJ potential and obtained mixing rules, which we have rearranged to:

$$\sigma_{ij}^6 = \frac{1}{\sqrt{\epsilon_{ii}\epsilon_{jj}}\sigma_{ii}^3\sigma_{jj}^3} \left[\frac{(\epsilon_{ii})^{\frac{1}{13}} (\sigma_{ii})^{\frac{12}{13}} + (\epsilon_{jj})^{\frac{1}{13}} (\sigma_{jj})^{\frac{12}{13}}}{2} \right]^{13} \quad \text{and}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}\frac{\sigma_{ii}^3\sigma_{jj}^3}{\sigma_{ij}^6}, \tag{5.5}$$

where ϵ_{jj} and σ_{jj} are the Lennard-Jones energy and size parameters between two molecular segments of the same kind j, while subscript ij refers to the unlike pairs. Waldman and Hagler (WH) [228] applied mathematical scaling and symmetry rules, leading to:

$$\sigma_{ij}^6 = \left(\frac{\sigma_{ii}^6 + \sigma_{jj}^6}{2} \right) \quad \text{and} \quad \epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}\frac{\sigma_{ii}^3\sigma_{jj}^3}{\sigma_{ij}^6}. \tag{5.6}$$

We mention that the WH size combining rule yields always larger values than the IC size combining rule, when we replace the b_j by σ_{jj}^3 and b_{ij} by σ_{ij}^3.

The IC, Kong and WH combining rules can be evaluated by comparing the calculated results of the unlike size and energy parameters of the noble gases ranging from He to Xe with the accurate parameter values of Kestin *et al.* [111]. The WH combining rules give the highest accuracy in calculating the unlike parameters of noble gases (See appendix F). Kong is second best. In our work we will further apply only the classical and the WH combining rules for the cEoS. In order to do so, we have to convert the WH rules into terms of the cEoS parameters a and b. For monoatomic spherical molecules this results in:

$$b_{ij} = \sqrt{\frac{b_i^2 + b_j^2}{2}} \quad \text{and} \quad a_{ij} = \sqrt{a_i a_j} \frac{2 b_i b_j}{b_i^2 + b_j^2}. \tag{5.7}$$

For polyatomic molecules Eq. 5.4 and 5.7 are not applicable, because the molecular shape is no longer spherical as is schematically shown in Fig. 5.1. The excluded covolume of a polyatomic molecule and a monoatomic molecule is defined by half the volume enclosed by a surface, which is defined by the center of the monoatomic molecule that is rolled around the polyatomic molecule. Here, the factor half is the correction that eliminates double count in the summation of the enclosed volumes of all the molecules in the mixture. Fig. 5.1 implies that the excluded covolume depends on the size of the monoatomic molecule that touches the polyatomic molecule and the size mixing rule. We denote the excluded covolume by the type of the rolling sphere; e.g. b_H is the excluded covolume of a molecule calculated by rolling a hydrogen atom around it. The core or the van der Waals volume of a molecule is readily calculated by the method described in the appendix of Bondi's book [18]. For this calculation one needs the van der Waals radii of the atoms and the bond lengths between the atoms. Pauling [176] tabulated lengths of a large set of bonds, but we will restrict ourselves to carbon-carbon and hydrogen-carbon bonds, which have an average length of L_{CC} = 1.54 Å and L_{CH} =1.10 Å, respectively. The van der Waals radii of the carbon and hydrogen atoms are R_C = 1.70 Å and R_H =1.20 Å, respectively. The volume defined by touching an atom of a polyatomic molecule around another polyatomic molecule, is calculated in the same way, but now the radii are no longer the atomic radii, but the sum of the radius of the atom of the core molecule and that of the touching atom. In case the two atoms are alike, we have for two hydrogen atoms R_{HH} = 2.40 Å, and for two carbon atoms R_{CC} = 3.40 Å. In case the touching atom is different from the atom of the core molecule, i.e. the unlike situation, the sum of the radii is defined by a size mixing rule. For the IC and WH mixing rules this yield in case of alkanes R_{CH} = 2.90 and 2.94 Å, respectively. In Fig. 5.1 the gap between the carbon and the hydrogen atom illustrates the effect of the mixing rule, which makes the center to center distance larger than the sum of the two atom radii. The left and right panels of Fig. 5.1 show that the carbon atom, as part of a molecule, gives

a larger covolume to the polyatomic molecule than the hydrogen atom does. The
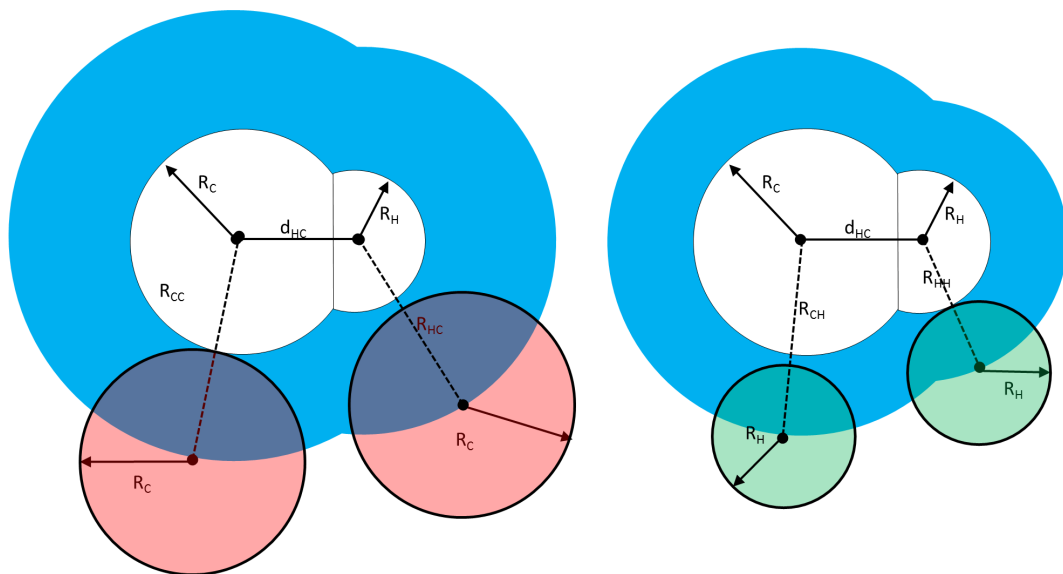
Figure 5.1: Schematic picture that illustrates the calculation of the excluded covolumes. A hypothetical molecule, formed by a carbon and a hydrogen atom and having bond length d_{HC}, is scanned by a touching carbon (left) and a hydrogen (right) of another polyatomic molecule. The center to center distance of an unlike pair, here $R_{CH} = R_{HC}$ is computed by the size mixing rule and is equal or larger that the sum of the two radii. The center to center distance of an alike contact is simply the sum: $R_{HH} = 2R_H$ and $R_{CC} = 2R_C$. The blue area depicts the excluded volume added to the core volume, where the center of the red or green sphere defines the outer surface.

obtained volumes are divided by two to get the excluded covolumes for a touching carbon and hydrogen atom, b_C and b_H, respectively. These are listed in Table 5.2 in the form of linear relations in terms of the carbon number N_C, which were obtained by a linear regression of the obtained results.

Method	b_C	b_H	b_{jj} (pure)
$4 \times v_{vdW}$	-	-	$67.8\ N_C + 45.9$
IC	$37.5\ N_C + 64.8$	$26.9\ N_C + 38.2$	$30.5\ N_C + 43.5$
WH	$39.1\ N_C + 66.5$	$26.6\ N_C + 39.1$	$30.9\ N_C + 44.2$

Table 5.2: Excluded covolume between an atom (Carbon and Hydrogen) of n-alkanes (Å^3) as a linear function of the carbon number N_C (Eq. 5.8).

The covolume of a central molecule, as pure liquid b_{jj} or as solute in a mixture, b_j, can be defined by the external contact fraction of the atoms in the surrounding

liquid. As a first order approximation the carbon and hydrogen contact probability, $\mathcal{P}_C$ and $\mathcal{P}_H$, can be set equal to the mole fraction of the atoms in the surrounding. This yields the following equation for the covolume of a molecule b_j:

$$b_j = \mathcal{P}_C b_C\left(N_C\right) + \mathcal{P}_H b_H\left(N_C\right) \approx x_C b_C\left(N_C\right) + x_H b_H\left(N_C\right), \tag{5.8}$$

where x_C and x_H are the mole fractions of the carbon and hydrogen atoms in the surrounding medium, respectively. The functions $b_C\left(N_C\right)$ and $b_H\left(N_C\right)$ are linear relations of N_C, which were obtained by a least-squared regression of the results obtained by contacting a carbon and hydrogen atom around an n-alkane, respectively.

The covolume of the pure compound is calculated by setting the carbon and hydrogen fractions equal to the composition of the central molecule. The obtained results for the covolumes of the n-alkanes in the pure state are regressed by a linear fit. This leads to the b_{jj} parameters given in the right column of Table 5.2. When we compare these values, we see that the IC and WH combining rules give nearly the same result (Fig. 5.2a). However, these linear relations have a lower slope than those coming from the simple calculation of 4 times the van der Waals volume. In fact, the ratio of covolume b and van der Waals volume v_{vdW} drops with increasing chain length to an asymptotic level, just below the value known for infinite cylinders, where $b/v_{vdW} = 2$ (Fig. 5.2b).
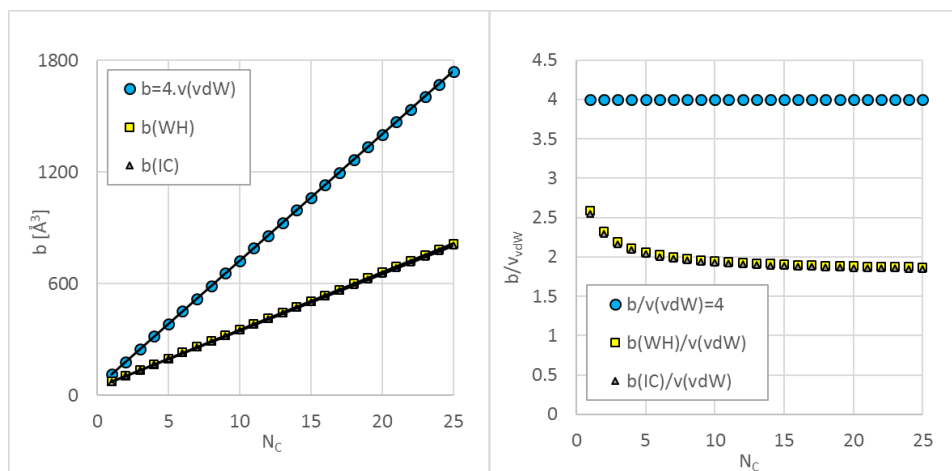
Figure 5.2: Left panel: Covolume of pure n-alkane as function of the alkane number, N_C, calculated by the classical hard sphere result $b = 4v_{vdW}$ (circles), and the rolling ball method, where the unlike size parameter, R_{CH}, is calculated with the WH (squares) and IC (triangles) combining rules. Lines are linear regressions with parameters given in table 5.2. Right panel: Ratio of the covolume and the van der Waals volume as function of N_C.

Because the relations of Table 5.2 are linear, the average covolume of a mixture

can be written as:

$$\bar{b} = x_C b_C \left(\overline{N}_C \right) + x_H b_H \left(\overline{N}_C \right), \tag{5.9}$$

where $b_C \left(\overline{N}_C \right)$ and $b_H \left(\overline{N}_C \right)$ are the carbon- and hydrogen-based covolume, and the system average alkane number $\overline{N}_C$ is given by:

$$\overline{N}_C = \sum_j x_j N_{C,j}. \tag{5.10}$$

Eq. 5.9 can also be written as:

$$\bar{b} = \beta_0 + \beta_1 \sum_j x_j N_{C,j}, \tag{5.11}$$

where β_0 and β_1 are the intercept and the slope of the linear relations given in the right column of Table 5.2. This implies:

$$\bar{b} = \beta_0 + \beta_1 \sum_j x_j N_{C,j} = \sum_j x_j \left(\beta_0 + \beta_1 N_{C,j} \right) = \sum_j x_j b_{jj}. \tag{5.12}$$

In other words, in hydrocarbon mixtures the (atomic) size combining rule is not important for the cubic equation of state, because the high number of atoms average its value to the result of the classical combining rule. We observe that the covolume b of a long linear molecule is no longer $4v_{vdW}$, but is reduced to about $2v_{vdW}$ when we apply for the contact probability the mole fraction of the atoms in the mixture. This behavior changes a little, when the external contact probability is defined by surface fractions instead of mole (See supporting material: Excel workbook). Also in this case the covolume will decrease with increasing chain length. This is an important finding, which helps us to understand the systematic error of activity coefficient equations, which are derived from the cubic equation of state.

Gibbs energy of mixing at infinite pressure

The molar Gibbs energy of mixing, g^{mix}, for the two-parameter cEoS is given by the equation:

$$\frac{g^{\text{mix}}}{RT} = Z - \sum_j x_j Z_j + \left[\sum_j x_j \ln \left(\frac{V_j - b_j}{V - b} \right) \right]$$

$$+ \frac{1}{RT(\lambda_1 - \lambda_2)} \left[\sum_j x_j \frac{a_j}{b_j} \ln \left(\frac{V_j + \lambda_1 b_j}{V_j + \lambda_2 b_j} \right) - \frac{a}{b} \ln \left(\frac{V + \lambda_1 b}{V + \lambda_2 b} \right) \right]. \tag{5.13}$$

Here Z and Z_j are the compression factors of, respectively, the mixture and that of component j. The second line of the equation is the dispersive energy term, while the first part describes the Flory-Huggins contribution. In the limit of infinite pressure the volume shrinks to the covolume, and all atoms make maximum contact with the surrounding atoms. We mention that this expression does not hold for the aforementioned three-parameter model of Cismondi and Mollerup (See appendix H). Here we use their notation to denote a two-parameter cEoS, which lies between the SRK and the PR cEoS. In the high pressure limit, introduced by Vidal [225], the dispersive part of Eq. 5.13 becomes:

$$\frac{g^{\text{mix},\infty}}{RT} = \frac{\kappa}{RT}\left[\sum_j x_j \frac{a_j}{b_j} - \frac{a}{b}\right], \tag{5.14}$$

where the parameter κ is a cEoS constant defined by the parameters λ_1 and λ_2 according to:

$$\kappa \equiv \frac{\ln\left(1+\lambda_1\right) - \ln\left(1+\lambda_2\right)}{\lambda_1 - \lambda_2} \tag{5.15}$$

Values of the κ parameter are given in Table 5.3. The characteristic pure compound parameters a and b of the cEoS are defined by the critical volume V_{c}, the critical temperature T_{c}, and the critical pressure P_{c}. We mention that the repulsive part goes to infinity, due to the infinite pressure. It places the molecules into close contact and therefore this part of the Gibbs energy of mixing can be expressed by the generalized Guggenheim combinatorial model [129].

Dispersive activity coefficient and enthalpy of mixing from cEoS

The activity coefficient equation can be derived from the excess Gibbs energy g^{E}, by applying the thermodynamic relation:

$$\ln\gamma_k = \frac{1}{RT}\left(\frac{\partial n_t g^{\text{E}}}{\partial n_k}\right)_{T,P,n_j \neq k} \approx \frac{1}{RT}\left(\frac{\partial n_t a^{\text{E}}}{\partial n_k}\right)_{T,V,n_j \neq k}, \tag{5.16}$$

where n_t is the total number of moles, n_k the number of moles of molecule k in the mixture, and a^{E} is the excess Helmholtz energy. The approximation made in the last part of Eq. 6.15 holds for mixtures where the excess volume is negligible. In order to arrive at an expression for the activity coefficient, we write Eq. 5.14 into the form:

$$\frac{n_t g^{\text{mix}}}{RT} = \frac{\kappa}{RT}\left[\sum_j n_j \frac{a_j}{b_j} - n_t \frac{\sum_{j,k} n_j n_k a_{jk}}{\sum_{j,k} n_j n_k b_{jk}}\right]. \tag{5.17}$$

Substitution of Eq. 5.17 into Eq. 6.15 yields the following general expression for the activity coefficient of a two-parameter cEoS:

$$
\ln \gamma_j^{\text{disp, cEoS}} = \frac{\kappa}{RT} \left[\frac{a_j}{b_j} - \frac{2 \sum_k x_k a_{jk}}{\sum_{k,m} x_k x_m b_{km}} \right]
$$
$$
+ \frac{\kappa}{RT} \left[\frac{\left(2 \sum_k x_k b_{jk} - \sum_{k,m} x_k x_m b_{km} \right) \sum_{k,m} x_k x_m a_{km}}{\left(\sum_{k,m} x_k x_m b_{km} \right)^2} \right], \quad (5.18)
$$

where a_{jk} and b_{jk} are defined by one of the aforementioned combining rules. We mention that a similar relation is obtained for the three-parameter cEoS of Cismondi and Mollerup (See appendix H). The excess enthalpy of component j is given by the relation:

$$
h_j^{\text{E}} = -RT^2 \frac{\partial \ln \gamma_j}{\partial T} \quad (5.19)
$$

With the general expression for the activity coefficient of a cEoS, and assuming that the cEoS parameters a and b are temperature- independent, we obtain for the enthalpy change on mixing:

$$
H_{\text{mix}} = \sum_j x_j h_j^{\text{E}} = \kappa \left[\sum_j x_j \frac{a_j}{b_j} - \frac{\sum_{j,k} x_j x_k a_{jk}}{\sum_{k,m} x_k x_m b_{km}} \right] \quad (5.20)
$$

It shows that, under the assumption that a and b are temperature-independent, the enthalpy of mixing of a cEoS is also temperature-independent. Next, we discuss the effect of the difference types of combining rules for the activity coefficient equation derived from the two-parameter cEoS.

A. Ideal classical combining rules

The classical combining rules (see Eq. 5.4) with the approximation $l_{jk} = k_{jk} = 0$, yields the following expression for the activity coefficient of component j:

$$
\ln \gamma_j^{\text{cEoS-IC}} = \frac{\kappa}{RT} \left[\frac{a_j}{b_j} - 2 \frac{\sum_k x_k \sqrt{a_j a_k}}{\sum_{j,k} x_j b_j} + \frac{b_i \sum_{j,k} x_j x_k \sqrt{a_j a_k}}{\left(\sum_{j,k} x_j b_j \right)^2} \right]. \quad (5.21)
$$

By defining the volume fraction as:

$$
\phi_j = \frac{x_j b_j}{\sum_k x_k b_k}, \quad (5.22)
$$

eq. 5.21 can be written as:

$$\ln \gamma_j^{\text{IC}} = \kappa \frac{b_j}{RT} \left[\frac{a_j}{b_j^2} - 2 \frac{\sqrt{a_j}}{b_j} \sum_k \phi_k \frac{\sqrt{a_k}}{b_k} + \sum_{j,k} \phi_j \phi_k \frac{\sqrt{a_j a_k}}{b_j b_k} \right]$$

$$= \kappa \frac{b_j}{RT} \left[\frac{\sqrt{a_j}}{b_j} - \sum_k \phi_k \frac{\sqrt{a_k}}{b_k} \right]^2 , \qquad (5.23)$$

which, we recognize that the equation attains the van Laar equation for $\kappa = 1$. We also mention that Hildebrand's solution theory result is obtained [88], when the van der Waals covolume is replaced by the volume of the pure liquid at 25°C and the term $\sqrt{a_j}/b_j$, usually denoted as the solubility parameter, is optimized to experimental data. This approach has also been applied in the Hansen solubility model [84, 10], where besides dispersion, also solubility parameters for polar and hydrogen bonding are included.

In PC-SAFT the size of molecules are defined by a number of spheres, where methane has a normalized size of $m_1 = 1$. In a similar way we can take the covolume of the two-parameter cEoS as a measure to define the relative length of an alkane. As is explained in the work of Kontogeorgis and Folas [124], the covolume can be obtained from the van der Waals volume and the ratio of the critical temperature and pressure, $b \propto RT_c/P_c$. We opt for the relation $b = y.V_c$, where y is a cEoS specific constant (Table 5.3), to reduce the deviations arising from the difference between experimental and mean-field theory Z_c values, which can be explained by crossover theory [226, 238] (see appendix I). By doing this, we can define the number of units in an alkane by the ratio in covolumes. Therefore, with the critical volume of methane, $v_{c,1}$, we can define the relative size of an alkane by:

$$m_j = \frac{b_j}{b_1} = \frac{V_{c,j}}{V_{c,1}}. \qquad (5.24)$$

The a and b parameters in Eq. 5.23 can be replaced by the critical pressure $P_{c,j}$, the critical temperature $T_{c,j}$, and the critical volume $V_{c,j}$ of compound j (see appendix D) and we obtain the following expression for the dispersive activity coefficient equation derived from a cEoS.

$$\ln \gamma_j^{\text{cEoS-IC}} = K \, m_j \left(\frac{T_0}{T} \right) \left[\Delta_j - \sum_k \phi_k \Delta_k \right]^2 , \qquad (5.25)$$

where Δ_j is the dimensionless Hildebrand parameter of compound j, defined as:

$$\Delta_j = \sqrt{\frac{P_{c,j} \, V_{c,1}}{RT_0}}, \qquad (5.26)$$

with the reference temperature T_0 set at 298.15 K. The parameter K is a cEoS specific component independent constant and follows from the cEoS parameters λ_1 and λ_2. In appendix E we show how the value of K can be calculated and Table 5.3 shows typical values for each cEoS. The model of Cismondi and Mollerup yields for

EoS	Ω_a	Ω_b	y [a]	Z_c	$\kappa(Eq.5.15)$	K [b]	K/K_{vdW}
vdW	0.4219	0.1250	0.333	0.375	1	9	1
SRK	0.4275	0.0866	0.259	0.333	0.6931	10.3	1.14
PR	0.4527	0.0778	0.253	0.307	0.6232	11.9	1.32

[a] Note: y see appendix E.
[b] Note: $K = \kappa y \Omega_a / \Omega_b^2 = \kappa \Omega_a / (\Omega_b Z_c)$.

Table 5.3: Cubic equation of state constants.

the pure components values for K, which run along the curve as depicted in Fig. 5.3, starting from $K = 9.97$, when $\lambda_1 = \sqrt{2} - 1$. Values for the size parameter m_j and the dimensionless solubility parameter Δ_j can be calculated directly from the experimental critical pressure and volume. Experimental values of the critical parameters are for most compounds available in the DIPPR database [23]. An accurate series of critical volume measurements of n-alkanes between pentane and octadecane has been measured by Ansalme *et al.* [4]. Values of m_j and Δ_j for some alkanes are given in Table 5.4.

In case no experimental critical parameters are available, an appropriate group contribution (GC) method, such as the method of Marrero and Gani [161], can be applied to estimate the critical parameters P_c and v_c to calculate Δ and m. Instead

Alkane	Source	$m_j = V_{c,j}/V_{c,1}$	Δ_j
Methane	exp.	1	0.428
Ethane	exp.	1.476	0.440
n-Pentane	exp.	3.174	0.366
n-Hexane	exp.	3.763	0.346
n-Heptane	exp.	4.341	0.330
n-Octane	exp.	4.929	0.315
n-Hexadecane	pred.	9.574	0.235
Squalane	pred.	16.63	0.186

Table 5.4: Alkane activity model parameters at $T_0 = 298.15K$. Experimental (exp.) and predicted (pred.) values

of this method, we found it more convenient to create a predictive method for the size parameter m and the parameter Δ, since they show a clear relationship with
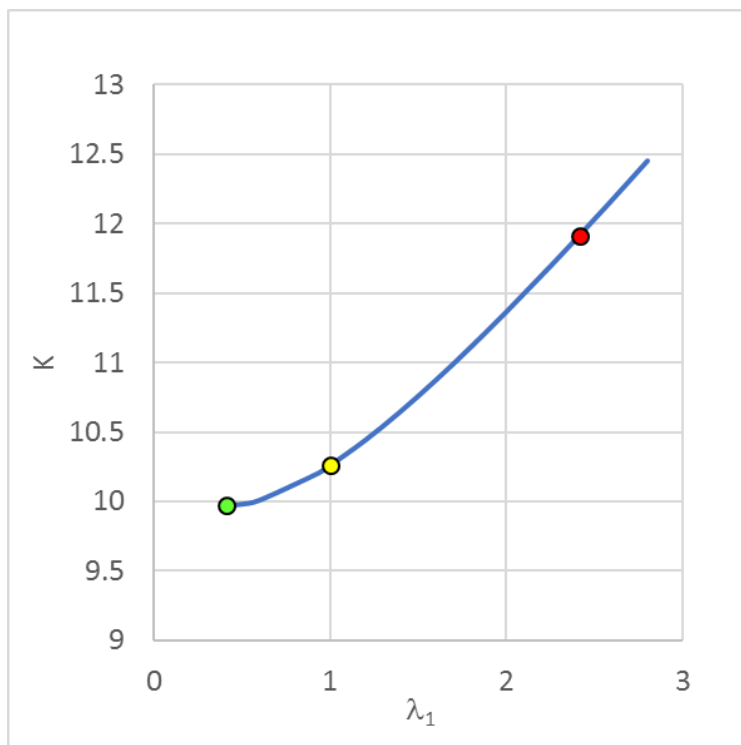
Figure 5.3: Value of parameter K using the CM model parameter λ_1 (Eq. 5.1). The green circle is the minimum value, where $\lambda_1 = \sqrt{2} - 1$, the yellow and red circles denote the K-value for the SRK and PR cEoS, respectively.

the carbon number and type of atom groups. Fig. 5.4a shows that size parameter m increases linearly with carbon number N_C. Because of this, we can use the same
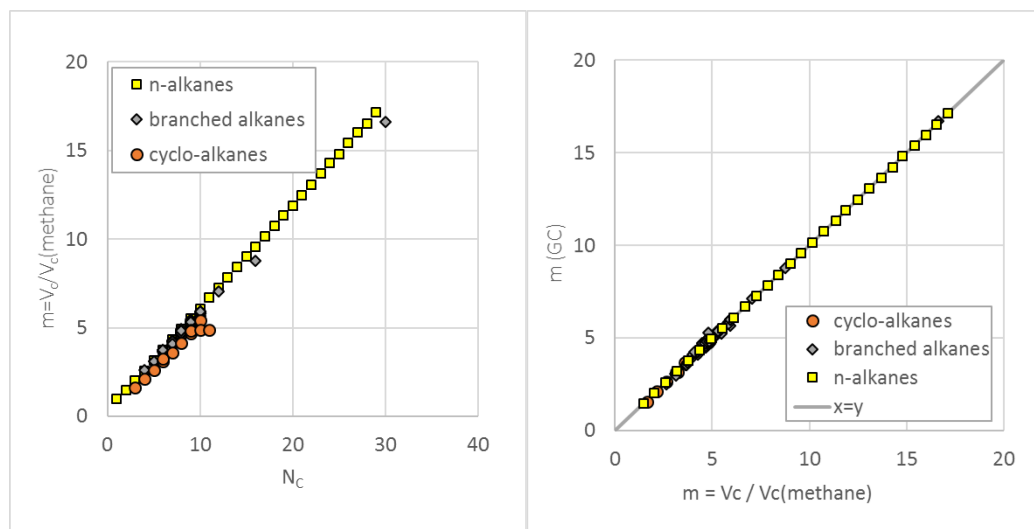
Figure 5.4: Size parameter of n-alkanes, cyclo-alkanes and branched alkanes as function of carbon number (left panel). Parity plot of m calculated by ratio in critical volume and group contribution method after optimization of primary and secondary increments (right panel).

methodology as was applied by Marrero and Gani, and use:

$$m = \sum_j p_j N_j + \sum_k s_k N_k, \tag{5.27}$$

where p_j is the increment of a primary group j (i.e. an atom group), s_k the increment of a secondary group k (i.e. a proximity correction), and N_j and N_k are the number of atom groups or proximity corrections in the molecule, respectively. Values for the primary and secondary increments were obtained by a least-squared regression of the deviation in m and Δ data, which were calculated from the critical pressure and volume data of alkanes retrieved from the DIPPR database [23] (See appendix E). The increments for the group contribution (GC) formula are given in Table 5.5.

Fig. 5.5 shows that the values of the parameter Δ decrease with increasing relative size m. Eventually Δ vanishes for $m \to \infty$ due to the fact that an infinitely large molecule has zero critical pressure. For this parameter we opted for a power-law description of the following form:

$$\Delta = \left[\frac{t_1}{1 + \sum_j p_j N_j + \sum_k s_k N_k} \right]^{t_2}, \tag{5.28}$$

Group	Type	Order [a]	m	Δ
CH$_3$-	linear	p	0.728	0.268
-CH$_2$-	linear	p	0.580	0.206
>CH-	linear	p	0.318	0.083
>C<	linear	p	0.153	-0.004
-CH$_2$-	cyclic	p	0.526	0.134
>CH-	cyclic	p	0.340	0.096
-C(CH$_3$)$_2$-	linear	s	-0.095	-0.065

[a] p: primary group, s: secondary group.

Table 5.5: GC increments for m and Δ.

where the parameters t_1 and t_2 are independent of the molecular structure. The equation is based on the GC model for the critical pressure as proposed by Marrero and Gani [161], since the dimensionless solubility parameter, Δ, is in fact the critical pressure times a constant $V_c(\text{methane})/(R.T_0)$. The parameters of Eq. 5.28 are obtained by a least-squared fit of the experimental based Δ values of alkanes. We obtained $t_1 = 0.392$ and $t_2 = 0.594$. The group contribution increments for the dimensionless solubility parameter are given in Table 5.5.
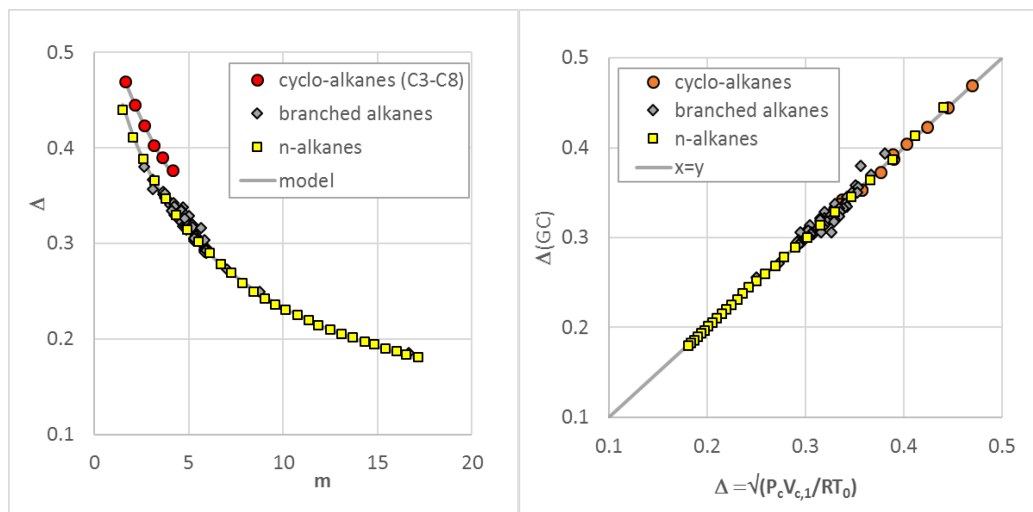
Figure 5.5: Left panel: Δ of n-alkanes, cyclo-alkanes and branched alkanes as function of size parameter m. Right panel: Parity plot of Δ calculated by the ratio in critical volume and the group contribution method after optimization of primary and secondary increments.

The activity coefficient equation of a binary mixture now follows from Eq. 5.25 as:

$$\ln \gamma_j^{\mathrm{IC}} = K \, m_j \, \left(\frac{T_0}{T} \right) [\Delta_j - \Delta_k]^2 \, (1 - \phi_j)^2 . \tag{5.29}$$

Subsequently, we can deduce the equation for the case where solute j is dissolved at infinite dilution in solvent k:

$$\ln \gamma_j^{\mathrm{IC},\infty} = K \, m_j \, \left(\frac{T_0}{T} \right) [\Delta_j - \Delta_k]^2 . \tag{5.30}$$

An important observation can be made from Eqs. 5.25, 5.29 and 5.30. The parameters m_j, Δ_j, and Δ_k are invariant, because they are determined by the critical parameters of the compound. As a result, K is the only parameter that could be used to downscale the value of the activity coefficients. But K is determined by the λ_1 (and λ_2) values for the equation of state, and therefore only has a limited range. The vdW cEoS has $K = 9$, while the SRK and the PR cEoS give a value of 10.26 and 11.92, respectively. Cismondi and Mollerup [32] and later Cismondi *et al.* [33] showed that a very good phase equilibria description is obtained when the λ_1 parameter is optimized for each pure component. The results of Cismondi *et al.* [33] show that the λ_1 parameter of the pure alkanes between methane and n-hexatriacontane ranges between 0.92 and 2.88, which sets the range of K for the two-parameter cEoS somewhere between 10.2 and 12.5.

The expression for the enthalpy of mixing of a binary mixture is derived by applying Eq. 5.20 to Eq. 5.29 yielding:

$$\begin{aligned} h_{\mathrm{mix}}^{\mathrm{IC}} &= \kappa \left[x_1 \frac{a_1}{b_1} + x_2 \frac{a_2}{b_2} - \frac{x_1 x_1 a_1 + 2 x_1 x_2 \sqrt{a_1 a_2} + x_2 x_2 a_2}{x_1 b_1 + x_2 b_2} \right] \\ &= \kappa \, \bar{b} \left[\frac{\sqrt{a_1}}{b_1} - \frac{\sqrt{a_2}}{b_2} \right]^2 \phi_1 \phi_2, \end{aligned} \tag{5.31}$$

where $\bar{b}$ is the average covolume of the binary mixture. The enthalpy of mixing can also be written in terms of the average chain length $\overline{m}$ and the dimensionless solubility parameter Δ_j:

$$h_{\mathrm{mix}}^{\mathrm{IC}} = K \, \overline{m} \, [\Delta_1 - \Delta_2]^2 \, R T_0 \, \phi_1 \phi_2. \tag{5.32}$$

We mention that this expression is temperature-independent and has a maximum at: $\phi_1 = x_2 = 1/(1 + \sqrt{m_2/m_1})$ or, equivalently, at $x_1 = \phi_2 = 1/(1 + \sqrt{m_1/m_2})$.

B. The Waldman and Hagler combining rules

The WH combining rule describes the unlike LJ-parameters of noble gases the best (See appendix F). Substitution of this rule into Eq. 5.18 gives for polyatomic molecules an expression that cannot be simplified into terms of $\sum_j n_j b_j$ due to the square root terms. To gain insight, we can elaborate on a binary mixture. The equation of a compound j in a binary mixture with compound k is:

$$
\ln \gamma_j^{\text{WH}} = K\ m_j \left(\frac{T_0}{T}\right) \Delta_j^2 \Bigg[1 - \frac{2x_j + x_k \frac{\Delta_k}{\Delta_j} \frac{4r^2}{1+r^2}}{x_j^2 + x_j x_k \sqrt{2 + 2r^2} + rx_k^2}
$$
$$
+ \frac{\left(2x_j + x_k^2 \sqrt{2 + 2r^2} - x_j^2 - rx_k^2\right)\left(x_j^2 + x_j x_k \frac{\Delta_k}{\Delta_j} \frac{4r^2}{1+r^2} + x_k^2 r^2 \left(\frac{\Delta_k}{\Delta_j}\right)^2\right)}{\left(x_j^2 + x_j x_k \sqrt{2 + 2r^2} + rx_k^2\right)^2} \Bigg],
$$

$$(5.33)$$

where r is the solvent-size ratio, $r = m_k/m_j$. At infinite dilution, where solute j is dissolved in solvent k, Eq. 5.33 reduces to:

$$
\ln \gamma_j^{\text{WH},\infty} = K\ m_j \left(\frac{T_0}{T}\right) \left[\Delta_j^2 - \frac{4r}{(1 + r^2)} \Delta_j \Delta_k + \left(\sqrt{2 + 2r^2} - r\right) \Delta_k^2 \right]. \quad (5.34)
$$

The second and third terms within the square brackets of Eq. 5.34 are different from Eq. 5.30, due to the functions with argument r. When the molecules are of equal size, i.e. $r = 1$, Eq. 5.34 reduces to Eq. 5.30. However, for $r >> 1$, Eq. 5.33 and 5.34 diverge exponentially. It demonstrates in a way that the use of this equation is limited to monoatomic molecules, such as the noble gases, and that for polyatomic molecules the classical combining rule is preferred.

Systematic deviation

Eq. 5.29 yields unrealistic high values for the activity coefficient, due to the high value for K. This can be repaired by using the approach of Hildebrand [87, 89], who proposed to replace the covolume b_j by the molar liquid volume $v_{\text{liq},j}$ of compound j and the ratio of $\sqrt{a}$ and b by the cohesion energy. Because $v_{\text{liq},j}$ about 1.9 v_{vdW}, while the covolume b_j is 4 v_{vdW} for spherical molecules, the logarithmic activity coefficient is by this replacement improved by a factor two The introduction of the cohesive energy, which reflects the effective intermolecular interaction, further reduces the deviation.

We can understand the systematic deviation also from a theoretical point of view. Fig. 5.2 shows that the ratio of covolume and van de Waals volume decreases with increasing chain length; the covolume changes from 4 v_{vdW} to about 1.9 v_{vdW}. This

implies that for long molecules it looks as if the K parameter should reduced by a factor two as well. The second correction comes from the assumption, that all molecules experience the same average concentration of the mixture. It neglects that the shape of a molecule affects the intermolecular interaction. In case of polyatomic molecules some atoms of the molecule are shielded or hindered by other atoms of the same molecule. This can also be interpreted by the fact that part of the intermolecular interaction is replaced by intramolecular interaction and covalent bonds, which do not contribute to the activity coefficient. This local field effect reduces the value of the parameter a. The overall effect is that the parameter K is more than two times smaller than the λ_1 and λ_2 or a and b parameters suggest. In the application of cEoS-based activity coefficient models, we are thus forced to optimize the K parameter, since the theoretical values are useless for real systems.

5.2.2 The total activity coefficient

In general, the total activity coefficient of molecules, which interact by dispersion forces only, is given by the sum of a combinatorial and a dispersive contribution.

$$\ln \gamma_k^{\text{tot}} = \ln \gamma_k^{\text{comb}} + \ln \gamma_k^{\text{disp}}. \tag{5.35}$$

For the combinatorial activity coefficient we take the generalized Guggenheim combinatorial activity coefficient of Krooshof *et al.* [129]:

$$\ln \gamma_k^{\text{comb}} = \ln \left(\frac{\phi_k}{x_k} \right) + \left(1 - \frac{\phi_k}{x_k} \right) \left[\frac{\ln \left(\frac{\phi_k}{\theta_k} \right)}{\frac{\phi_k}{\theta_k} - 1} \right], \tag{5.36}$$

where ϕ_k and θ_k are the volume and surface fraction of compound 1 in the mixture, respectively. This expression in combination with the residual part of the UNIQUAC model or COSMOSPACE model gives better results for binary equilibria of alkane-alkane and alkane alcohol systems (Chapter 4) than the combination with original Staverman-Guggenheim combinatorial activity coefficient model. Substitution of Eqs. 5.29 and 5.36 into Eq. 5.35 leads to the following expression for the cEoS-based total activity coefficient equation:

$$\ln \gamma_k^{\text{tot,cEoS}} = \ln \left(\frac{\phi_k}{x_k} \right) + \left(1 - \frac{\phi_k}{x_k} \right) \left[\frac{\ln \left(\frac{\phi_k}{\theta_k} \right)}{\frac{\phi_k}{\theta_1} - 1} \right]$$
$$+ K m_k \left[\Delta_k - \Delta_j \right]^2 \left(1 - \phi_k \right)^2 \tag{5.37}$$

5.3 Results

To optimize the K parameter of Eq. 5.37, we need activity coefficient data, which are consistent according to the rule of Gibbs-Duhem, and which are sufficiently accurate to reveal deviations between different activity coefficient models. Therefore, vapor-liquid equilibria (VLE) data of alkane mixtures with a difference in carbon number below three, which yield activity coefficients close to unity, are not useful to observe differences between the models. On the other hand, isothermal or isobaric VLE data of asymmetric systems are in most cases incomplete in vapor composition or are unreliable with respect to the largest compound, because the low partial pressure of this compound is difficult to quantify. Besides VLE data, literature provides us with a large set of activity coefficients at infinite dilution. For such data the standard deviation of measurement is in the order of 10%. Note that in some cases the experimental activity coefficient of volatile compounds are subjected to systematic errors of calculation, due to the assumption that the gas is ideal. Excess enthalpy data, which measures the temperature dependency of the activity coefficient, is in general of good quality and sufficiently available, and enables us to determine the temperature dependence of activity coefficients.

5.3.1 The total activity coefficient of the cEOS

As explained above, we can not apply the cEoS-based activity coefficient equation, Eq. 5.18, directly in the form of Eq. 5.37, due to the systematic deviation in the parameter K. The theoretical minimum value for parameter K is 9 for the van der Waals cEoS, while the two-parameter cEoS between the SRK and the PR model give values in the range $10 \lesssim K \lesssim 12$. The three-parameter cEoS of CM gives values of the same magnitude (See appendix H). In general these K-values yield too large activity coefficients, and thus re-evaluation of K is inevitable. The optimization of K was performed for two data sets, that were retrieved from DDBST [38]. Data set 1 contains the infinite dilution activity coefficients of linear n-alkane binary mixtures. Data set 2 contains data of alkane binary mixtures with at least one them branched. We use these two classes, because they enable to quantify differences in the value of K and help us to understand the deficiencies of the cEoS-based activity model. From both experimental data sets we excluded experimental data, which shows a systematic deviation. These are 1) data having limiting activity coefficients above 1, 2) data having limiting activity coefficients lower than 0.8, while the difference in carbon number is smaller than 2 carbons, and 3) data, where the activity coefficient is too close to unity than one can expect from a physical point of view. These excluded data points are shown as crosses in the left and right panel of Fig. 5.6. The left panel depicts data set 1 covering a carbon range from butane to tetracosane. The right
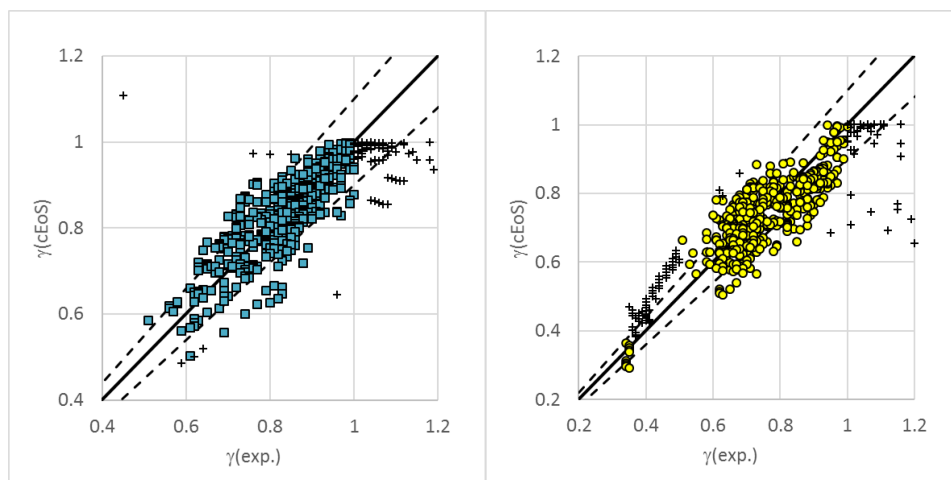
Figure 5.6: Correlation plot of activity coefficients at infinite dilution for data set 1 of linear alkanes (left panel) and data set 2 of branched alkanes (right panel) between experimental data and results obtained with Eq. 5.37 Data retrieved from DDBST [38]. Cross symbols are rejected experimental data points, as explained in text.

panel shows the correlation for data set 2 within the carbon range of dimethylbutane to dioctadecyldotetracontane. Table 5.6 shows that data set 2 requires a 26% lower value for K than data set 1. Fig. 5.7 shows the results of the limiting activity

Table 5.6: Fit results for cEoS parameter K using Eq.5.37 as depicted in Fig. 5.6.

Set	N(total)	N(excluded)	K	AAD%
1	690	95	4.13	4.9
2	655	109	3.04	7.8

coefficients as function of alkane number for linear and branched alkane mixtures. The description of the experimental limiting activity coefficients with the cEoS-based activity model looks fine at first sight, the average error is 5 to 8 %, although the model is not optimal, as is illustrated by the experimental activity coefficient data of Ashworth [6] depicted in Fig. 5.8. We observe that with increasing chain length the calculated values deviate more from the experimental values. This shows that the constant K needs adjustment for each binary.

Secondly, as has been noted in the theory section, the cEoS based activity models can not describe the temperature dependency of the enthalpy of mixing. This is illustrated by the binary system n-hexane - n-hexadecane (Fig. 5.9). The cEoS-based activity model gives a correct endothermal description, but the isothermal curves at
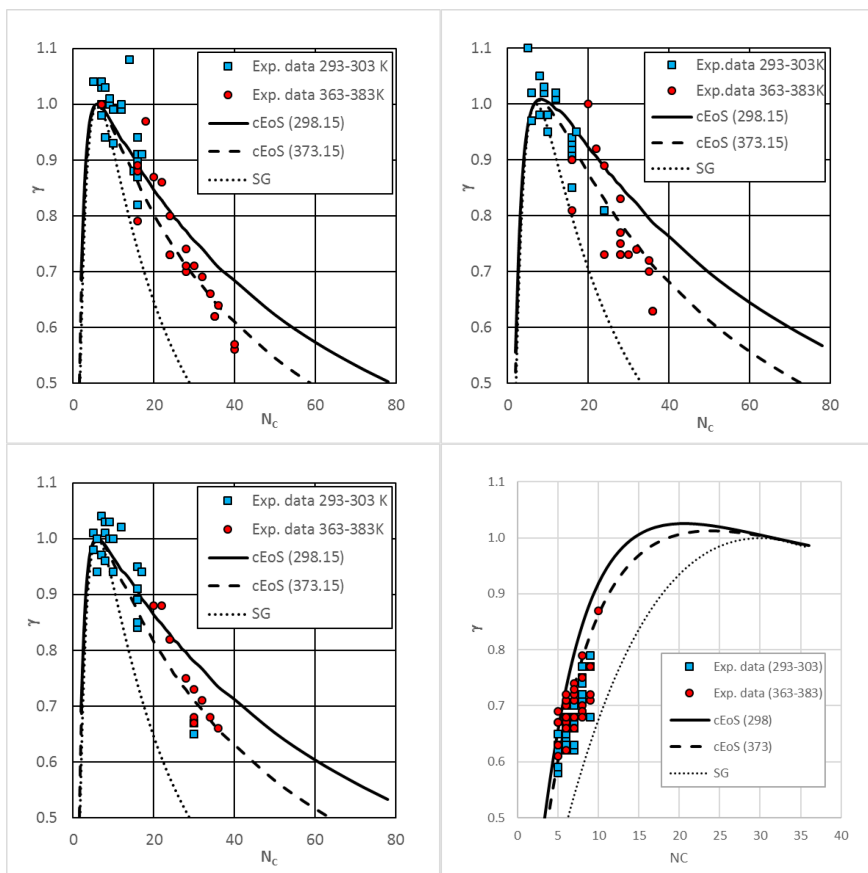
Figure 5.7: Experimental data (symbols) compared to cEoS activity model prediction (curves) for hexane in n-alkanes (top left), heptane in n-alkanes (top right) with $K = 4.13$, and 2-methylpentane in alkanes (bottom left) and alkanes in squalane (bottom right) with $K = 3.04$. In these plots the top and middle curves (Eq. 5.37) are the results at indicated lower and higher temperatures, respectively. The bottom dashed curves are the predictions of the combinatorial contribution Eq. 6.36.

298.15, 303.15 and 313.15 K coincide, while in reality there is a clear reduction in enthalpy with increasing temperature, and the maximum of the model is shifted to a higher concentration than is measured. This implies that the parameter K is not only temperature dependency, but also concentration dependent.
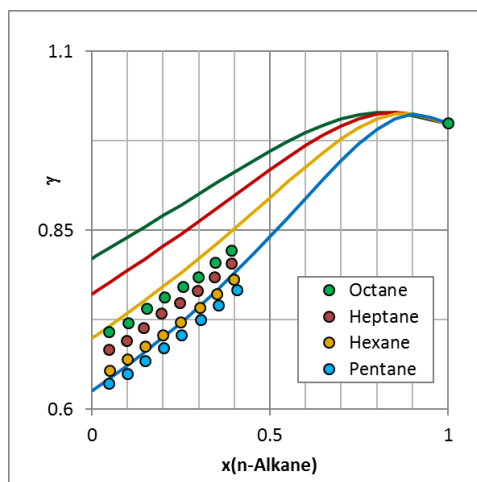
Figure 5.8: Experimental activity coefficient data from [6] (symbols) and theoretical cEoS activity model predictions Eq. 5.37, with $K = 3.04$, for the binary systems n-pentane, n-hexane, n-heptane and n-octane in squalane at 298.15 K.
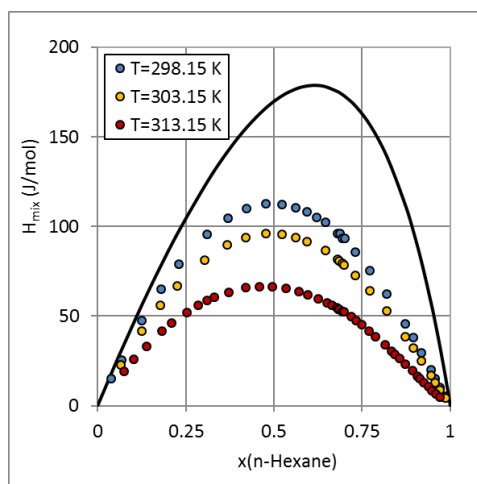
Figure 5.9: Experimental enthalpy of mixing data in comparison to cEoS activity model for the binary hexane - hexadecane. Solid curve is Eq. 5.32 which is temperature independent. Data from [167, 162].

5.4 Concluding remarks

A general expression for the activity coefficient of molecules that interact by dispersion only was derived from the cubic equations of state (cEoS). We have shown that

the cEoS based activity models can be characterized by a single parameter K, which for the van der Waals, Soave-Redlich-Kwong and Peng-Robinson equations of state has the values 9, 10.3 and 11.9, respectively. Experimental data, however, indicate that $K = 4.13$ for linear and $K = 3.04$ for branched alkanes. With these parameters we obtain realistic predictions without the necessity to optimize the size parameter m_j and the solubility parameter Δ_j. Both dimensionless parameters can be calculated from the normalized critical experimental parameters of the compounds using methane as a spherical reference, or by using a group contribution method. However, we also encountered small, but systematic deviations. Binary mixtures of n-alkanes with squalane increasingly deviate with increasing chain length of the n-alkane.

The cause for the systematic differences between the cEoS based activity coefficient model and the experimental results lie in two assumptions that were made in the derivation of the cEoS. The largest impact has the assumption that all molecules are spherical. This leads to a systematic error, because the ratio of the covolume and the van der Waals volume is taken to be 4, while it decreases to nearly 1.9 for long chains. In relation to this we also showed that the combining rules of Waldman and Hagler, and of Kong reduce to the classical combining rule of Lorentz and Berthelot, as soon as the mixture contains molecules which are polyatomic. The use of the classical combining rules for dispersion therefore appears to be sufficient, at least for alkanes. It might not apply to mixtures containing molecules with large atoms, such as the halogenated hydrocarbons.

The assumption that all molecules interact with the same extent with the surroundings is another assumption that is not valid for polyatomic molecules. In the van der Waals mean-field approximation the average interaction between two particle centers is taken as the integral of interaction of a particle in a uniform spherical potential field in analogy of a charge in a uniform electric potential field. Applying this methodology neglects the fact that polyatomic molecules behave differently. For polyatomic molecules one part of the molecule can prevent another part to interact with its immediate surrounding; end groups have more space to interact than chain groups, or groups located near branches. Equivalently, the intramolecular interaction and the covalent bonds reduce the number of intermolecular interactions. This is reflected by the reduced parameter K, which drops further when the molecule contains branches. This difference between mono- and polyatomic molecules becomes more pronounced when the density of the fluid is high, because the effective number of interactions increases with density. Effectively this results in a smaller value of cEoS parameter a for polyatomic molecules than the value of a obtained from the critical parameters of the compound, because proximity effects are less prominent at the critical point.

Another deficiency of the cEoS-based activity coefficient contribution is that,

although the model yields a reasonable result for the enthalpy of mixing, it can not describe its temperature dependence. This is cause by the Huron-Vidal method, which sets the thermal expansion of the liquid to zero. One way to address this, would be to make the parameter K temperature-dependent, for instance, by scaling it with the liquid volume at given temperature and the reference temperature T_0.

The generalized form of the cEoS-based activity coefficient model demonstrates that there is a physical way to close the gap between the combinatorial and the experimental activity coefficient for systems, in which molecules interact by dispersion only. The introduced GC-method could be used in models, like the original UNIFAC [59] and the family of COSMO-RS models, to account only for the dispersion interaction between alkyl groups.

In part 2 we will derive from the perturbed hard-sphere chain equations of state a more accurate dispersion activity coefficient equation.

Chapter 6

Dispersion ACMs.
Part II: Perturbed Chain
Equations of State[1]

6.1 Introduction

In this chapter we aim at developing a contribution for the activity coefficient that accurately accounts for the dispersion interaction between molecules. We have shown in part 1 that the cubic equations of state (cEoS) give a van Laar type of activity model, and that these models systematically predict too high activity coefficients. This systematic deviation is caused by the assumptions embedded in the van der Waals cEoS. It is assumed that all molecules are spherical, which sets the normalized excluded covolume to 4, while linear molecules have a value close to the covolume of cylinders, which is 2. Van der Waals further assumed that the effective number of interactions around a molecule is not influenced. However, the polyatomic character of most molecules effectively reduce intermolecular interaction, because of proximity effects of the same molecule. The change of the cEoS parameters a and b, as used in the van Laar activity coefficient equation, into the liquid volume and the Hildebrand solubility constant, respectively, is an ad-hoc way to eliminate this systematic error. We showed in part 1 that all van Laar type equation can be characterized by a unique parameter, K, which is directly related to the critical parameters and volume correction parameters of the cEoS. This parameter, K, however, needs adjustment

---

[1]Based on the publication: Dispersion activity coefficient models. Part 2: Perturbed Chain equations of state. Fluid Phase Equil. (2019), by Gerard J.P. Krooshof, Remco Tuinier, and Gijsbertus de With.

in order to give reasonable predictions, but it does not fully address the differences in molecular shape and polyatomic character of molecules. In this work we evaluate whether this issue can be solved by using as starting point the perturbed chain equations of state. It is noted that Lee and Sandler [134] derived an activity coefficient model from a perturbed mean sphere equation of state. In this model small molecules are regarded as spheres. This approach contains a systematic error for non-spherical molecules in the form of overestimating the number of interaction pairs around a molecule, because spheres do not have intramolecular interactions and covalent bonds that hinder the intermolecular interaction.

Another dispersion activity coefficient model is that of Hsieh *et al.* [93], who applied this within the COSMO-SAC model. However, their proposed dispersion activity coefficient model is a crude simplification. First, they use the single parameter Margules model, which indicates that dispersion is treated as an interaction of entities of the same size. Secondly, their choice of a fitted coordination number of 0.275 is physically impossible. Third, the unlike energy parameter for alkanes is zero, thereby yielding no dispersion contribution for alkane mixtures, which is incorrect. To this we mention that the MOSCED model [214] of Thomas *et al.* is an activity coefficient model that also includes dispersion interaction. The model is limited its use, as it provides the activity coefficient at infinite dilution of binary mixtures only. UNIQUAC [2], UNIFAC [59], COSMO-SAC [145] and COSMO-RS [116] model are activity models that do not have a dispersion interaction contribution. In particular the UNIFAC, COSMO-SAC and COSMO-RS models can not describe accurately the phase equilibrium behavior of alkane mixtures. In the modified UNIFAC model [231] this omission has been repaired by introducing an ad-hoc modification of the combinatorial contribution proposed by Donahue and Prausnitz [43]. This correction has, however, no physical meaning.

In this work we start with the derivation of activity coefficient models from the perturbed chain equations of state, using the method of Huron and Vidal [97, 225], in which the pressure is set at infinity. Infinity in this context means that the molecules are set into close contact. The models for the dispersive activity coefficient are combined with an off-lattice combinatorial contribution [129] to obtain the total activity coefficient. We show that the existing parametrization of the perturbed chain models is not applicable in the limit of infinite pressure, and that the theory can be improved by using another method to compute the first perturbation integral at this limit. This method can be formalized by topology theory, where the first Zagreb index is the key property. In addition, we introduce a new method to calculate the interaction energy between methyl groups of hydrocarbons, in which only the number of hydrogen atoms per carbon atom is required. This model is evaluated using three combining rules, compared to experimental data and applied to predict

phase equilibria of systems that were outside the set of used experimental data. In the last section we summarize our findings and discuss the key elements obtained.

6.2 Theory

6.2.1 Perturbed chain equations of state

Besides the cEoS, a whole series of equation of state models has been developed that specifically account for the chain-like structure of molecules [14, 44, 29, 28]. One of the successful models is the Perturbed Chain Statistical Associating Fluid Theory (PC-SAFT), developed by Gross and Sadowski [78] in 2001. This model yields excellent vapor-liquid phase equilibria predictions of hydrocarbon mixtures. Therefore, it is expected that PC-SAFT provides a better way to define the dispersive term in activity coefficient models. One of the main differences between the cEoS models and PC-SAFT is that the cEoS, due to the spherical nature of the molecule, treats the composition and density of the space around a molecule as uniform, while PC-SAFT accounts for local composition. Because PC-SAFT considers chains instead of spheres, a part of the space around a sphere in a chain is filled by neighbor spheres of the same chain, which increase the density locally, while the remaining part is uniform and determined by the statistics of chain collisions. We note further that prior to the PC-SAFT model Gross and Sadowski [77] developed a perturbed hard-sphere chain (PHSC) model. It differs from the PC-SAFT model in the definition of the segment diameter, which is in PHSC temperature independent.

In PC-SAFT the Helmholtz energy for dispersive interaction a^{disp} is expressed by:

$$\frac{a^{\text{disp}}}{RT} = -\frac{1}{2} \left[4\pi\rho I_1 \right] \overline{m^2\epsilon\sigma^3} - \pi\rho\overline{m}C_1 I_2 \overline{m^2\epsilon^2\sigma^3}, \tag{6.1}$$

where ρ is the total number density of the fluid, I_1 and I_2 are the first and second order perturbation integrals, $\overline{m}$ is the the average chain length of the molecules in the system, $\overline{m^2\epsilon\sigma^3}$ and $\overline{m^2\epsilon^2\sigma^3}$ are energy averages of the system energy and C_1 is a function. The first perturbation integral is defined by:

$$I_1(\eta, m) = -\int_0^\infty \tilde{u}(x) g^{\text{hc}}(m; x, \rho) x^2 dx, \tag{6.2}$$

where $\tilde{u}$ is the reduced pair potential and g^{hc} the pair correlation function with x the scaled center-to-center distance between spheres. The second perturbation integral I_2 and the function C_1 are defined by:

$$I_2(\eta, m) = \frac{\partial}{\partial\rho} \left[\rho \int_0^\infty \tilde{u}^2(x) g^{\text{hc}}(m; x, \rho) x^2 dx \right], \tag{6.3}$$

and

$$C_1 = \left(1 + Z^{\text{hc}} + \rho \frac{\partial Z^{\text{hc}}}{\partial \rho}\right)^{-1}$$
$$= \left(1 + \overline{m}\frac{8\eta - 2\eta^2}{(1-\eta)^4} + (1-\overline{m})\frac{20\eta - 27\eta^2 + 12\eta^3 - 2\eta^4}{[(1-\eta)(2-\eta)]^2}\right)^{-1}, \qquad (6.4)$$

respectively, where Z^{hc} is the compression factor of a hard-sphere chain fluid and η is the total packing fraction given by:

$$\eta = \frac{\pi}{6}\rho \sum_j x_j m_j d_j^3, \qquad (6.5)$$

where d_j is the segment size of compound j, which is a constant in the PHSC model and temperature dependent in the PC-SAFT model:

$$d_j = \begin{cases} \sigma_j & \text{PHSC} \\ \sigma_j \left[1 - 0.12\exp\left(-\frac{3\epsilon_j}{k_{\text{B}}T}\right)\right] & \text{PC-SAFT.} \end{cases} \qquad (6.6)$$

In the PC-SAFT model alkanes larger than ethane have $\epsilon_j/k_{\text{B}} > 200$ K, which implies that the segment diameter shrinks at most 2% going from 0 to 350 K. In other words, the temperature dependence is very weak and will only be noticeable in isobaric equilibria with large temperature differences. The averages involved are defined by:

$$\overline{m} = \sum_j x_j m_j$$
$$\overline{m^2\epsilon\sigma^3} = \sum_{j,k} x_j x_k m_j m_k \frac{\epsilon_{jk}}{k_{\text{B}}T}\sigma_{jk}^3$$
$$\overline{m^2\epsilon^2\sigma^3} = \sum_{j,k} x_j x_k m_j m_k \left(\frac{\epsilon_{jk}}{k_{\text{B}}T}\right)^2 \sigma_{jk}^3, \qquad (6.7)$$

where ϵ_{jk} is the average square-well potential depth for the dispersion interaction between a segment of molecule j and a segment of molecule k, σ_{jk} is the center-to-center distance between these segments, m_j and m_k are the number of segments in a chain of molecule type j and molecule k, respectively, and k_{B} is Boltzmann's constant. The calculation of ϵ_{jk} and σ_{jk} of the unlike pairs will be explained in section 6.2.3.

For a pure component, of which $d_j \approx \sigma_j$, we can define the number of interacting segments around a central segment by:

$$Z_j = 4\pi\rho I_1 m_j \sigma_j^3 = 24\eta I_1 \frac{m_j \sigma_j^3}{m_j d_j^3} \approx 24\eta I_1, \tag{6.8}$$

where Eq. 6.1 is used. This number is also known as the coordination number of the average segment of compound j.

In order to arrive at an expression for the activity coefficient from the perturbed hard-sphere chain models, we set the system to infinite pressure according to method that was applied by Huron and Vidal [97, 225] for cEoS. At this condition the liquid is compressed to its maximum value ($\eta_{\text{LCP}} \simeq 0.637$). We note that solid structures, for which the perturbed hard-sphere chain models do not apply, can have a packing fraction between that of a random close-packed (RCP) solid ($\eta_{\text{RCP}} \simeq 0.637$), and a face-centered cubic (FCC) ordered close-packed structure ($\eta_{\text{FCC}} = \pi/\sqrt{18}$). The value for the solid depends on the process path towards infinite pressure. For the liquid state it is important to understand that at LCP the very large compression factor of the hard-sphere chain, Z^{hc}, brings C_1 to zero (see appendix M). Hence, at infinite pressure only the first perturbation integral matters.

First perturbation integral

In the PC-SAFT model the first perturbation integral is represented by a power series introduced by Liu and Hu [148]:

$$I_1 = \sum_{j=0}^{6} \hat{a}_j \eta^j. \tag{6.9}$$

The constants, $\hat{a}_j$, are given by the expressions:

$$\hat{a}_j = \hat{a}_{0,j} + \hat{a}_{1,j}\frac{m-1}{m} + \hat{a}_{2,j}\frac{m-1}{m}\frac{m-2}{m} \tag{6.10}$$

in which m is the number of segments in a chain, and $\hat{a}_{0,j}$, $\hat{a}_{1,j}$, and $\hat{a}_{2,j}$ are model constants. Substitution of η by η_{LCP} into Eq. 6.9 gives:

$$I_1\left(\eta_{\text{LCP}}\right) = \tilde{a}_1 + \frac{\tilde{a}_2}{m} + \frac{\tilde{a}_3}{m^2}, \tag{6.11}$$

where $\tilde{a}_1$, $\tilde{a}_2$, and $\tilde{a}_3$ are constants given in Table 6.3. The PHSC model [77] contains the same equations as the PC-SAFT model, but uses a square-well potential for

the interactions, instead of the softer, modified square-well potential that has been defined for the PC-SAFT EoS (see Eq. 6.6). The 21 parameters for the first perturbation integral of the PHSC model were optimized by a set of generated data points, for which equation 6.2 was used, in the ranges of $0 < \eta < 0.6$ and $1 < m < 1000$, while those of the PC-SAFT model parameters were determined by fitting experimental vapor pressures and liquid densities of the n-alkanes up to decane, except ethane, in. The temperature range given in Table 2 of the PC-SAFT paper [78], see also [76], indicate that the experimental data covers a packing ranges of $0 < \eta < 0.45$ (though a valid range to 0.74 is stated). Due to the difference in the potential well and the applied strategy to determine the parameters of Eq. 6.9, the values of the parameters of Eq. 6.11 for the PHSC and the PC-SAFT model are different (see Table 6.3). The top panels of Fig. 6.1 show the value of I_1 as function of η for both models from zero to 0.74. The bottom panels show the coordination number of a pure component j when $d_j = \sigma_j$ as function of the packing fraction. We have plotted results for various m values up to 20, because for longer chains the results are very close to $m = 20$. We observe that the first perturbation integral and the coordination number of the PHSC model run nearly parallel and that with increasing chain length the difference in I_1 becomes smaller. This is not the case for the PC-SAFT model. First, we observe that the curves of the PC-SAFT model start at a significantly higher value at $\eta = 0$. From a theoretical point of view it is expected that the relative difference would be at most 3%, as explained in J. Another point of concern is that all curves of the PC-SAFT model, except for methane, converge just outside the experimental data range at $\eta \approx 0.46$ to a value of $I_1 \approx 1.05$, cross and diverge in the high η limit. This is physically impossible, because long chains have more covalent bonds, and therefore miss more sites for intermolecular interaction. This crossing point lies outside the range of experimental data and, therefore, should be regarded as an artifact of the parametrization process. For the use of the PC-SAFT EoS model in fluid phase equilibria description this issue is not a problem, because the model is meant to be applied for fluids with $\eta < 0.46$. But for solid-liquid equilibria or gas-liquid phase equilibria at very high pressures, the PC-SAFT EoS could give erroneous results for the liquid phase, because the packing fraction of the liquid can be larger than 0.46. This problem has been studied by Yelash *et al.* [241], and Privat *et al.* [181]. On the other hand Polishuk [179] showed that also SAFT contains certain pitfalls. In short, these publications indicate that re-optimization of the I_1 parameters is required to obtain a proper liquid description beyond $\eta = 0.45$. This has been done by Liang and Kontogeorgis [140]. They have proposed a new set of model parameters, but as can be noticed in Fig. 7 of this publication, also here exists a cross-over of the I_1 curves, which makes extrapolation to the solid state impossible. We therefore need to optimize the $\tilde{a}_1$, $\tilde{a}_2$, and $\tilde{a}_3$ parameters to define the
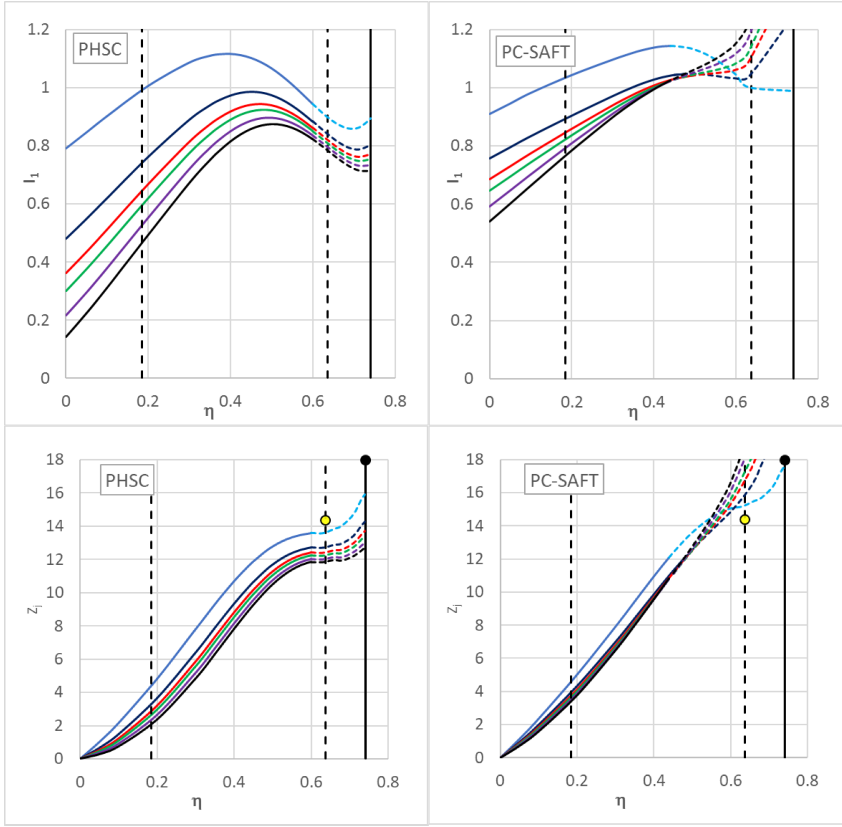
Figure 6.1: First perturbation integral, I_1, (top panels), and coordination number, Z_j (bottom panels), of a pure component as function of the packing fraction given by PHSC and PC-SAFT (tags). Curves show the results for chain lengths of (from top to bottom) $m = 1, 2, 3, 4, 7,$ and 20 segments. The solid part of each curve indicates the experimental data range, while the dashed part is an extrapolation of Eqs. 6.9 and 6.8, respectively. The vertical dashed lines indicate the critical volume ($\eta \approx 0.18$) and LCP ($\eta \approx 0.64$). The solid vertical line denotes $\eta \approx 0.74$. The yellow and black points indicate the coordination number of single spheres, $m = 1$ at LCP. These are $Z_j = 14.4$, according to Klumov's work [119], and at FCC $Z_j = 18$, respectively.

first perturbation integral at maximum density. The work of Klumov [119] points out that the number of hard spheres is 14.4 at liquid closed packed condition ($\eta = 0.64$), when the integration runs between $1 < r/\sigma < 1.5$. This is indicated by the yellow data point in the bottom panels of Fig. 6.1. Interesting to observe is that the PHSC model gives $Z_j = 13.7$ at $m = 1$, which is close to Klumov's value. We will use Klumov's value, $Z_j = 14.4$, as the limit for monosegmented molecules.

Perturbed chain activity coefficient model

At infinite pressure the Helmholtz energy of mixing for the PC-SAFT model is determined by I_1 only (see Appendix M). With the definition for the coordination number, Eq. 6.8, it has the following expression:

$$\frac{n_\text{t} a^{\text{E,disp}}}{RT} = \left(\frac{T_0}{T}\right)\left[\sum_j \frac{Z_j}{2} n_j m_j \tilde{\epsilon}_j \left(\frac{\sigma_j}{d_j}\right)^3 - \frac{\sum_{\text{j,k}} Z_j n_j n_k m_j m_k \tilde{\epsilon}_{jk} \sigma_{jk}^3}{2\sum_j n_j m_j d_j^3}\right], \quad (6.12)$$

where $\tilde{\epsilon}_j$ is the reduced dimensionless interaction energy at reference temperature, T_0, defined by:

$$\tilde{\epsilon}_j = \frac{\epsilon_j}{k_\text{B} T_0}, \quad (6.13)$$

and

$$Z_j = 24\eta_{\text{LCP}}\left(\tilde{a}_1 + \frac{\tilde{a}_2}{m_j} + \frac{\tilde{a}_3}{m_j^2}\right). \quad (6.14)$$

For single spheres we have $Z_j = 24\eta_{\text{LCP}}(\tilde{a}_1 + \tilde{a}_2 + \tilde{a}_3)$. This gives for the PHSC and PC-SAFT model $Z_j = 13.7$ and $Z_j = 15.3$ respectively. For infinite homonuclear chains, where $m = \infty$, the last two terms of Eq. 6.14 vanish, and we get $Z_j = Z_\infty = 24\,\eta_{\text{LCP}}\,\tilde{a}_1$. For this case the PHSC and PC-SAFT model have $Z_j = 11.88$ and 19.10, respectively. Since the face centered cubic (FCC) or hexagonal close-packed (HCP) packing of spheres gives the maximum value of the coordination number $Z_j = 18$, the result of PC-SAFT model for the LCP condition is physical impossible. It is a mathematical error caused by the extrapolation beyond the singular point $\eta > 0.46$, as we pointed out when discussing the results of Fig. 6.1.

Assuming that the weak dispersion gives negligible excess volume, we have $g^E \approx a^{\text{E,disp},\infty}$, and the activity coefficient γ_k can be calculated from:

$$\ln\gamma_k = \frac{1}{RT}\left(\frac{\partial n_\text{t} g^\text{E}}{\partial n_k}\right)_{T,P,n_j\neq k} \approx \frac{1}{RT}\left(\frac{\partial n_\text{t} a^\text{E}}{\partial n_k}\right)_{T,P,n_j\neq k}, \quad (6.15)$$

Substitution of Eq. 6.12 in Eq. 6.15 gives the expression

$$\ln\gamma_k^{\text{disp}} = m_k \frac{Z_k}{2}\left(\frac{T_0}{T}\right)\left(\left(\frac{\sigma_k}{d_k}\right)^3 \tilde{\epsilon}_k - 2\frac{\sum_j x_j m_j \tilde{\epsilon}_{jk}\sigma_{jk}^3}{\overline{md^3}} + \frac{d_k^3\overline{m^2\epsilon\sigma^3}}{\left(\overline{md^3}\right)^2}\right)$$
$$+ m_k\frac{Z_\infty}{2}\left(\frac{T_0}{T}\right)\left[\frac{\tilde{a}_2}{\tilde{a}_1\overline{m}^2} + \frac{2\tilde{a}_3}{\tilde{a}_1\overline{m}^3}\right]\left(1 - \frac{\overline{m}}{m_k}\right)\frac{\overline{m^2\epsilon\sigma^3}}{\overline{md^3}}, \quad (6.16)$$

where we defined the average

$$\overline{md^3} = \sum_j x_j m_j d_j^3.$$ (6.17)

Note that the volume term is absent in Eq. 16, as at infinite pressure the fluid is incompressible ($\eta = \eta_{\text{LCP}}$). The activity coefficient of compound k at infinite dilution in solvent j then reads:

$$\ln \gamma_k^{\text{disp},\infty} = m_k \frac{Z_k}{2} \left(\frac{T_0}{T} \right) \left(\frac{\sigma_k^3}{d_k^3} \tilde{\epsilon}_k - 2 \frac{\sigma_{jk}^3}{d_j^3} \tilde{\epsilon}_{jk} + \frac{d_k^3}{d_j^3} \frac{\sigma_j^3}{d_j^3} \tilde{\epsilon}_j \right)$$

$$+ m_k \frac{Z_\infty}{2} \left(\frac{T_0}{T} \right) \left[\frac{\tilde{a}_2}{\tilde{a}_1 m_j} + \frac{2\tilde{a}_3}{\tilde{a}_1 m_j^2} \right] \left(1 - \frac{m_j}{m_k} \right) \left(\frac{\sigma_j}{d_j} \right)^3 \tilde{\epsilon}_j.$$ (6.18)

In here the segment diameter d_j is defined by Eq. 6.6. With $d_j = \sigma_j$ the finite and limiting activity coefficients of the PHSC model is obtained. In appendix K we demonstrate how the PC-SAFT based dispersion activity coefficient model is related to the equation of van Laar [223] and the regular solution theory of Hildebrand [86]. Table 6.1 provides an overview of PC-SAFT parameters of some n-alkanes as published by Gross and Sadowski [78], as well as the parameters of squalane, which we obtained by fitting experimental saturated vapor pressure and liquid density data taken from DIPPR [23]. More m, σ, and ϵ values can be found in the work of Gross [78] and Tihic *et al.* [215]. Tihic *et al.* provides also equations to predict these parameters for alkanes with $N_{\text{C}} > 20$, which we will use in the results section.

The last three columns of Table 6.1 list the calculated values for Z_j, showing how the coordination number of the alkanes vary with chain length and structure. Both PHSC and PC-SAFT use Eq. 6.14, but the results differ due to the difference in $\tilde{a}_j$ parameters given in Table 6.3. Squalane has a smaller coordination number than an infinite n-alkane, because the presence of methyl side groups along the squalane chain reduce the space of interaction. In general, small molecules should have more interaction per segment with the surrounding molecules than large molecules, because the space around such a small molecule is less hindered by other parts of the molecule. Therefore, methane should interact with more spheres than hexadecane per segment, which is also observed in Table 1 for the PHSC and the IPC model, but not for the PC-SAFT due to the error of extrapolation.

6.2.2 The improved perturbed chain activity coefficient model

Based on the results of the PC-SAFT model, Eq. 6.16 in combination with Eq. 6.36, as depicted in Fig. 6.4, we concluded that the $\tilde{a}_j$ parameters of the PC-SAFT model

	PC-SAFT			PHSC	PC-SAFT	IPC
	EoS and activity model			$Z_j\,(\eta_{\mathrm{LCP}})$		
Alkane	m	σ (Å)	$\tilde{\epsilon}$	Eq. 6.14	Eq. 6.14	Eq. 6.22
C1	1	3.7039	0.503	13.72	15.30	14.40
n-C5	2.6896	3.7729	0.775	12.59	16.41	11.60
n-C6	3.0576	3.7983	0.794	12.50	16.66	11.40
n-C7	3.4831	3.8094	0.800	12.43	16.89	11.26
n-C8	3.8176	3.8373	0.814	12.38	17.05	11.16
n-C12	5.306	3.8959	0.836	12.24	17.55	10.90
n-C16	6.6485	3.9552	0.854	12.17	17.83	10.78
n-C$_\infty$	∞	4.0829	0.882	11.84	19.10	10.40
Squalane	10.37 (3)	4.153 (6)	0.858 (1)	12.06	18.37	10.20

Table 6.1: PC-SAFT parameters for n-alkanes [78, 215] with $T_0 = 298.15$ K. Squalane parameters were obtained in this work by regression of liquid density and saturated vapor pressure data. Between parentheses the standard deviation in the last digit. The coordination numbers are obtained by Eq. 6.14 with the $\tilde{a}_j$ parameters from Table 6.3. In the last column the results of the IPC model where Z_j is defined by Eq. 6.22

.

for I_1 (Eq. 6.11) are not suitable to obtain an accurate value for Z_j at LCP condition. In order to arrive at a more suitable description, we take a closer look at I_1, which effectively means that we investigate the way in which the coordination number of a sphere in a chain is defined. This leads to a dispersion activity coefficient model, which we call the improved perturbed chain activity coefficient model, abbreviated to IPC model.

Perturbation integral at infinite pressure

The general idea for the IPC model is that under high pressure the molecules will be at their most compact conformations. Both the PHSC and the PC-SAFT model take the ad-hoc interaction range $1 \leq x/\sigma \leq 1.5$ to define a shell of interaction around a sphere. This choice gives in a hexagonal close-packed structure (HCP) 18 spheres within this shell. There are 12, respectively, 6 spheres in the first and second shell around a central sphere. The third shell of spheres is located at $\sqrt{3}\sigma > 1.5\sigma$, which falls outside this range. This situation is valid for solids.

For liquids the number of spheres within the interaction range is lower. According to the works of Bennett, Parisi and Klumov [175, 13, 119] there are 6 spheres in direct contact with the central sphere. Scott [201] showed that besides these 6 there are on average 3.3 spheres, which are are not in direct contact with the central

sphere, but still belong to the first shell. Klumov *et al.* [119] calculated with a modified Lubachevsky-Stillinger algorithm the radial and cumulative distribution of hard spheres in the packing fraction range of 0.64 to 0.68. In Fig. 1 of Klumov's paper we can read that at LCP the average total number of hard spheres within the PC-SAFT defined shell is $Z_j = 14.4$. Hence, the average number of spheres, which are not in direct contact with the central sphere are $N_{s2} = 14.4 - 6 = 8.4$. The value of $Z_j = 14.4$ is valid for disconnected spheres. We denote these covalent bonded spheres in the first and second coordination shell by N_{x1} and N_{x2}, respectively. The number of first (N_{x1}) and second (N_{x2}) bonded spheres in a linear chain is given by:

$$N_{x1} = 2\frac{m-1}{m} \quad \text{and} \quad N_{x2} = 2\frac{m-2}{m}, \tag{6.19}$$

respectively. These equations hold for linear molecules with $m \geq 2$. Methane has no carbon bonds; $N_{x1} + N_{x2} = 0$.

For non-linear chains, or in general poly-segmented structures, the total number of spheres in the range of interaction of a central sphere will be higher than that of a linear chain due to the presence of branches. At this point our model starts to become different from the PC-SAFT and the PHSC model, where all molecules are treated as a homonuclear linear chains and where the effects of more space at chain ends and less space at branches are ignored. This approach is known as the Percus-Yevick 2 approximation [30]. In our proposed model, denoted as improved perturbed chain (IPC), we will account for branching and proximity effects, for which we use the carbon skeleton of the molecule. This approach is known as the full Percus-Yevick approximation [30]. Hence, in the IPC model $m = N_C$, where N_C is the carbon number, while in the PC-SAFT model, according to Table of the original PC-SAFT paper [78], m assigns several methyl groups, which is about 2.775 for n-alkanes.

Fig. 6.2 schematically illustrates examples of poly-segmented chains, when they are at close packed condition.

Interestingly, topology theory [83, 37] provides a way to generalize Eq. 6.19 to branched structures, where proximity effects play no role. For $m \geq 2$, the sum of spheres bonded in the first and second coordination shell, which we will call the occupation number, can be calculated from the first Zagreb index [83, 37], ZM_1. This topology index was actually introduced for quantifying the polarity of molecules, but here it is the average occupation number of spheres in the interaction shell of a central sphere. The first Zagreb index for linear, branched and cyclic structure is calculated by:

$$ZM_1 = 0m_{v,0} + 1m_{v,1} + 4m_{v,2} + 9m_{v,3} + 16m_{v,4} = \sum_{j=0}^{4} j^2 m_{v,j}, \tag{6.20}$$
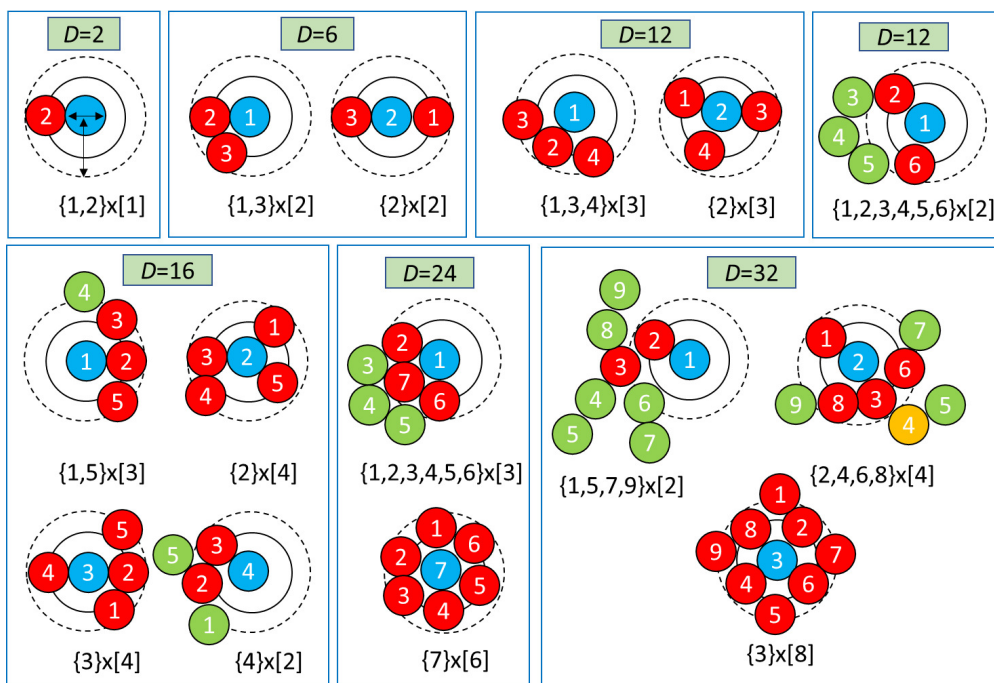
Figure 6.2: Schematic projected view of molecule configurations for the calculation of topology number D in the IPC model. Top row: ethane, propane, isobutane and cyclohexane. Bottom row: 2-methylbutane, methylcyclohexane, and 3,3-diethylpentane. Numbers between the accolades denote the carbon position in the molecule, which have the same number of spheres that occupy the square-well space (red spheres). The occupation number is denoted between square brackets. The green spheres are outside the interaction range of the central sphere (blue). The orange sphere is excluded due to proximity effect of adjacent spheres.

where $m_{v,j}$ is the number of carbon groups, which have j covalent bonds to other carbon groups. The subscript v comes from the topology term 'vertex', which is a connection point with a certain number of connections to other points. In chemistry a vertex means an atom, which has a certain number of bonds to other atoms. In our case the carbon-carbon bonds. For instance, CH_4 has no carbon connection, CH_3- has one, -CH=, =C= CH_2< groups have two bonds, -CH< three, and >C< four. By counting these carbons Eq. 6.20 gives ZM_1. The first Zagreb index can not be applied for molecular structures, where rigidity or proximity effects prevent that a next nearest neighbor sphere of the same molecule is within the shell of interaction as is defined in the PHSC and PC-SAFT model. It also happens for cyclic structures with no alkyl branches $N_{x1} = 2$ and $N_{x2} = 0$. It yields for cyclobutane, cyclopentane and cyclohexane $D = 8$, 10 and 12, while the Zagreb indices are 16,

20, and 24, respectively. To extend the ZM_1 we define the quantity D, which we call the dispersion topology number:

$$D = m_{v,1} + 4m_{v,2} + 9m_{v,3} + 16m_{v,4} + 2m_{v,c2}, \qquad (6.21)$$

where $m_{v,c2}$ denotes the number of cyclic carbons in a ring, which has no branches. There are however structures, for which it is more difficult to express D and which are more easily obtained graphically. For instance for structures where the methylene group can be pressed into the dimple of a cyclic structure, such as the alkylcyclohexanes (see Fig. 6.2). In Table 6.2 we give molecular structures that have a D number that differs from Eqs. 6.20 and 6.21, due to the rigidity of a ring or the proximity of groups. The effective coordination number Z_j can now be written as:

Compound	m	ZM_1	D	Z_j
Cyclohexane	6	24	12	12.40
Cycloheptane	7	28	14	12.40
Methyl cyclohexane	7	30	21	11.40
2,2,3-trimethyl butane	7	33	29	10.26
1,1-dimethyl cyclohexane	8	38	28	10.90
2,2,3-trimethyl pentane	8	34	30	10.66
2,3,3-trimethyl pentane	8	34	30	10.66
2,2,3,3-tetramethyl butane	8	38	30	10.66
1,2,4,-trimethylcyclohexane	9	42	30	11.06
3-ethyl,2,3-dimethyl pentane	9	38	34	10.62
3,3-diethyl pentane	9	36	32	10.84
3,3,4-trimethyl hexane	9	38	34	10.62
2,2,3,3-tetramethyl pentane	9	42	34	10.62
2,3,3,4-tetramethyl pentane	9	40	32	10.84
Adamantane	10	60	24	12.00
Bicyclohexyl	12	58	26	12.24

Table 6.2: Examples of alkanes for which D differs from ZM_1, due to rigidity of a cyclic structure or proximity of branched groups.

$$Z_j = N_{s1} + N_{s2} - N_{x1} - N_{x2} = Z_0 - \frac{D}{m}, \qquad (6.22)$$

where Z_0 has a value of 14.4, which holds for methane and other single spheres. The subscript j to Z denotes a compound number, to indicate that it is a pure compound specific quantity.

It is noteworthy to mention that an infinitely long linear chain, which has $D/m = 4$, gives with Eq. 6.22 $Z_j = 10.4$. This value is closer to the value of the PHSC-model at LCP, where $Z_j = 11.84$, than the unrealistic value of the PC-SAFT based model, where $Z_j = 19.10$. Salient is that the coordination number in the UNIQUAC and UNIFAC models has always been set to the value of 10. This value is obtained by considering a HCP lattice, where a central cell has 12 nearest neighbors, but of which 2 sites are already occupied by 2 neighbors of the same chain. There are, however, two concerns with this choice for the UNIQUAC and UNIFAC models. Firstly, molecules are not always infinite long, thereby giving on average a lower correction than 2 neighbors. For instance, small molecules like water, ammonia and methane should have at HCP $Z_j = 12$, while ethane, methanol require $Z_j = 11$. Secondly, the choice for a HCP lattice is valid for crystals, but a liquid would be better described by a RCP lattice, which has a coordination number of about 9.2 for the first shell according to the work of Scott [201]. This would lead to a value of 7 for UNIQUAC and UNIFAC, if one takes into account the occupation of 2 sites by the same molecule. In other words the assumption in UNIQUAC and UNIFAC that $Z_j = 10$ for all molecules is incorrect. With Eq. 6.22 this could be corrected.

For pure components where $\sigma_j = d_j$, Eq. 6.22 yields for the perturbation integral at LCP ($\eta = 0.64 \approx 2/\pi$) the formula:

$$I_1 = \frac{Z_j}{24\eta} = \frac{\pi}{48}Z_j = \frac{14.4\pi}{48} - \frac{\pi}{48}\frac{D}{m} = 0.942 - 0.0654\frac{D}{m}. \qquad (6.23)$$

According to this expression, the coefficients of Eq. 6.11 for single spheres, where $D = 0$, would be $\tilde{a}_1 = 0.942$, $\tilde{a}_2 = 0$, and $\tilde{a}_3 = 0$, and for linear chains, where $D = 4m - 6$, $\tilde{a}_1 = 0.680$, $\tilde{a}_2 = 0.393$, and $\tilde{a}_3 = 0$. The m^{-2} term is zero, which implies that in the power series of Liu and Hu [148] the third term should vanish at RCP. The PHSC model seems to point to this criterion. It has a small value for this term; $\tilde{a}_3 = -0.006$. We have denote the values of a linear chain as theoretical values in column 4 of Table 6.3.

Derivation of a new dispersion activity model

For the derivation of the dispersion activity coefficient model proposed here, which we will denote as the improved perturbed chain (IPC) activity coefficient model, we consider a mixture composed of n_j molecules of type j. The average segment of molecule j has interaction with Z_j spheres as follows from Eq. 6.22. The total number of interactions of all molecules j with its local surrounding is therefore:

$n_j m_j Z_j$. When we define the volume fraction of a compound by:

$$\hat{\phi}_i = \frac{n_i m_i d_i^3}{\sum_j n_j m_j d_j^3},\tag{6.24}$$

we can calculate the contribution in interaction energy between molecule j and segment i by $n_j m_j Z_j \hat{\phi}_i$. We note that at infinite pressure all compounds have a pure molar volume of $n_i m_i d_i^3 / \eta_{\text{LCP}}$. The factor η_{LCP} drops out of equation 6.24. Subsequently, the following expression for the excess Helmholtz energy of mixing is obtained:

$$\frac{n_t a^{\text{E,disp}}}{RT} = \sum_j n_j m_j \frac{Z_j}{2}\left(\frac{T_0}{T}\right)\left(\tilde{\epsilon}_j - \sum_i \hat{\phi}_i \tilde{\epsilon}_{ij}\right).\tag{6.25}$$

Here the factor $\frac{1}{2}$ is used to eliminate double counting of interactions. Substitution of Eq. 6.25 into Eq. 6.15 yields the following expression for the dispersive activity coefficient:

$$\ln \gamma_k^{\text{disp,IPC}} = m_k \frac{Z_k}{2}\tilde{\epsilon}_k \left(\frac{T_0}{T}\right) \times$$
$$\left[1 - \sum_j \left[1 + \frac{Z_j}{Z_k}\left(\frac{d_k}{d_j}\right)^3\right]\hat{\phi}_j \frac{\tilde{\epsilon}_{jk}}{\tilde{\epsilon}_k} + \sum_{i,j} \frac{Z_i}{Z_k}\hat{\phi}_i\hat{\phi}_j \frac{\tilde{\epsilon}_{ij}}{\tilde{\epsilon}_k}\right].\tag{6.26}$$

In the derivation of the IPC dispersion activity coefficient model m_k is defined by the total number of methyl groups, d_k follows from the van der Waals volume, and Z_k from Eq. 6.22. This implies that in this model $\tilde{\epsilon}_k$ is the reduced energy of an average methyl group in the molecule, which needs another quantification than the one which is used in the PC-SAFT model, because we have another definition for the number of segments m. For this we take a closer look at the theory on dispersion as was introduced by London [151], who quantified the dispersion energy U by the expression:

$$U = -\frac{3}{4}\frac{h\nu_0 \alpha^2}{R_0^6},\tag{6.27}$$

where h is the Plank constant, while ν_0, R_0 and α are the characteristic frequency, the distance, and the polarizability of the two identical systems, respectively. In our case the two identical systems are two identical methyl groups. Since the polarizability of a methyl group is linear related to the number of hydrogens in it (see Fig. 6.3), while the dispersion energy is related to the square of the polarizability, the dispersion
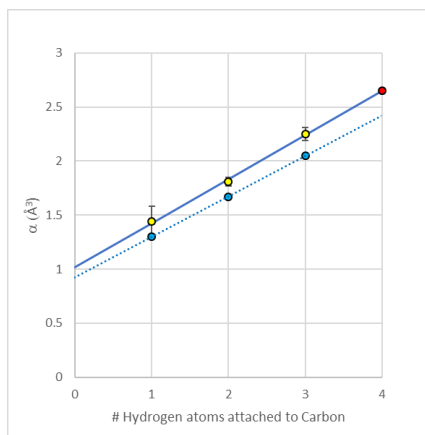
Figure 6.3: Polarizability of methyl groups as function of the number of hydrogen in a methyl group. Experimental data from [132, 169] yellow symbols, and from [222] red symbol. Computed data from [227] are indicated by blue circles.

energy should quadratically depend on the number of hydrogen atoms. We opted for the following relation:

$$\tilde{\epsilon}_j = \frac{\epsilon_0 + \epsilon_1 J_{\text{QH}}}{m_j}, \tag{6.28}$$

where ϵ_0 are ϵ_1 are IPC model constants. The parameter J_{QH}, which we have named the quadratic hydrogen number, is a constitutional index, and depends on the structure of the molecule. It is equated as:

$$J_{\text{QH}} = \sum_{j=0}^{4} (4-j)^2 m_{v,j}, \tag{6.29}$$

where j denotes the number of non-hydrogen bonds in a methyl group, and $m_{v,j}$ the number of methyl groups with j non-hydrogen bonds. The methyl groups -CH_3, $>CH_2$, $-CH<$ and $>C<$ have therefore $j = 1, 2, 3$ and 4, respectively. Let us exemplify Eq. 6.29 with the compound 2,2,4-trimethylpentane. This compound has 5 -CH_3 groups, 1 $>CH_2$ group, 1 -$CH<$ group and 1 $>CH$ group. Therefore, $J_{\text{QH}} = 9 \times 5 + 4 \times 1 + 1 \times 1 = 50$. This value is higher than that of the linear isomer, n-octane, where $J_{\text{QH}} = 42$. As one can observe from Eq. 6.29, in this approach carbons without hydrogen do not contribute to the total dispersion energy. The IPC activity coefficient equation differs from the PC-SAFT activity coefficient equation on three aspects. First, we use another definition for m. Second, we replaced in the numerator the hard-core segment size by a temperature-dependent diameter, because we consider volume fractions to define the number of segments of

a particular molecule within the range of interaction of a central segment. Third, the Z_j-term and the dispersion energy are calculated from two molecular descriptors, D and J_{QH}, respectively.

For convenience, we report that the infinite dilution activity coefficient (IDAC) of compound k in solvent j is given by:

$$\ln \gamma_k^{\mathrm{disp},\infty} = m_k \left(\frac{T_0}{T} \right) \left[\frac{Z_k}{2} \left(\tilde{\epsilon}_k - \tilde{\epsilon}_{jk} \right) + \frac{Z_j}{2} \left(\tilde{\epsilon}_j - \frac{d_k^3}{d_j^3} \tilde{\epsilon}_{jk} \right) \right]. \qquad (6.30)$$

For a mixture, where the molecules have the same diameter and $Z_k = Z_j$, and where the classical combining rules, i.e. Eq. 6.32, can be applied, Eq. 6.30 reduces to:

$$\ln \gamma_k^{\mathrm{disp},\infty} = \frac{m_k Z_k}{2} \left(\frac{T_0}{T} \right) \left[\sqrt{\tilde{\epsilon}_k} - \sqrt{\tilde{\epsilon}_j} \right]^2, \qquad (6.31)$$

which is a van Laar activity model. It points out that the difference between a dispersion activity coefficient models derived from a cubic and a perturbed chain dispersion activity coefficient model is a result of differences in segment diameters and coordination numbers, since branched structures exclude more segments from the spherical interaction volume than linear molecules do. Therefore Z_j is lower and by this the prefactor of the cEoS based dispersion activity model drops from 4.13 for linear alkane mixtures to 3.04 for mixtures involving branched alkanes.

6.2.3 Combining rules

The calculation of the unlike interaction dispersion energy, ϵ_{jk}, requires the use of a combining rule. In our previous paper on dispersion activity coefficient models derived from cubic equations of state [130], we evaluated three different rules. The generalized Lorentz [153] and Berthelot [17] combining rules, the Kong [123] combining rules and the Waldman and Hagler [228] mixture rules.

The generalized Lorentz size and Berthelot energy combining rules are:

$$\sigma_{ij} = (1 - l_{ij}) \left(\frac{\sigma_{ii} + \sigma_{jj}}{2} \right) \quad \text{and} \quad \epsilon_{ij} = (1 - k_{ij}) \sqrt{\epsilon_{ii} \epsilon_{jj}}, \qquad (6.32)$$

where σ_{jj} and ϵ_{jj} is the diameter of and the dispersion energy between two spheres of type j. The parameters k_{ij} and l_{ij} are the energy and size interaction parameters, respectively, which are used in the PC-SAFT equation of state to bring the calculated phase equilibrium results in agreement with experimental data. In the prediction of phase equilibria the parameters k_{ij} and l_{ij} are set to zero, which implies that for the unlike diameter the arithmetic mean, and for the unlike interaction energy, the

geometric mean is taken. We will denote this as the ideal classical (IC) combining rule. Other combining rules are those of Kong [123] and Waldman and Hagler [228].

Kong [123] sets three energy criteria for the repulsive and attractive part of the 6-12 Lennard-Jones potential and obtains combining rules, which we have rearranged into:

$$
\sigma_{ij}^6 = \frac{1}{\sqrt{\epsilon_{ii}\epsilon_{jj}}\sigma_{ii}^3\sigma_{jj}^3} \left[\frac{(\epsilon_{ii})^{\frac{1}{13}}(\sigma_{ii})^{\frac{12}{13}} + (\epsilon_{jj})^{\frac{1}{13}}(\sigma_{jj})^{\frac{12}{13}}}{2} \right]^{13} \quad \text{and}
$$

$$
\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}\frac{\sigma_{ii}^3\sigma_{jj}^3}{\sigma_{ij}^6}, \tag{6.33}
$$

with ϵ_{jj} and σ_{jj} are the Lennard-Jones energy and size parameters between two molecular segments of the same kind j, while subscript ij refers to the unlike pairs.

Waldman and Hagler (WH) [228] applied mathematical scaling and symmetry rules, which led to a less complicated equation for the unlike size parameter σ_{ij}:

$$
\sigma_{ij}^6 = \left(\frac{\sigma_{ii}^6 + \sigma_{jj}^6}{2} \right). \tag{6.34}
$$

The unlike energy parameter ϵ_{ij} is the same as that of Kong.

6.2.4   The total activity coefficient

In general, the total activity coefficient of molecules, which interact by dispersion forces only, is given by the following sum of a combinatorial and a dispersive contribution:

$$
\ln\gamma_k^{\text{tot}} = \ln\gamma_k^{\text{comb}} + \ln\gamma_k^{\text{disp}}. \tag{6.35}
$$

For the combinatorial activity coefficient we use the generalized Guggenheim-Staverman combinatorial activity coefficient derived previously by us [129]:

$$
\ln\gamma_k^{\text{comb}} = \ln\left(\frac{\phi_k}{x_k}\right) + \left(1 - \frac{\phi_k}{x_k}\right)\left[\frac{\ln(\phi_k/\theta_k)}{\frac{\phi_k}{\theta_k} - 1}\right], \tag{6.36}
$$

where ϕ_k and θ_k are respectively the volume and surface fraction of compound k in the mixture, which in the PC-SAFT and PHSC based activity models is defined by σ, while in the IPC model this is the segment diameter d (Eq. 6.24. Since the volume and area fractions in the perturbed chain model are equal the term within the square brackets of Eq. 6.36 becomes unity. This reduces Eq. 6.36 to the combinatorial term of the Flory-Huggins (FH) model. By combining Eqs. 6.36 with 6.16 and 6.26 one

obtains the total activity coefficient of the PC-SAFT based and the IPC model, respectively:

$$\ln\gamma_k^{\text{tot,PC-SAFT}} = \ln\left(\frac{\phi_k}{x_k}\right) + 1 + \frac{\phi_k}{x_k}$$

$$+ \frac{m_k Z_k}{2}\left(\frac{T_0}{T}\right)\left[\left(\frac{\sigma_k}{d_k}\right)^3 \tilde{\epsilon}_k - 2\frac{\sum_j x_j m_j \tilde{\epsilon}_{jk}\sigma_{jk}^3}{\overline{md^3}} + d_k^3 \frac{\overline{m^2\epsilon\sigma^3}}{\left(\overline{md^3}\right)^2}\right]$$

$$- \frac{m_k Z_\infty}{2}\left(\frac{T_0}{T}\right)\left(\frac{\tilde{a}_2}{\tilde{a}_1\overline{m}} + \frac{2\tilde{a}_3}{\tilde{a}_1\overline{m}^2}\right)\left(\frac{1}{m_k} - \frac{1}{\overline{m}}\right)\frac{\overline{m^2\epsilon\sigma^3}}{\overline{md^3}} \qquad (6.37)$$

$$\ln\gamma_k^{\text{tot,IPC}} = \ln\left(\frac{\hat{\phi}_k}{x_k}\right) + 1 + \frac{\hat{\phi}_k}{x_k}$$

$$+ \frac{m_k Z_k}{2}\left(\frac{T_0}{T}\right)\tilde{\epsilon}_k\left[1 - \sum_j\left[\hat{\phi}_j + \frac{Z_j}{Z_k}\left(\frac{d_k}{d_j}\right)^3\right]\frac{\tilde{\epsilon}_{jk}}{\tilde{\epsilon}_k} + \sum_{i,j}\frac{Z_i}{Z_k}\hat{\phi}_i\hat{\phi}_j\frac{\tilde{\epsilon}_{ij}}{\tilde{\epsilon}_k}\right] \qquad (6.38)$$

The equation for the PHSC dispersion activity coefficient model is obtained by setting $d_j = \sigma_j$ in Eq. 6.37.

6.2.5 The enthalpy of mixing

The excess enthalpy of component j is obtained by applying the relation:

$$h_j^{\text{E}} = -RT^2\frac{\partial \ln\gamma_j}{\partial T}, \qquad (6.39)$$

Subsequently, the enthalpy of mixing follows from:

$$H_{\text{mix}} = \sum_j x_j h_j^{\text{E}}. \qquad (6.40)$$

We used these expression to calculate the enthalpy of mixing. Because the PC-SAFT based as well as the IPC have a dispersion contribution, which is inversely proportional to temperature, the enthalpy of mixing is temperature independent. In order to make it temperature dependent the IPC model requires a temperature dependent Z_j. For instance one could opt for the equation:

$$Z_j = Z_0\frac{\rho(T)}{\rho(T_{\text{fus}})} - \frac{D}{m}, \qquad (6.41)$$

where Z_0 is scaled by the density of the liquid at system temperature, T, and melting point, T_{fus}. We will not investigate the temperature dependency of the enthalpy of mixing in this work.

6.3 Results

6.3.1 The total activity coefficient of the perturbed chain models

Fig. 6.4 displays the total activity coefficient model based on the PC-SAFT and PHSC models (Eq. 6.37) using for both models the original PC-SAFT compound parameters [78, 215] and using different size and energy combining rules to quantify the unlike segment and energy parameters. We note that there exist no compound parameters for the PHSC model, but since the temperature dependence of the segments size is weak, as we discussed at Eq. 6.6, the PC-SAFT parameters should be applicable for the PHSC model. The difference between the PC-SAFT and PHSC with IC combing rule are caused by the difference in the $\hat{a}_j$ parameters of Eq. 6.9. In the same figure the Flory-Huggins combinatorial activity model is plotted as reference. The difference between the perturbed chain activity models and the Flory-Huggins model is the effect of the dispersion contribution. We observe that the PHSC activity model with IC combining rules gives a better description of the activity coefficient than the PC-SAFT activity model with any of the combining rules, which is mainly due to the lower value of the parameter Z_∞. The results of the PHSC model lie close to those of the cubic equations of state, where theoretical result had to be reduced by a factor 3 to 4 [130]. It shows that the PHSC activity model has a better physical basis than the cubic equations of state, because no adjustment of the theoretical result is required. The results of the PC-SAFT based activity coefficient model with IC combining rule gives higher values than the PHSC based model. This was expected, as we indicated when we discussed Fig. 6.1. The WH and Kong combining rules lead to more deviation. This is an important observation, because it shows that small differences in the size and energy parameters of the compounds are amplified when using the WH and Kong combining rules. In other words, in order to obtain similar results with PC-SAFT with the WH or the Kong combining rules, the $\tilde{a}$ parameters have to be optimized in combination with the combining rules. We can not interchange the combining rules afterwards. To illustrate the effect of not doing so, we optimized the three model parameters, $\tilde{a}_1$, $\tilde{a}_2$ and $\tilde{a}_3$, using the consistent set of experimental activity coefficient data of Ashworth [6], who accurately measured and computed the activity coefficients of pentane, hexane, heptane and octane in squalane. The optimization of the $\tilde{a}_j$ parameters, while keeping the PC-SAFT compound parameters fixed, was carried out by least-squares
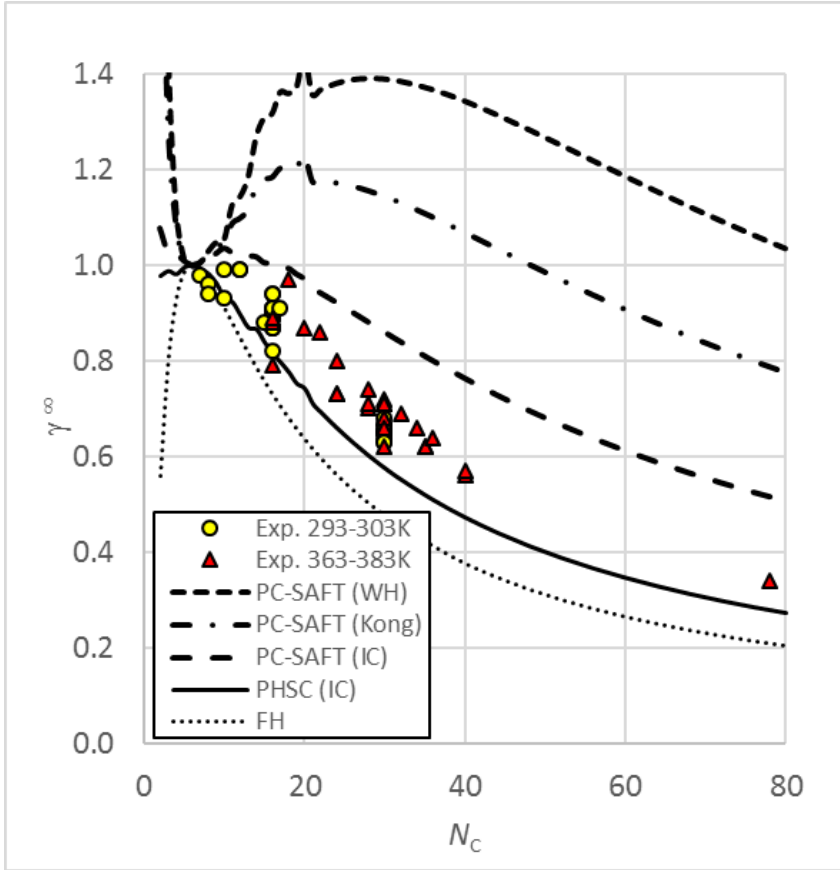
Figure 6.4: Prediction of the activity coefficients at infinite dilution of n-hexane dissolved in alkanes with different carbon numbers N_C. Experimental data at temperature range 293-303 K (yellow circles) and 363-383 K (red triangles) from [38]. Curves from bottom to top: FH combinatorial activity coefficient model (Eq. 6.36 dotted), PHSC based activity model, Eq. 6.37, with $d_j = \sigma_j$ and $Z_\infty = 11.88$ and IC combining rules (solid curve), and the PC-SAFT activity model, Eq. 6.37, with $Z_\infty = 19.10$ with IC (long dashed), Kong (dashed-dotted) and WH (short dashed) size and energy combining rules, respectively. The compound parameters m, σ, and ϵ are taken from [78, 215].

fitting. The optimized $\tilde{a}_j$ parameters with applied combining rules are given in the last 3 columns of Table 6.3. The results before and after fitting are displayed in Fig. 6.5. The differences between the 3 different combining rules are nearly imperceptible; the average absolute difference between the model and the experimental data is less than 1%.

The new parameters also give a good prediction for other alkane system as is
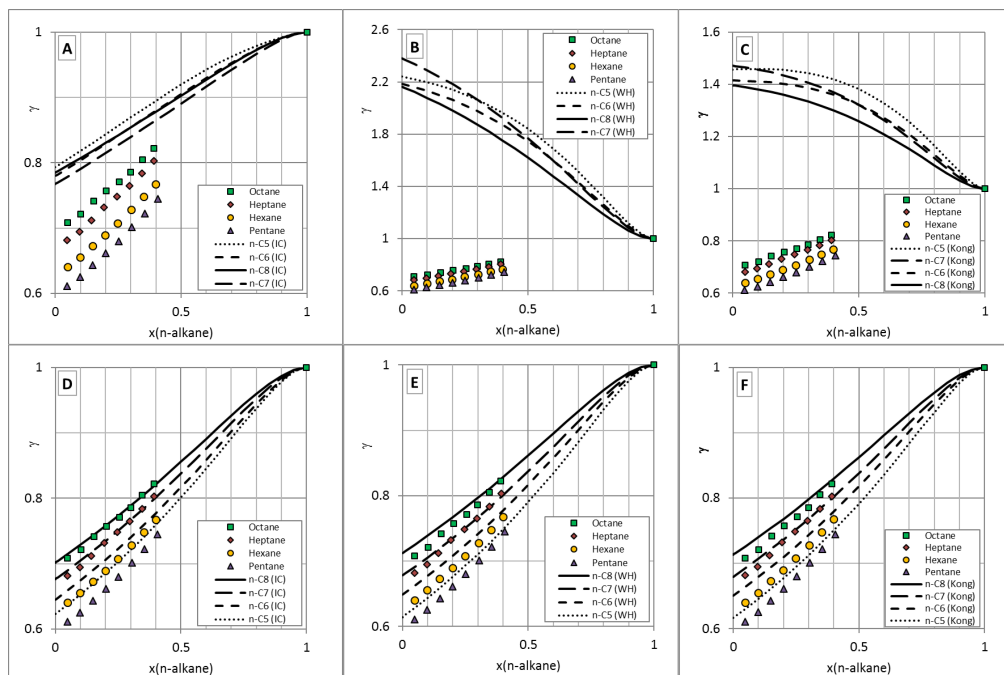
Figure 6.5: Composition dependency of the activity coefficients of n-pentane (triangles), n-hexane (circles), n-heptane (diamonds) and n-octane (squares) dissolved in squalane at 303 K from [6]. Results of Eq. 6.37 with IC, WH and Kong combining rules are depicted by curves for n-pentane (dotted), n-hexane (small dashed), n-heptane (large dashed), and n-octane (solid). Top row panels (A,B,C) and bottom row panels (D,E,F) are the results with the original and optimized PC-SAFT activity model parameters $\tilde{a}_j$ parameters of Table 6.3.

depicted in Fig. 6.6. Here we used for the calculation $T = 298$ K. Any influence of temperature on the curves is directly related to a reciprocal temperature dependence of the dispersion term, which moves the curves towards the Flory-Huggins model. When comparing the average absolute deviations (AAD) it follows that the most accurate prediction is obtained with the IC combining rules. On the other hand, the optimization of the PC-SAFT activity coefficient model in combination with the IC combining rules shows that there is still an inconsistency in the perturbation integral at LCP condition, because the Z-parameter of the PC-SAFT activity model in combination with the IC combining rule (see Table 6.3) is 4.8 instead of a value close to 11, and for single spheres is it even worse ($Z = 0$).

Another point to mention to Fig. 6.6 are the ripples on the curve for the regime $11 < N_{\rm C} < 20$. This is an artifact caused by the m, and σ parameters of the pure components. The density of a pure liquid is directly related to the hard-core volume

	Published $\tilde{a}_j$		This work	Optimized $\tilde{a}_j$ PC-SAFT		
Param.	PHSC	PC-SAFT	IPC	IC	WH	Kong
$\tilde{a}_1$	0.7778	1.2504	0.681	0.3163	0.0047	0.0048
$\tilde{a}_2$	0.1263	-0.6073	0.393	-0.452	-0.066	-0.067
$\tilde{a}_3$	-0.0062	0.3585	0	0.1370 [a]	0.2287	0.2312
$Z_j(m=1)$	13.7	15.3	14.4	0.0	2.6	2.6
$Z_j(m=\infty)$	11.9	19.1	10.4	4.8	0.07	0.07
AAD [%]	-	-	-	0.4%	0.8%	0.8%

[a] Note: Although $\tilde{a}_3$ should be zero, the used set of experimental data (C5-C8 in squalene) forces non-zero values. Either the PC-SAFT compound parameters are sub-optimal or the accuracy of the experimental data is not that accurate as assumed. A combination of both is possible as well.

Table 6.3: I_1 parameters of Eq. 6.11 at $\eta = 0.636$. Z_j calculated with Eq.6.14. AAD values for Fig. 6.5
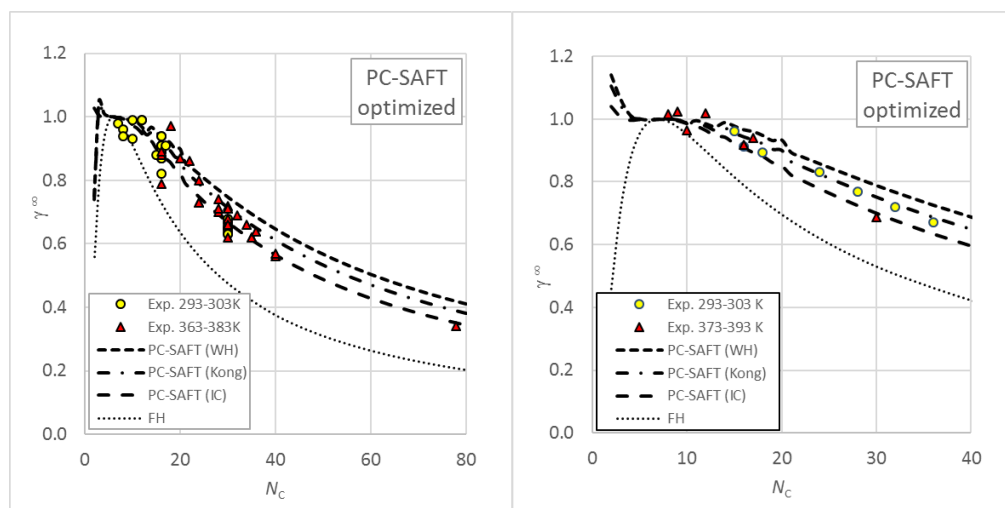
Figure 6.6: Activity coefficients of n-hexane (left panel) and n-heptane (right panel) in alkanes after optimization of the three PC-SAFT based activity model parameters $\tilde{a}_j$ (see Table 6.3). Experimental data from DDBST [38] at temperature ranges: 293 K - 303 K (yellow circles) and 373 K - 393 (red triangles). Curves from bottom to top: FH combinatorial activity model (dotted), and the PC-SAFT activity model (Eq. 6.37) with IC (long dashed), WH (dashed-dotted) and Kong (short dashed) combining rule, respectively, at 298 K. Compound parameters m, σ, and ϵ from [78, 215].

of the molecule. One can obtain the same density by increasing m and decreasing σ. In other words, the m and σ parameters are highly negative correlated. The

ripples are no longer present in the curve for $N_C > 20$, because in that regime the relations of Tihic *et al.* [215] for m, σ and ϵ/k_B were used. These relations smooth the deviations arising from parameter fitting. A strategy to avoid the occurrence of ripples is therefore to optimize the compounds parameters of the same class of compounds simultaneously, and using, for instance, for the alkanes a linear function that correlates the m parameter with the carbon number. The main message from the above results and observations is, that we need to revise the perturbation integral yielding an activity coefficient model and parameters that are physically meaningful. This is why we introduce a new model, which we will discuss in the next section.

6.3.2 The improved perturbed chain based activity coefficient model

As was explained above, the activity model derived from PC-SAFT gave disappointing results, which urged us to re-examine the perturbed chain model and to improve it. Here we show how well this improvement, the IPC model, worked out. The first step we made is that we set the parameter m equal to N_C. In the PC-SAFT EoS model the value of the m parameter is smaller than carbon number N_C. This is done to effectively reduce the flexibility of the chain to limit self-interaction. In the IPC model, the flexibility is controlled by parameter Z_j. In the Supplementary Material we show that this parameter can be used to predict the normal boiling points of alkanes ranching from methane to tetracosane with an accuracy of 3.3 K. This supports the idea that the Z_j parameter is a suitable measure to define the number of maximum possible intermolecular interactions per segment of molecule j. The second step in the development of the IPC model is the calculation of the segment diameter. This is done via the van der Waals volume, $V_{\text{vdW},j}$, which can be computed by Bondi's method [18] or retrieved from the DIPPR database [23]. From the van der Waals volume and the number of segments, in our case equal to the carbon number, $m_j = N_{C,j}$, we obtain the effective diameter σ_j:

$$\sigma_j = \left(\frac{6}{\pi} \frac{V_{\text{vdW},j}}{m_j} \right)^{\frac{1}{3}}. \tag{6.42}$$

The third step is the determination of the energy parameters ϵ_0 and ϵ_1 by fitting the total activity coefficient data of alkane mixtures with the IPC activity model and the three combining rules. From more than 1000 infinite dilution activity coefficient (IDAC) data points retrieved from the DDBST [38], 157 data points were excluded because a) the experimental temperature was above 400 K which gives values that differ too much from the γ-values at 298 K, b) the value was above 1, or c) the γ-value deviated too much from other sources. Simultaneously, we regressed the experimental finite activity coefficient data of alkanes dissolved in squalane [6] and

the IDACs of long n-alkanes in small alkanes, which were compiled by Kniaz [120]. Table 6.4 summarizes the obtained model parameters, the average absolute deviation of the models for the three types of data, and the overall accuracy of the fit expressed by the reduced χ-squared at 5%. The table shows that the three combining rules give

		combining rules		
Energy parameters of Eq. 6.28	No. data	IC	WH	Kong
ϵ_0		125.24	107.35	97.73
ϵ_1		12.69	9.28	7.40
AAD% C5-C8 in squalane [6]	32	1.4%	2.1%	1.5%
AAD% IDAC [38]	1188	4.2%	4.4%	4.4%
AAD% IDAC [120]	52	25%	25%	25%
χ^2_{red}	1262	2.85	2.91	2.95

Table 6.4: IPC model parameters and performance of combining rules.

nearly the same accuracy, though with different energy equation parameters. The average absolute deviation (AAD) of the IDAC of the IPC activity model with IC combining rules is 4.2%. This is at the level of the UNIFAC(Do) model [240] (3.6% for n-alkanes, 4.9% for saturated hydrocarbons). Fig. 6.7 depicts the correlation plot of the experimental and computed activity coefficients and the description of the alkane-squalane binaries for the IPC model (Eq. 6.38) with IC combining rules (Eqs. 6.33). Similar results are obtained with the Kong and WH combining rules (see Supplementary Material). In comparison to the results of Fig. 6.5 it follows that the description of the squalane systems by the IPC model with the IC combining rules is similar as those obtained by the PC-SAFT based activity model with IC combining rules. The main difference is the more tangible calculation method, yielding realistic values for Z_j.

Kniaz [120] calculated activity coefficients of long alkanes dissolved in small alkanes from solubility data. The correlation plot of these data and the IPC model results is depicted in the right panel of Fig. 6.7. We observe a larger scatter in results than obtained with small alkanes in long alkanes.

Figure 6.8 depicts the series of measurement of Madsen and Boistelle [156, 157]. We observe that the experimental results of dotriacontane lie outside expected range of octacosane and hexatriacontane, which are well described by the IPC model. There is no theoretical molecular explanation for the higher solubility and thereby lower activity coefficient of dotriacontane. In appendix O we take a closer look at the data of Madsen and Boistelle [156, 157] to understand the larger deviation of the IPC model from the parity line and explain possible causes for systematic errors.

Fig. 6.9 shows the limiting activity coefficients of hexane and heptane in alkanes
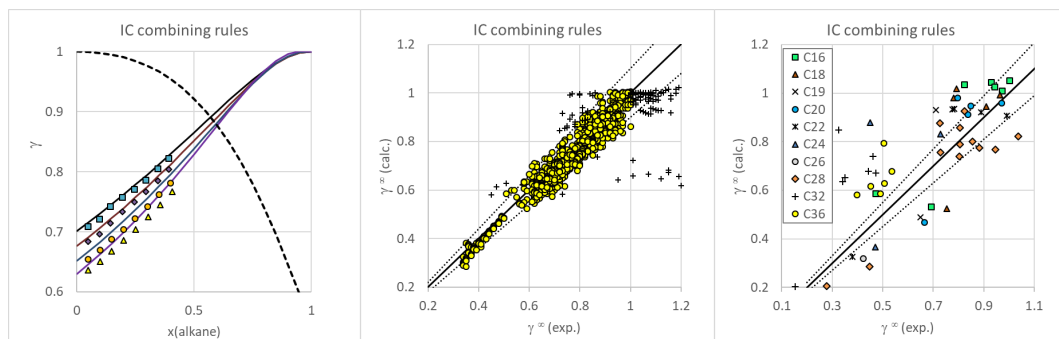
Figure 6.7: Left panel: Activity coefficients of n-pentane (triangles), n-hexane (circles), heptane (diamonds), n-octane (squares) in squalane at 303 K from Ashworth [6]. Solid curves depict IPC model Eq. 6.38 with IC energy and size combing rules for the alkanes dissolved in squalane. Dashed curve depicts IPC model for squalane in n-octane. Middle/Right panel: Parity plot of IDAC of binary alkane mixtures between experimental data and IPC model with IC combining rules, where dashed lines indicate the 10% error level from the y=x solid line. Middle panel: Yellow dots and crosses represent, respectively, included and excluded experimental IDAC data from DDBST [38] used for fitting of the energy parameters ϵ_0 and ϵ_1. Right panel: Experimental IDAC data compiled by Kniaz [120] for long n-alkanes in small alkane solvents.

and the results obtained with the IPC activity model with Kong combining rules. The results are similar as obtained with the PC-SAFT based activity model (Fig. 6.6).
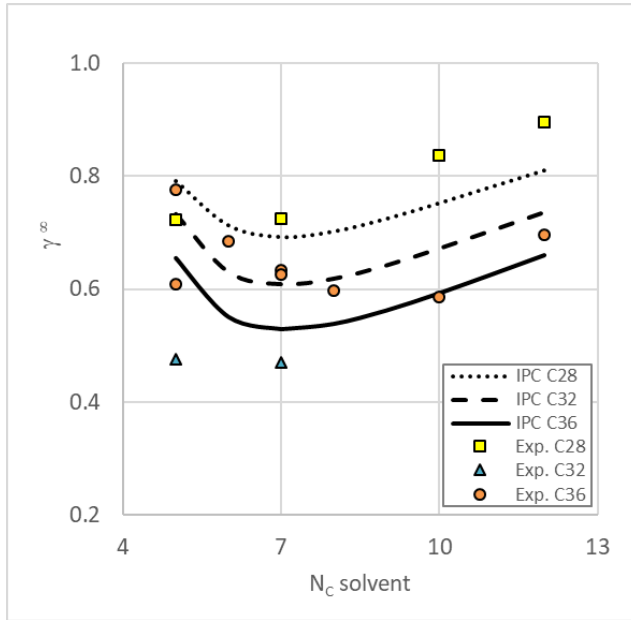
Figure 6.8: IDAC of long n-alkanes (see legenda) as function of the carbon number N_C of the n-alkane solvent. Data points are the values at 300 K of table O.2. Curves are the results of the IPC activity coefficient model (Eq. 6.38) at 300 K.
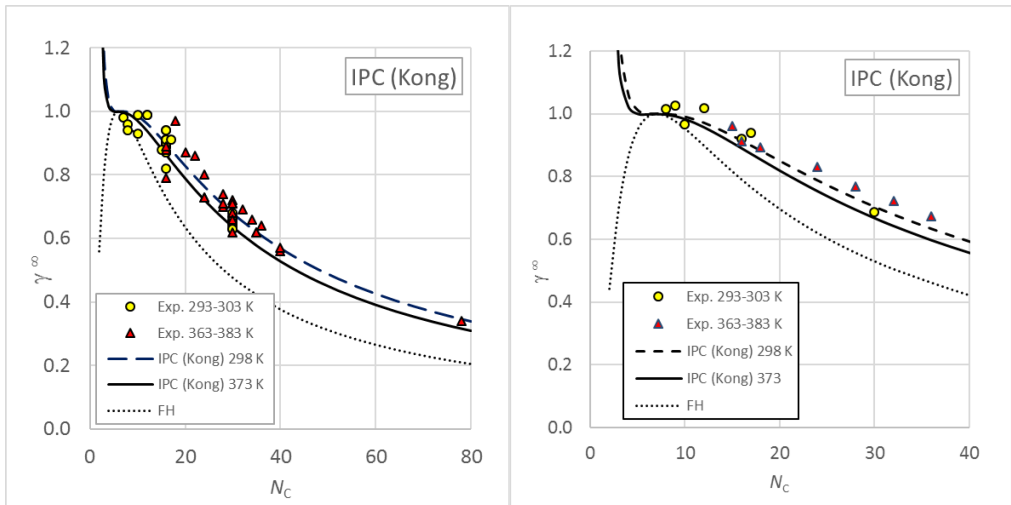
Figure 6.9: Limiting activity coefficients of n-hexane (left panel) and n-heptane (right panel) in alkanes. Experimental data as in Fig. 6.6. Curves are the FH combinatorial activity model (dotted), and the IPC activity model Eq. 6.38 with Kong size and energy combining rules at 298 K (solid) and 373 K (dashed).

6.3.3 Enthalpy of mixing

The PC-SAFT activity model (Eq. 6.16) with the original perturbation parameters predicts too high values for the enthalpy of mixing for the binary systems n-hexane - n-dodecane and n-hexane - n-hexadecane as can be observed by the small dashed curves in left and right panel of Fig. 6.10. With the optimized model parameters
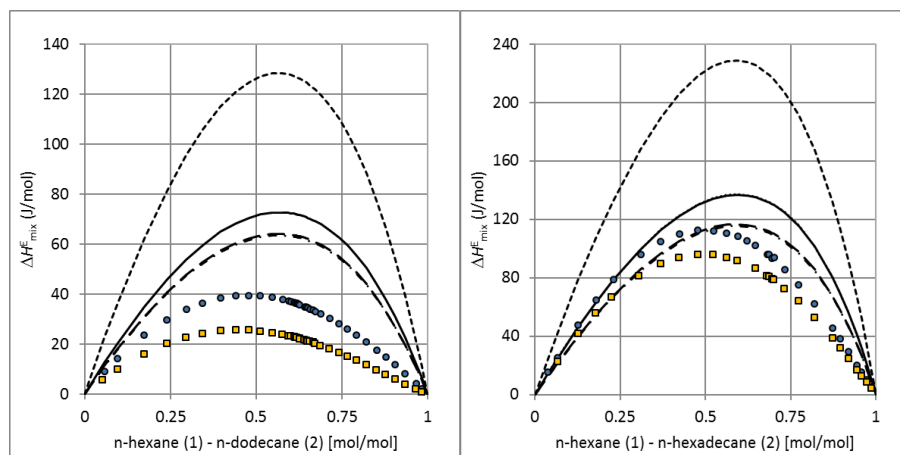
Figure 6.10: Enthalpy of mixing of the binary systems n-hexane - n-dodecane (left panel) and of n-hexane - n-hexadecane (right panel) as function of n-hexane concentration. Experimental data of n-hexane - n-dodecane at 298 K (circles) and 308 K (triangles) from [174], and of n-hexane - n-hexadecane at 298 (circles) and 303 K (triangles) from [167]. Small dashed curves are the results of Eqs. 6.39, 6.40 and 6.16 with IC combining rules and the original $\tilde{a}_j$ parameters at 298 K. Solid curves, and the overlapping large dashed and dashed-dotted curves are the results with optimized $\tilde{a}_j$ parameters for the IC, WH and Kong combining rules, respectively, at 298 K.

(Table 6.3) a good prediction of the experimental results is obtained. The WH and Kong combining rules yield curves that overlap. These curves lie closer to the experimental data, than those obtained with the IC combining rules. The model with optimized parameters shows correctly the trend of increasing mixing enthalpy with increasing difference in molecular size, but a temperature dependency is hardly discernible. The maximum difference between the results of 298 and 303 K for the binary system of n-hexane and n-dodecane is only 0.3 J/mol. This indicates that the temperature-dependency of the diameter has little influence on the enthalpy of mixing.

In Fig. 6.11 the results of the IPC model (Eq. 6.26) with the IC, WH and Kong combining rules are compared to the same experimental data as shown in Fig. 6.10. The IPC activity model yields results, which are of similar quality as those of the PC-SAFT activity model with optimized $\tilde{a}_j$ parameters. Here the IPC activity model
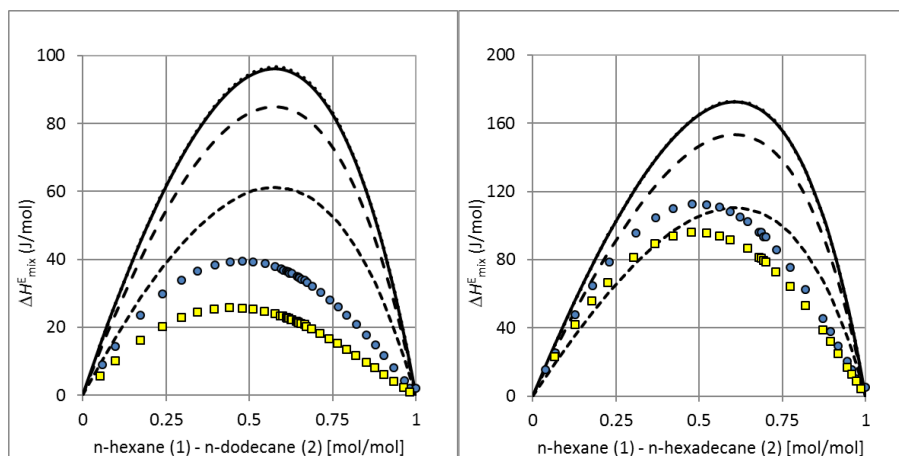
Figure 6.11: Enthalpy of mixing of the binary systems n-hexane - n-dodecane (left panel) and of n-hexane - n-hexadecane (right panel). Experimental data as in Fig. 6.10. Curves are the results of the IPC model (Eqs. 6.39, 6.40 and 6.26) with the WH (small dashed curve), the Kong (large dashed curve) and the IC (solid curve) combining rules at 298 K. The dotted curve on top of the solid curve is the IPC model with IC combining rules at 303 K.

with the Kong combining rules provides the best results. The IPC activity model with IC combining rules gives less accurate results than the PC-SAFT activity model with optimized $\tilde{a}_j$ parameters. We expect that the curves can be further improved by introducing a temperature dependence of the Z_j parameter as proposed in Eq. 6.41.

6.4 Concluding remarks

In this paper dispersion activity models were derived from perturbed chain equations of state by using the limit of infinite pressure. In this limit molecules are set into close contact. At this point the second perturbation integral vanishes, while the first perturbation integral simplifies to a function of molecular chain lengths only.

It was shown that the original set of perturbation integral parameters of the PC-SAFT equation of state is not applicable in the derivation of a predictive dispersion activity coefficient model, due to the impossibility to extrapolated the first perturbation integral beyond a packing fraction $\eta = 0.46$. Optimization of the $\tilde{a}_j$ model parameters, while keeping the compound parameters fixed, for this close packed liquid condition was possible and yielded a good activity coefficient description. However, the resulting average coordination numbers, Z_j, are physically impossible for single spheres and infinite long chains. This implies that the PC-SAFT compound

parameters of the EoS are not applicable for an activity coefficient model.

Therefore, we introduced another method to quantify the first perturbation integral. In here the effective coordination number of a molecule is quantified by means of a topological descriptor. For branched and linear chains this can be done by the first Zagreb index, which requires the carbon structure of a molecule. For cyclic structures and hyperbranched structures, in which proximity effects exclude more intramolecular configurations, the first Zagreb index was adjusted to a new topology number called the dispersion index, D. For the quantification of the dispersion energy a hydrogen based topology index was introduced, which is a simplification based on the London dispersion theory. The derived dispersion activity coefficient model with these descriptors gives in combination with the Flory-Huggins combinatorial activity coefficient model excellent results for the prediction of activity coefficients of alkane mixtures. The average absolute deviation of this new model is at the level of the UNIFAC(Do) model, which is regarded as the best today.

In the limit of molecules having the same coordination number, which happens for long linear chains, the improved perturbed chain model yields a similar expression as the van Laar type of activity models. This similarity can be understood from the fact that a mixture of molecules having equal coordination number is similar to a cubic equation of state, where all molecules are equal sized spheres, which have identical coordination number. In other words, in systems where molecules have the same shape and proximity effect the van Laar type of activity coefficient model can be used to predict phase equilibria. In general branching reduces the number of intermolecular interaction and thereby changes the van Laar type of modelling to a model that quantifies differences in shape and interaction. This is done in the proposed IPC model.

The evaluation of three different combining rules to quantify unlike interaction energies and center-to-center distances, which are needed in the calculation of activity coefficients, shows that nearly the same accuracy in activity coefficients is obtained, though with different model energy parameters. Therefore, the classical combining rules of Lorentz and Berthelot seem to be sufficient in the prediction of the dispersion activity coefficient contribution.

The enthalpy of mixing calculated by the proposed dispersion activity coefficient models gives qualitatively good results. It yields a correct dependency in chain length. Binary systems with a larger difference in chain length show a larger enthalpy of mixing. But the models lack a correct temperature dependency description. We attribute this deficiency to the absence of a term that accounts for the reduction in the number of interacting spheres as a result of volume expansion. A way to improve this could be by scaling the maximum number of interacting spheres with the liquid density, as proposed by Eq. 6.41.

In conclusion, the dispersion activity coefficient model as introduced in this paper is a new way to improve the existing activity coefficient models, where an explicit term for dispersion interaction has been neglected, and offers the possibility to improve existing predictive models, such as UNIFAC(Do) and COSMO-RS, in a more robust way.

Chapter 7

Dispersion ACMs.
Part 3: Group Contribution
Methodology [1]

7.1 Introduction

We have shown in the previous chapter how the dispersion contribution of an activity model can be derived from the PC-SAFT equation of state, developed by Gross and Sadowski [78]. In combination with an off-lattice combinatorial contribution [129] this yields the total activity coefficient model for hydrocarbon mixtures. This model, which we called the improved perturbed chain (IPC) activity model, can be extended to systems containing polar and hydrogen bond forming molecules by using the energetic contributions derived from the original UNIFAC [59], COSMO-RS [116] or COSMOSAC [145] model. The benefit of the IPC activity model is that only the topology of the molecule is required [130].

In this work we will refine the IPC acitivity model by introducing a group contribution method which includes the topology of the molecule. This information is stored by means of a local topology index. A further improvement is made by using Boltzmann weights to account for differences in submolecular interactions by using the concept of the COSMOSPACE [117] and COSMO-RS [116] models. This implies that the interaction energy of a molecule is not averaged by the local mole fraction, but by the local activity.

---

[1]To be submitted for publication as "Dispersion activity coefficient models. Part 3: A structural preserving group contribution method.", by Gerard J.P. Krooshof, Remco Tuinier, and Gijsbertus de With.

First we summarize the IPC activity dispersion model and point out the key features. Subsequently, we will derive the group contribution method for dispersion contribution in the activity coefficients of alkanes. The obtained GC model is compared to experimental data and compared to earlier models. Finally, we discuss the obtained results.

7.2 Theory

7.2.1 The dispersion activity coefficient model

The dispersion contribution can be calculated by the improved perturbed chain (IPC) activity coefficient model [130]. This model gives for component k, which consists of m_k segments having an average diameter d_k, the activity coefficient:

$$\ln \gamma_k^{\mathrm{disp}} = m_k \frac{Z_k}{2} \tilde{\epsilon}_k \left(\frac{T_0}{T} \right) \left[1 - \sum_j \left[1 + \frac{Z_j}{Z_k} \left(\frac{d_k}{d_j} \right)^3 \right] \hat{\phi}_j \frac{\tilde{\epsilon}_{jk}}{\tilde{\epsilon}_k} + \sum_{i,j} \frac{Z_i}{Z_k} \hat{\phi}_i \hat{\phi}_j \frac{\tilde{\epsilon}_{ij}}{\tilde{\epsilon}_k} \right],$$
(7.1)

where the hard-core volume fraction $\hat{\phi}_k$ is calculated with the mole fraction x_k:

$$\hat{\phi}_k = \frac{x_k m_k d_k^3}{\sum_j x_j m_j d_j^3}.$$
(7.2)

The quantity Z_k defines the average number of interacting spheres per molecule segment:

$$Z_k = Z_{\max} - \frac{D_k}{m_k},$$
(7.3)

where $Z_{\max} = 14.4$ is the maximum number of spheres around a sphere at random close-packed (RCP) density. This parameter could be made temperature dependent to account for expansion of the liquid. We will mot explore this in this work. D_k is the dispersion number of molecule k, which is temperature independent. For linear and branched alkanes this topology index is equal to the Zagreb topology index [83] and is given by the equation:

$$D = \sum_{j=1}^{4} m_{v,j} \times j^2,$$
(7.4)

where the $m_{v,j}$ denotes the number of vertices, in our case non-hydrogen atoms, that have j covalent bonds with other non-hydrogen atoms. In the case of alkanes, for

instance, the 3-methyl-pentane molecule has 3 vertices (-CH$_3$) with 1 non-hydrogen bond, 1 vertex (>CH$_2$) with 2 non-hydrogen bonds, and 1 vertex (>CH-) with 3 non-hydrogen bonds. In this case Eq. 7.4 reads: $D = 3 \times 1^2 + 1 \times 2^2 + 1 \times 3^2 = 16$. Subsequently, Eq. 7.3 gives the average number of segments that interact with a segment of 3-methyl-pentane at random close-packed condition: $Z_k = 14.4 - 16/(6) = 11.74$.

The quantity $\tilde{\epsilon}_k$ is the reduced interaction energy of pure compound k at reference temperature, T_0, and given by:

$$\tilde{\epsilon}_{\mathrm{k}} = \frac{\epsilon_{\mathrm{k}}}{k_{\mathrm{B}} T_0}, \tag{7.5}$$

where k_{B} is the Boltzmann constant, and ϵ_{k} the dispersion energy of molecule k, which is calculated by the equation:

$$\epsilon_k = \frac{\epsilon_0 + \epsilon_1 J_{\mathrm{QH}}}{m_k}, \tag{7.6}$$

where ϵ_0 are ϵ_1 are constants, which were derived by fitting experimental data and given in table 6.4. The parameter J_{QH} is a constitutional index, called the quadratic hydrogen number, and is defined by the number of hydrogen atoms attached to carbons in molecule k.

$$J_{\mathrm{QH}} = 3^2 N_{\mathrm{CH3}} + 2^2 N_{\mathrm{CH2}} + 1^2 N_{\mathrm{CH}} + 0^2 N_{\mathrm{C}}. \tag{7.7}$$

In here the parameter N_{CH3}, N_{CH2}, N_{CH}, N_{C} is the number of carbons that have 3, 2, 1 and 0 hydrogen atoms, respectively. For instance, the compound 2,2,4-trimethylpentane, which has 5 -CH$_3$, 1 >CH$_2$, 1 -CH< and 1 >C< group. This gives $D = 5 + 4 + 9 + 16 = 34$ and $J_{\mathrm{QH}} = 9 \times 5 + 4 \times 1 + 1 \times 1 = 50$. The calculation of the unlike interaction energy in Eq. 7.1 requires a mixing rule, which has been discussed in detail elsewhere [130].

The main feature of the IPC model is the parameter Q_{j}, which contains topological information of the molecule. One of the weak point in the PC-SAFT model is the calculation of the average interaction energy. It is an average, treating molecules as randomly distributed. As has been pointed out by Guggenheim [81] interaction creates non-randomness. His quasi-chemical approach, in which molecules make a quasi-bond between contacts, leads to explicit equations for binaries of molecules, which have one or two types of contacts. In COSMOSPACE [117] equivalent equations as those of Guggenheim are obtained, but also formula's that can handle molecules having more than two different contact types. This formulation will be used in this chapter.

7.2.2 Improvement towards a group contribution method

The IPC model requires as input for the dispersion contribution of the activity coefficient model only the molecular formula of the alkanes in the mixture. From this information the topology indices D and the constitutional indices J_{HQ} can be calculated, and subsequently Q_k and ϵ_k. For the development of a group contribution method we have to define molecular groups. In the next sections we show how this is done. To address the weakness of the PC-SAFT model, where the mole average instead of a Boltzmann weighted average is used for the calculation of the average interaction enrgy, we apply the concept of COSMOSPACE.

Definition of group increments

From the molecular model we learned that a key parameter is the dispersion number, defined by Eq. 7.4. It defines the average number of interacting segments around a segment of molecule by Eq. 7.3. To transpose this into a group contribution model, we define main groups, which are characterized by the same dispersion energy, and subgroups which are defined by the connectivity. The topology index D_d of the molecule is now reduced to a summation of group increments. Each subgroup k is defined by the connectivity with its neighbors:

$$D(m_{v,k}) = \sum_{i=1}^{k} j(i), \tag{7.8}$$

where $m_{v,k}$ defines the number of vertices of a segment with k covalent bonds to non-hydrogen atoms, and $j(i)$ denotes the type of the subgroup to which is bonded to. For example a -CH$_3$ group makes only one covalent bond with another methyl group. This case reduces Eq. 7.8 to just one value depending to which type of methyl group it is bonded. When a -CH$_3$ group is bonded to a -CH$_3$, to >CH$_2$,-CH , or >C< group, the value of D$_d$ is 1, 2, 3 and 4, respectively. A >CH$_2$ group has two covalent bonds with non-hydrogen atoms, $k = 2$. Therefore the summation is over two types of methyl groups. This yields values ranging from 2 (bonded to 2 CH$_3$) to 8 (bonded to 2 >C<). In appendix P a full table of 68 possibilities for branched alkanes and the D increments are given. Table 7.1 shows the groups in n-alkanes. We see that at the chain ends the number of interactions Z is higher than in the chain middle due to the presence of neighboring carbons. Tables 7.1 gives for the entire molecule nonane $D = 30$ and $Z = 99.6$. An example of a branched alkane is given in Table 7.2. We note that the value for D of a molecule can be calculated by summing up the contributions of the groups, as was done in the tables above, or directly by Eq. 7.4 or Eq. 7.8. In the case of 2,2,4-trimethylhexane Eq.7.4 gives $D = 5 \times 1^2 + 2 \times 2^2 + 1 \times 3^2 + 1 \times 4^2 = 38$.

Position	1	2	$3 \ldots N_C$-2	N_C-1	N_C	
Group	CH_3	CH_2	CH_2	CH_2	CH_3	Sum
D_j	2	3	4	3	2	$4N_C$-6
$Z_j = Z_{max} - D_j$	12.4	11.4	10.4	11.4	12.4	$6+10.4N_C$

Table 7.1: Example: Increments of D for an n-alkane with $N_C > 4$ and the calculation of Z_j when $Z_{max} = 14.4$.

Position	1	2	3	4	5	6	7	8	9	
Group	CH_3	C	CH_2	CH	CH_2	CH_3	CH_3	CH_3	CH_3	Sum
D_j	4	5	7	5	4	2	4	4	3	38
$Z_j = Z_{max} - D_j$	10.4	9.4	7.4	9.4	10.4	12.4	10.4	10.4	11.4	91.6

Table 7.2: Example: Increments of D for 2,2,4-trimethylhexane and the calculation of Z_j when $Z_{max} = 14.4$.

Eq. 7.8 plays an important role in the calculation of interacting spheres per main group, which denotes a type of an induced dipole. Therefore, all subgroups, which belong to the same main group, have the same dispersion energy. The Q values of these subgroups can be summed to quantify the total amount of interactions that will give the same interaction energy with another main group. Table 7.3 gives some examples of this procedure.

Main group	n-alkane $(N_C > 3)$	n-Nonane	2,2,4-trimethylhexane
$Z_{k3} = \sum Z_{-CH3}$	$24.8 = 4 + 10.4 \times 2$	24.8	55.0
$Z_{k2} = \sum Z_{>CH2}$	$2 + 10.4 \, (N_C - 2)$	74.8	17.8
$Z_{k1} = \sum Z_{>CH-}$	0	0	9.4
$Z_{k0} = \sum Z_{>C<}$	0	0	9.4
Z_{tot}	$6 + 10.4N_C$	99.6	91.6
$Z_{tot}/N_C = Z_k(Eq.7.3)$	$10.4 + 6/N_C$	11.06	10.18

Table 7.3: Dispersion profile of n-alkanes and 2 nonane isomers.

The table shows that for nonane the two CH_3 groups and the seven CH_2 groups interact in total with 24.8 and 74.8 methylene groups from the surrounding, respectively. By dividing these number with the number of groups, we get the average value per group: 12.40 and 10.68, respectively. In general for n-alkanes with $N_C \geq 3$, we have always 12.4 for the CH_3 groups and $10.4 + 2/(N_C - 2)$ for the CH_2 group. The last results shows that the interaction with CH_2 is dependent from chain length of the molecule. It is a topology effect, for which the UNIFAC and COSMO-RS family of models do not account. It plays a role for relatively short chains, since long chains

have a contribution of 10.4 neighbors per CH_2 group.

Definition of interaction energy

In the COSMOSPACE and COSMO-RS models interactions are computed on the basis of pairwise interacting surface areas. In these models a molecule is divided in different surface types and the local interaction energy is dependent from the charge density between surface patches. We will use this idea for the induced dipole interaction between a pair of methylene groups. Further, instead of segment surface fractions, we take Z_j, to quantify the fraction of induced dipole interactions. This fractions is defined by the expression:

$$\psi_j = \frac{\sum_k x_k Z_{k,j}}{\sum_{k,j} x_k Z_{k,j}}, \tag{7.9}$$

where x_k is the mole fraction of compound k in the mixture and $Z_{k,j}$ is the number of alkyl groups of type j in molecule k. We will use subscripts $j = 3, 2, 1$ and 0 as short-cut to denote the CH_3, CH_2, CH, and C group, respectively.

Let us take as example pure n-nonane. From Table 7.3 it follows that $\psi_3 = \frac{24.8}{99.6} = 0.249$ and $\psi_2 = \frac{74.8}{98.6} = 0.751$, while $\psi_1 = \psi_0 = 0$. A 50:50 mixture of n-nonane and 2,2,4-trimethylhexane has $\psi_3 = \frac{39.90}{95.6} = 0.417$, $\psi_2 = \frac{46.30}{95.6} = 0.484$, and $\psi_1 = \psi_0 = \frac{4.70}{95.6} = 0.049$. The activity coefficient of an methylene group in a mixture, $\hat{\gamma}_j$, is found by using the concept of the COSMOSPACE model [117]. With the group fractions, ψ_k the group activity coefficients are defined by the expression:

$$\frac{1}{\hat{\gamma}_j} = \sum_k \psi_k \hat{\gamma}_k \tau_{kj}, \tag{7.10}$$

where j and k run from 0 to 3 to denote the four alkyl groups by their hydogen number. The model can easily be extended to other groups that interact by dispersion, but for this work we limit ourselves to linear and branched alkanes, to show that this concept works. The above set of four equations is solved iteratively starting with $\gamma_k = 1$ at the right side of Eq. 7.10. The parameter, τ_{ij}, is the coefficient of interaction between two alkyl groups and is defined by:

$$\tau_{ij} = \exp\left[-\frac{\epsilon_{ij} - (\epsilon_{ii} + \epsilon_{jj})/2}{k_B T}\right] = \exp\left[-\frac{\Delta \epsilon_{ij}}{k_B T}\right], \tag{7.11}$$

where ϵ_{ij} is the dispersion energy between an alkyl pair, which is computed by the dispersion energy of the groups. In appendix Q we show that in the approximation

that all carbons have the same induced dipole strength, that Eq. 7.11 reduces to:

$$\tau_{ij} = \exp\left[-\frac{(i-j)^2 \epsilon_{HH}}{2k_BT}\right], \tag{7.12}$$

where i and j denote the number of hydrogen atoms in the segments that interact, and ϵ_{HH} is the induced dipole interaction energy between two hydrogens of different alkyl groups. This implies that the energy differences of unlike pairs are correlated by the ratio:

$$(\Delta\epsilon_{1,2} : \Delta\epsilon_{2,3} : \Delta\epsilon_{3,4}) : (\Delta\epsilon_{1,3} : \Delta\epsilon_{2,4}) : \Delta\epsilon_{1,4} = (1:1:1):(4:4):9 \tag{7.13}$$

We can expect that, due the bonding between the carbon atoms, the induced dipole energy is shifted. We optimize the $\Delta\epsilon_{ij}$ of Eq. 7.11 to the experimental infinite dilution activity coefficient data by assuming that there is not such a correlation between these energy differences.

In chapter 6 on the derivation of an improved perturbed chain activity model, we demonstrated that topological information improves the description of non-ideal behavior of alkane mixtures, and with the work on the nearest neighbor theory [129] we showed that the molecular activity of a compound k is calculated from surface activity coefficients by using the number of nearest neighbors, Z. Therefore, with the Z_{kj} of the alkyl groups in this work, calculated by the D parameters in P, we get:

$$\ln\left(\gamma_k^{disp}\right) = \sum_j \frac{Z_{kj}}{2} \ln\left(\frac{\hat{\gamma}_j^x}{\hat{\gamma}_j^p}\right) \tag{7.14}$$

where $\hat{\gamma}_j^x$ and $\hat{\gamma}_j^p$ denote the activity coefficient of group j in the mixture and in the pure compound, respectively, which are calculated by Eq. 7.10. An important difference with the COSMOSPACE method is that the Q_{kj} in this work contains topological information of the molecule, while in COSMOSAC [145] and COSMO-RS [116] this information is lost by the computation of sigma-profiles. In the next section the dispersion energy between two hydrogen atoms of different segments will be obtained by fitting the limiting activity coefficients of linear and branched alkanes.

7.3   Results

7.3.1   Activity coefficients

Fig. 7.1 depicts the results obtained with the GC-IPC model, which is the combination of Eq. 7.14 and Eq. 3.36. We observe a good correlation between experiment and model for the limiting activity coefficients of binary hydrocarbon mixtures with equation and for the activity coefficients of pentane to octane in squalane. In fact, it
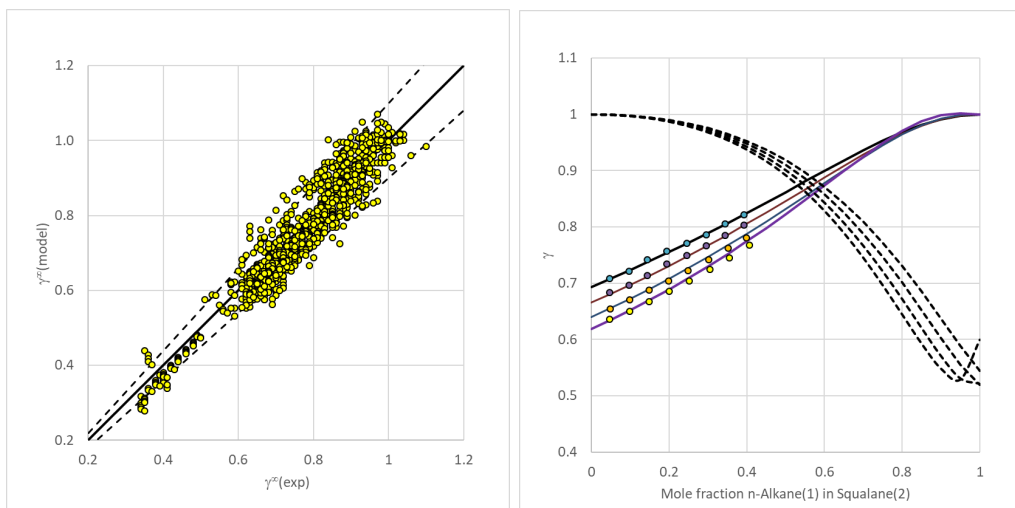
Figure 7.1: Left panel: Parity plot between experimental data and GC-IPC model data for the limiting activity coefficients of binary alkane mixtures. Right panel: Activity coefficients of alkanes in squalane as function of alkane mole fraction. Symbols and solid curves as in Fig. 5.8. Dashed curves represent the activity coefficient of squalane dissolved in octane (top), heptane, hexane and pentane (bottom), respectively.

is slightly better than the molecular approach. The average error for the activity at infinite dilution is for the GC model 4.4%, while it is 4.5% for the molecular model with Kong mixing rules. The average error for the activity coefficients of alkanes dissolved in squalane $<0.1\%$, while it is 0.8% for the molecular model. Fig. 7.2 shows the activity coefficient at infinite dilution of n-hexane and n-heptane in alkanes as function of carbon number. We see that model and experiment agree well. The level of accuracy is the same as the molecular model. Table 7.4 shows at the left side the absolute interaction energies, $\Delta\epsilon_{ij}$, and on the right the values relative to the interaction energy between a CH_3 group and a CH_2 group, where we have denoted the expected theoretical values within parentheses.
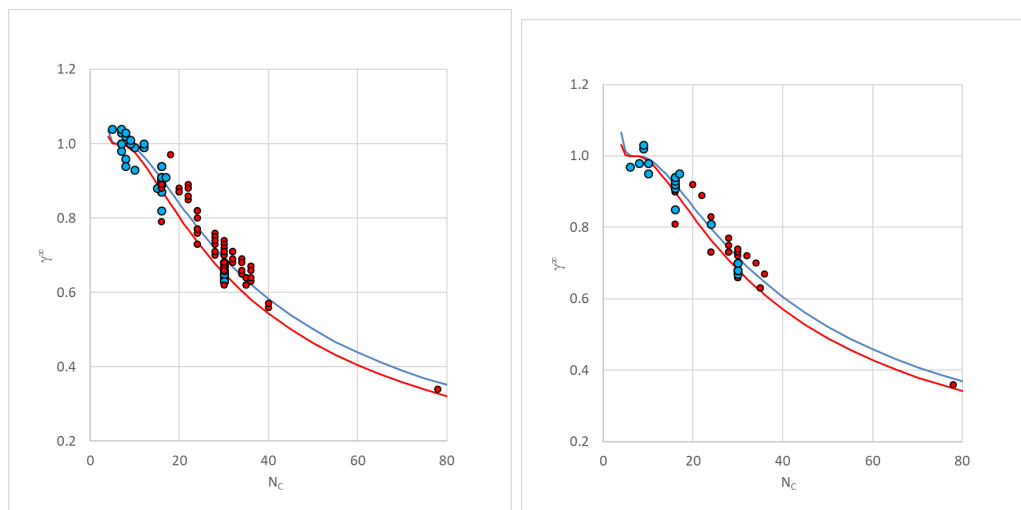
Figure 7.2: Limiting activity coefficients of hexane (left panel) and heptane (right panel) in alkanes. Blue and red symbols are experimental data at low and high temperature collected from DDBST [38]. Red and blue curves are the results of GC-IPC model at 313 and 373 K.

$\Delta\epsilon_{ij}[K]$	-CH$_3$	-CH$_2-$	-CH<	>C<
-CH$_3$	0	42	177	294
-CH$_2-$	42	0	36	71
-CH<	(168)	(42)	0	0
>C<	(378)	(168)	(42)	0

$\Delta\epsilon_{ij}/\Delta\epsilon_{12}$	-CH$_3$	-CH$_2-$	-CH<	>C<
-CH$_3$	0	1	4.24	7.07
-CH$_2-$	(1)	0	0.87	1.69
-CH<	(4)	(1)	0	0
>C<	(9)	(4)	(1)	0

Table 7.4: Absolute and relative dispersion interaction energies.

We observe that the regression gives for the relative interaction of CH$_3$ with the other alkyl groups the ratios of 1:4.24:7.07, which roughly follows the tendency of the expected ratio of 1:4:9. The difference indicates that the induced dipole of each alkyl group does not follow the quadratic relation with hydrogen number, which indicates that Eq. 7.12 can not be applied or requires a modification involving the effect of carbon-carbon and carbon-hydrogen interaction.

7.3.2 Excess enthalpy

The prediction of the excess enthalpy, calculated from Eq. 6.26 and 6.39, is shown in Fig. 7.3. It is qualitatively correct, and comparable to the results depicted in Fig. 5.9 and 6.10, when the classical combining rules of Lorentz and Berthelot are used. Since Kong's model gives better results, it might be that Eq. 7.12 requires a mixing rule for the interaction energy. Comparing these results with the results obtained
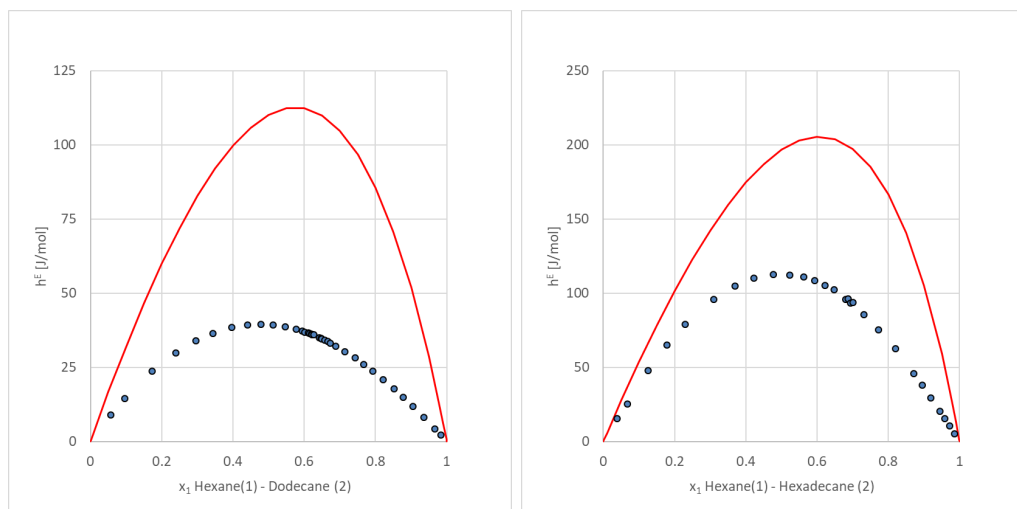
Figure 7.3: Enthalpy of mixing of the binary systems n-hexane - n-dodecane (left panel) and of n-hexane - n-hexadecane (right panel). Experimental data measured at $T = 298$ K from [174] and [167]. Curves are the results of Eq. 6.26 and 6.39 at $T = 298$ K.

with Eq. 5.31 and Eq. 6.26 with 6.39, as depicted in Fig. 5.9 and 6.10, respectively, show similar deviation from experimental results when the classical mixing rules of Lorentz and Berthelot are used.

7.4 Discussion and Conclusions

In this chapter we have shown that the IPC model for the calculation of the dispersion activity coefficient contribution can be converted into a group-contribution method. Key is the calculation of $Z_{k,j}$, which is the number of interacting spheres of a group j in molecule k. The interaction energies for these groups obtained by regressing experimental infinite dilution activity coefficient data follow roughly a quadratic relation with the difference in hydrogen atoms between two groups. The origin of the difference between the regressed and the expected values for the interaction energies (Table 7.4) is not yet clear. The assumption that the carbon atoms in a chain are not spherical but rather ellipsoids [11], might change the ratio in area and surface fraction of compounds, but is expected not to exceed a change of 5% in the combinatorial activity contribution, as was discussed in chapters 3 and 4. For that reason shape effects are regarded as negligible. On the other hand, the work of Schneider *et al.* [199] could be important. They showed that steric crowding of carbons, such as is present in molecules that contain, for instance, tertiary butyl

side groups, stretches the adjacent C-C bond, but does not lead to a lower stability of this covalent bond, because the crowded side groups give a compensating force that stabilizes the stretched bond. This finding could imply that the intramolecular bonding between adjacent methyl groups makes these groups less probable available for intermolecular bonding. This hypothesis could be explored with the current model set-up by introducing a second order effect. In practice this implies that the dispersion index, D, of methyl and methylene side groups, which are attached to neighboring $>$C$<$ or $>$C- groups, should be higher than the covalent connectivity of the molecule indicates. Effectively, the number of interactions has to be reduced. The current work neglects this effect and therefore, in order to compensate the too high number of interacting spheres, the regression of the activity coefficient data possibly yields a lower value for the interaction energy parameter than is expected from theory.

Chapter 8

COSMO-3D[1]

8.1 Introduction

In the design of chemical and pharmaceutical processes, the calculation of activity coefficients is essential to know the volatility or solubility of compounds [180, 194, 9, 60, 230]. In the past, many methods have been developed for this purpose such as group contribution methods (GCM) [146, 147, 142, 41, 34, 35, 91, 161, 243], or quantitative structure property relationships (QSPR) [70, 108, 183, 187, 207]. One of the most successful approaches is based on the UNIFAC theory [178, 112, 133, 231, 21]. Quantum-mechanically based predictive models, like COSMO-RS [114, 116] and COSMOSAC [144] methods form an important alternative when experimental data is missing and group contribution methods like UNIFAC fail due to missing atom groups or interaction energies. Typically, COSMO-based methods can be used to screen a long list of solvents to select a suitable candidate to extract a solute. Recently, COSMOSAC has become available in process simulators, where it can be used to calculate the equilibrium distribution of compounds in distillation, extraction or crystallization [7, 45]. The screening process can save time and money since experiments tend to be time consuming and expensive. However, the efficacy of computational methods hinges upon the accuracy and robustness of their predictions. Also computer simulation methods like Monte Carlo (MC) and Molecular Dynamics (MD) can be used to compute activity coefficients [149, 131, 150, 198, 197]. It is possible to obtain the center-of-mass averaged pair distribution function $g(r)$ of system using MC and MD methods [61]. The activity coefficients can be directly derived from $g(r)$ via the Kirkwood-Buff theory [113]. Liu *et al.* [149] showed that

---

[1]Based on the publication: COSMO-3D: Incorporating Three-Dimensional Contact Information into the COSMO-SAC Model. Ind. Eng. Chem. Res. 54(2015) 2214-2226 by Juan J. Gutiérrez-Sevillano, Kai Leonhard, Jan M. van der Eerden, Thijs J.H. Vlugt, and Gerard J. P. Krooshof.

MD calculations using the OPLS force field [104] provides more accurate thermodynamic factors than COSMO-based methods, but also that these calculations are computationally more expensive. The COSMOSAC [144] model is based on the conductor-like screening model (COSMO) [118, 114, 116]. In this approach, molecules are modeled with a quantum mechanical theory such as density functional theory[36]. This results in a charge distribution for the molecules, which represents the average electronic structure in the ground state. This distribution is then placed in a molecule-shaped cavity of a conductor (the so-called COSMO surface) which screens the atomic charges by accumulating charge on the cavity surface. This COSMO surface is composed of "segments" with a constant charge density and the charge distribution is determined by solving Poisson's equation. The screening charge is then used as an external field to recompute the charge distribution of the molecule. This is repeated until the two calculations are self-consistent. The outcome of this calculation is the charge distribution that the solvent would have to provide in order to screen a single molecule. From the screening charge distribution it is possible to compute the interaction energy between molecules by placing the COSMO charges into a thermodynamic ensemble in which they are allowed to freely associate. This is the basis for the COSMO-RS [114, 115, 116], COSMOSAC [94, 142, 141, 229, 239], and COSMO-RS(Ol) [74] algorithms. While these methods are based on the quantum mechanical descriptions of the mixture constituents, they still require a set of parameters that are fitted to experimental data. As these parameters seem to be universal for many liquid mixtures, the methods can be used to make predictions for mixtures containing novel components.

Here, we introduce the COSMO-3D model which is a modification of the COSMOSAC model. The purpose of this modification is to incorporate 3-dimensional geometric information related to the shape and size of molecules in a mixture. Depending on the shape and size of molecules, the contact area between them will be different. So far, in COSMO-based models the pairing area of the segments (a_{eff}) has an universal value. Our idea is that, in order to obtain a better thermodynamic description, this area should vary according to the molecular shape near the contact point of molecules. Since the effective contact area is reciprocally related to the coordination number of molecules, the choice of a universal effective area is like setting the coordination number (z) to a universal value. In the UNIFAC models the coordination number value is set to 10. In reality the value varies between 6 and 12, depending on the average packing structure that molecules have in the liquid; 6 for cubic packed molecules and 12 for close packed spheres. The simplification of $z = 10$ has little consequences in the UNIFAC model, since the developers had the option to repair a deviation between model and measurement by adjusting the group surface parameter and the binary interaction parameters. In the COSMOSAC model this

option is not available. As a result the COSMOSAC model will yield less accurate results for systems, in which the molecules have an average contact area that deviates from the proposed universal value. Another approach to improve activity models is given by the work on the MOQUAC activity coefficient model [22]. In this model molecular orientation of two molecules is used to improve the Gibbs energy calculation. The developers show that the MOQUAC model gives better results than the GEQUAC model. The latter has the same residual activity coefficient equation as the COSMOSAC model. The work on MOQUAC indicates that 3-dimensional structure information improves activity models. With the COSMO-3D model we will show how 3-dimensional information, as is provided by the ADF COSMO output files, can be used to improve the COSMOSAC model [94, 141, 145, 229, 239], where information on the 3-dimensional structure of molecules is not explicit. In COSMO-3D, 3-dimensional information is included by a new molecular descriptor, the so-called "molecule effective contact area". This descriptor is unique for each species and it is calculated in a preprocessing step. Their values are stored in a database and can be employed in future calculations without additional cost. Using the concept of molecule effective contact area allow us (1) to eliminate two parameters from the COSMOSAC (2010) model (A_{ES} and B_{ES}), (2) to replace a parameter (a_{eff}) by a new one (R_3), and (3) to add one parameter (b_{es}). For mixture calculations with non-hydrogen-bonding compounds, COSMOSAC (2010) has four adjustable universal parameters [94] (A_{ES}, B_{ES}, a_{eff}, and f_d), while our new method has three (R_3, b_{es} and f_d). Below we show that, although COSMOSAC uses B_{ES} as an independent parameter to describe mixtures over a range of temperatures and COSMO-3D does not need an additional parameter for temperature effects, COSMO-3D is significantly more accurate in describing experimental data.

This paper is organized as follows. In the next section we introduce the key concept: the molecule effective contact area, and how it is incorporated into the COSMOSAC model. In section 8.3, we explain the applied numerical procedures. In section 8.4 we show vapor-liquid equilibria predictions and activity coefficients obtained using our model. To show the higher accuracy of this method, a comparison with experimental data and with COSMOSAC model is also provided and discussed. Our findings are summarized in section 8.5.

8.2 Theory

8.2.1 Molecular Contact Areas

A premise of previous COSMO-based models is that the area used for segment pairings, typically referred to as a_{eff}, has a universal value [114, 144]. This contact area is used to determine the electrostatic interaction energy for any given segment pairing via Onsager's theory. In general, the area of interaction is much larger than a typical segment area on the COSMO surface, because electrostatic forces are long ranged. In the revision of COSMOSAC model, as proposed by Xiong *et al.* [239], two contact areas: a_{avg}, used in the charge density averaging (see section 2.2) and a_{eff}, to compute the restoring free energy of a solute molecule. In this work, we follow a different approach by keeping the same parameter a_{eff} for both purposes but making it dependent on the components in the mixture. Since charges are distributed over the 3-dimensional COSMO surface of the molecules, the interaction energy will depend on the shape and size of the interacting molecules. Our new method provides information in the form of pure component effective contact areas, which are derived by a procedure that uses the 3-dimensional structure of the interacting molecules. The pure component effective contact area can be seen as a molecular descriptor, which is unique for every species. This new parameter will be used to improve the COSMOSAC statistical thermodynamic calculations.

The effective contact area of a molecule is calculated using the output provided by the ADF COSMO code. The ADF COSMO output file gives not only the position, the surface charge, the area, and the potential of the surface segments at the inside of the cavity surface, but also the position of the atom centers. To calculate the effective contact area we start with two identical molecules. The choice for taking two identical molecules (cavities) is because the reference state is the pure liquid, which is expressed later in the sigma profile of the pure liquid. The procedure to determine the molecule effective contact area continues by choosing a surface segment of the first molecule. This surface segment and the center of the corresponding atom is translated and rotated to the x-axis. Subsequently a segment of the second molecule is placed against the first segment, in such a way that the vector between the segment center and the atomic center lies anti-parallel to the x-axis (Fig. 1a). This operation defines two translations and two rotations by which the coordinates of all the atom centers and all the other segments of the two molecules are calculated. When this is done a probing sphere with radius R_3 is introduced (Fig. 1b). It rotates around the x-axis over the surface of the two contacting atoms and defines on each atom a spherical cap. The size of the caps is dependent on the radii of the rolling sphere, and the two contacting atoms. By collecting the surface area of all the segments that have the coordinates within the two spherical caps, we get the local contact area of

the chosen pair of segments. Outside the spherical caps there are pairs of segments (i, j) that are in close contact. To find them fast, we first collect atom-atom center distances, r_{ij}, which fall within the range $R_i + R_j < r_{ij} < R_i + R_j + 2R_3$. Subsequently the pairs of close contacting segments are found by $d_{ij} < r_{ij} - R_i \cos \theta_i - R_j \cos \theta_j$ (Fig. 1c). In here r_{ij} is given by the position of the atom centers, R_i and R_j by the atom radii and the angles are given by the position of the probing sphere. The sum
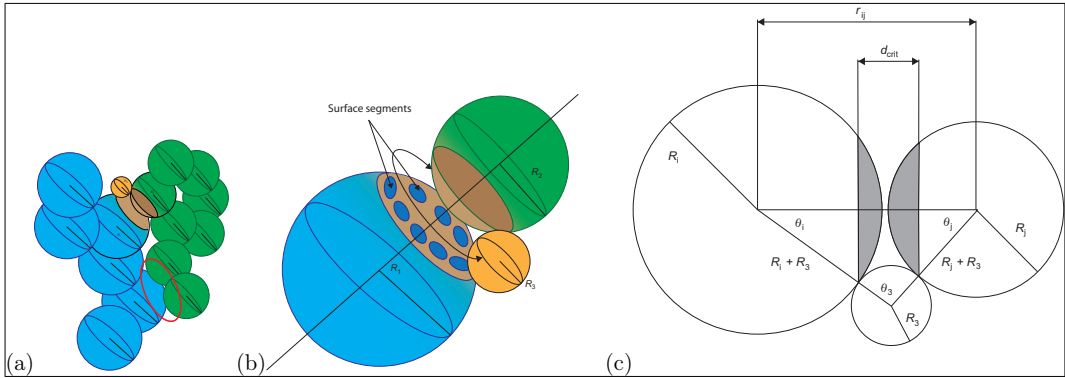
Figure 8.1: Schematic picture of the calculation of the molecule effective contact area. (a) Two identical molecules, colored green and blue, at one of the many contact configurations. (b) Close-up of Fig 8.1.a. An atom of the blue molecule with radius R_1 is within contact range of an atom of the green molecule with R_2. The probe sphere (orange) with radius R_3 defines the collection of surface pair segments (blue patches in the orange caps). (c) Plane projection of the two spheres of (Fig. 8.1 b) explaining the distance criterion $d_{ij} < d_{\text{crit}}$ by which the area of 'contact' (gray) and the blue patches shown in Fig. 8.1 b are defined .

of the local contact area and the external area is the total area of the established configuration, a_{uv}^θ. Subsequently, while keeping the contact point fixed, the second molecule is rotated to 5 other angles ($\theta = \pi/3$, $2\pi/3$, π, $4\pi/3$, $5\pi/3$) to define 5 other configurations, for which the contact a_{uv}^θ, is calculated. This operation will not change the local contact area, since the spherical caps will not change. But the rotation will change the distance between the segments of the two molecules and thereby the area of the external contacts. This is repeated for all contact pairs that can be made between the two molecules. So, for a molecule with N_S surface segments we evaluate in total $6N_S^2$ of configurations. The sum of all contact areas is divided by $12N_S^2$ to give the effective contact area of one molecule as embedded in its own pure liquid, the reference state. The molecule effective contact area in this work is the average of all possible pairings between the atoms of two identical molecules (i

and i):

$$a_i = \frac{1}{2N_S^2 N_\theta} \sum_{u,v,\theta} a_{uv}^\theta, \tag{8.1}$$

where a_{uv}^θ is the contact area calculated for the pairing of segments (u, v) from molecules i and i', respectively at angle θ, N_S is the number of segments, and N_θ is the number of different configurations as a result of the rotation of the molecules around the axis through the contact point, which in our case is 6. The factor 2 arises from the fact that we summed the area of both caps in the calculation. The aforementioned algorithm requires many translations, rotations, and distance calculations, e.g. in the computation of the effective contact area of acetone there are $6 \times (853)^2 = 4365654$ configurations to evaluate. For each configuration the distance between all pairs of segments have to be checked, which imply 853^2 calculations. In order to reduce the computation time quaternions [106, 71] are used for the 5 additional rotations around the axis of contact. The calculation of the effective contact area is done with two identical molecules and the result is stored in a database. For the effective contact area of a mixture, we opted for the mixing rule of Eq. 8.2, instead of calculating all pairs of molecule types that are in the mixture, in order to keep the computational time acceptable in engineering flow-sheeting programs. Table 1 lists the number of surface segments (NSS), the cavity volume, the cavity area from the ADF COSMO file, and the molecule effective contact area as found by the above explained algorithm. With the fitted value $R_3 = 0.5$ Å(the fit of R_3 is not very sensitive and therefore we report just one digit), the contact areas of the molecules of the investigated binary systems range from 5.6 to 6.5 Å^2 and agrees very well with the optimum value range of 4 to 7 Å^2 for the group of neutral compounds in the effective surface area sensitivity analysis of Wang *et al.* [229]. This is smaller than the COSMOSAC effective area (7.25 Å^2). Table 1 also shows that flat molecules have larger contact area than bulkier molecules, as a result of the multiple contacts outside the local contact area. For each mixture a system effective contact area is calculated, according to

$$a_{\text{eff}} = \frac{1}{N} \sum_{j=1}^{N} a_j, \tag{8.2}$$

where a_j is the molecule effective contact area of pure component j, and N the number of species in the mixture. We mention that the above mixing rule is only system dependent and not concentration dependent. This is a requirement that follows from the Gibbs-Duhem criterion. The new universal parameter R_3, i.e. the radius of the probing sphere, substitutes the old parameter a_{eff}. The benefit is that

Molecule	ADF COSMO NSS	ADF COSMO Volume (Å^3)	ADF COSMO Area (Å^2)	COSMO-3D a_{eff} (Å^2)
2-methoxy-2-methylpropane	1166	134.20	141.87	5.62
acetone	843	85.47	103.28	5.66
methylcyclohexane	1286	150.18	154.50	5.67
propanal	853	86.40	103.91	5.67
cyclohexane	1138	130.15	138.20	5.68
pentane	1172	124.01	140.11	5.68
oxolane	973	100.45	114.94	5.70
diisopropyl ether	1375	156.02	164.93	5.72
2-methylpropanal	1005	107.21	121.63	5.73
methyl acetate	930	97.15	114.98	5.75
hexane	1362	145.42	160.38	5.78
2-butanone	1002	107.04	122.86	5.79
1,4-dioxane	1022	110.99	123.10	5.81
heptane	1499	166.15	179.11	5.81
1-hexene	1271	139.35	154.42	5.83
3-pentanone	1177	128.78	143.08	5.84
ethyl acetate	1124	118.99	136.79	5.85
pentanal	1206	129.48	145.35	5.90
ethylbenzene	1308	154.16	160.57	6.14
benzene	958	111.88	123.02	6.27
benzaldehyde	1167	137.63	145.93	6.50

Table 8.1: Number of surface segments (NSS), cavity volume and cavity area from the ADF COSMO QM calculation, and effective area of molecules obtained in this study. The areas were computed using the method described in section 2 with $R_3 = 0.5$ Å.

geometric information about the effective contact area of molecules is taken into account in a more realistic way.

8.2.2 COSMO-3D model

In the COSMOSAC model, the computation of activity coefficients is based on the charge density distribution of the molecular species. This distribution is obtained from Quantum Mechanics (QM) calculations and incorporated in the form of so-called σ-profiles. As in most previous COSMO-models, we smooth the raw QM segment charge distribution to generate σ-profiles [144]. This procedure is introduced to remove artifacts, i.e. high charge densities, which can occur near sharp zones of the enclosing conductor surface. The applied function is actually a filter that smooths out these outliers. We use the apparent screening charge distribution of previous

models, which gives for the smoothed segment charge densities.

$$\sigma_u = \frac{\sum_{v=1}^{N_S} \left(\sigma_v^* \frac{a_v a_{\text{eff}}}{a_v + a_{\text{eff}}} \exp\left[-f_{\text{d}} \frac{\pi \lambda_{uv}^2}{a_v + a_{\text{eff}}} \right] \right)}{\sum_{v=1}^{N_S} \left(\frac{a_v a_{\text{eff}}}{a_v + a_{\text{eff}}} \exp\left[-f_{\text{d}} \frac{\pi \lambda_{uv}^2}{a_v + a_{\text{eff}}} \right] \right)} \tag{8.3}$$

where N_S is the total number of segments of a COSMO surface, σ_v^* and a_v are the raw COSMO charge densities and areas of each segment v forming the COSMO surface, λ_{uv} is the distance between segments u and v, and f_{d} is a universal parameter that defined the decay of the filter function. We mention that in Eq. 8.3 the system effective contact area, a_{eff}, is no longer a universal value, as it was in previous COSMO models [94]. This a_{eff} is a constant calculated by the system average (Eq. 8.3) and is used for all mixture concentrations of the investigated system, including the pure components.

With the smoothed charged densities obtained using the a_{eff} we can define the σ-profiles for a given molecule,

$$p_i(\sigma_m) = \frac{A_i(\sigma_{\text{m}})}{A_i} \tag{8.4}$$

where $A_i(\sigma_{\text{m}})$ is the area of molecule i with charge density σ_{m}, and A_i is the total area of the molecule. The σ-profile for a mixture is obtained from:

$$P(\sigma_{\text{m}}) = \frac{\sum_{i=1}^{N} x_i A_i p_i(\sigma_{\text{m}})}{\sum_{i=1}^{N} x_i A_i} \tag{8.5}$$

where x_i and A_i are the mole fraction and surface area of component i, respectively. Therefore the σ-profile of the mixture is the mole average of the σ-profiles of pure compounds. According to the original COSMOSAC models [146, 144, 143], the activity coefficients of solute i in the solution S ($\gamma_{i/S}$) are given by:

$$\ln\gamma_{i/S} = \ln\gamma_{i/S}^{\text{res}} + \ln\gamma_{i/S}^{\text{comb}} \tag{8.6}$$

where $\ln\gamma_{i/S}^{\text{res}}$ is the residual activity coefficient and $\ln\gamma_{i/S}^{\text{comb}}$ is the combinatorial contribution from mixing. This term, given by Huyskens [99], as discussed in chapter 2, reads:

$$\ln\gamma_{i/S}^{\text{comb}} = \frac{1}{2} \left[\ln\frac{\phi_i}{x_i} + 1 - \frac{\phi_i}{x_i} \right] \tag{8.7}$$

where x_i and $\phi_i = V_i x_i / \sum_i^N V_i x_i$ are the concentration and volume fraction of component i, respectively, N the number of components of the mixture, and V_i is taken

to be the COSMO volume of each species. In our work, this volume is approximated with the volume of the molecule-shaped cavity used in the QM calculations [200]. Essentially, Eq. 8.7 comprises an interpolation between the entropy of mixing for an ideal gas and a Flory-Huggins [53, 96] term which is typically used to describe fluid phases [21]. Note that we could use other, more complex forms of the combinatorial term, but Eq. 8.7 shows to be sufficiently accurate. E.g. acetone and heptane are the compounds in our set that have the largest difference in area and volume and thereby show the largest deviation from ideal mixing. The limiting combinatorial activity coefficients of acetone and heptane are 0.91 and 0.87, respectively, according to Eq. 8.7. For the Staverman-Guggenheim combinatorial activity coefficient, which is used in COSMOSAC model, the values are 6% lower (0.87 and 0.82 respectively). The difference between the two combinatorial models is much lower than the overall improvement made by the COSMO-3D model as depicted in Fig. 3a, where we observe a 36% difference in activity coefficients. For molecules of similar size the difference in the combinatorial models is even smaller. Then, we turn to the residual activity coefficients in Eq. 8.6. For component i in a mixture we write

$$\ln \gamma_{i/S}^{\text{res}} = \frac{A_i}{a_{\text{eff}}} \sum_{\sigma_{\text{m}}} P(\sigma_{\text{m}}) \left\{ \ln \left[\Gamma_{i/S}(\sigma_{\text{m}}) \right] - \ln \left[\Gamma_{i/i}(\sigma_{\text{m}}) \right] \right\} \qquad (8.8)$$

where a_{eff} is the effective contact area for the mixture as defined in Eq. 8.2 and A_i is the surface area of component i. The summation is taken over all smoothed segment charge densities σ_{m} in the σ-profile of the mixture, $P(\sigma_{\text{m}})$. The terms $\ln \left[\Gamma_{i/S}(\sigma_{\text{m}}) \right]$ and $\ln \left[\Gamma_{i/i}(\sigma_{\text{m}}) \right]$ are defined as the segment activity coefficients for the mixture and for the pure component at some reference state. Following the work of Wang and Sandler [229], the segment activities are functions of the species in the system via the interaction energy and effective contact area, and they are given by

$$\ln \left[\Gamma_{i/S}(\sigma_{\text{m}}) \right] = -\ln \left\{ \sum_{\sigma_{\text{m}}} P(\sigma_{\text{m}}') \ln \left[\Gamma_{i/S}(\sigma_{\text{m}}') \right] \exp \left[\frac{-E(\sigma, \sigma')}{RT} \right] \right\} \qquad (8.9)$$

In this equation, the segment-segment interaction energies are given by

$$E(\sigma, \sigma') = c_{\text{ES}}(\sigma + \sigma')^2 \qquad (8.10)$$

The constant c_{ES} in Eq. 8.10 is typically obtained from fitting to experimental data. Hsieh and Sandler [94] define c_{ES} as a function of the temperature and two fitting parameters:

$$c_{\text{ES}} = A_{\text{ES}} + \frac{B_{\text{ES}}}{T^2} \qquad (8.11)$$

Previously, Wang and Sandler [229] defined c_{ES} as follows:

$$c_{\text{ES}} = f_{\text{pol}} \frac{0.3 \left(a_{\text{eff}}\right)^{3/2}}{2\varepsilon_{\text{o}}}, \tag{8.12}$$

where ε_{o} is the permittivity of the vacuum, f_{pol} is the polarization factor (which is a universal constant coming from the used QM-model), and a_{eff} is a fit parameter. At this point we introduce a modification in Eq. 8.12 by using a system dependent a_{eff} parameter calculated by Eq. 8.2 and incorporating a new fit parameter b_{es}. The resulting expression equals

$$c_{\text{ES}} = f_{\text{pol}} \frac{0.3 \left(a_{\text{eff}}\right)^{3/2}}{2\varepsilon_{\text{o}}} b_{\text{es}} \tag{8.13}$$

The polarization factor f_{pol} can be calculated from the COSMO surfaces and the molecular ground state energies for a given set of compounds [141]. For our database of compounds, $f_{\text{pol}} = 0.71$ [200]. The parameter b_{es} should be unity according to Onsager's theory, but there are two steps in the computation procedure that are sources for deviations. First, the discretization of the COSMO surface is purely artificial and therefore inexact. Secondly, the σ averaging procedure applies an empirical formula, and although it is meant to smooth out inconsistencies in the charge densities on the COSMO surface, it loosens the boundary condition of a perfect conductor. After all, changing a surface charge density is changing the local electric field near the surface of a conductor according to Gauss law [102]. The systematic error of the smoothening function can be seen from the change in total charge of the molecule, which should be zero. The sum of the original COSMO segment charges is of the order of 10^{-9} e. But after smoothening this has become of the order of 10^{-3} e. To minimize this error, b_{es} is introduced as a fit parameter (see section 8.3). Note that in Eq. 8.13, opposite to the work of Wang and Sandler where a_{eff} is an universal constant, a_{eff} depends on the molecule effective contact areas of the molecules in the mixture (Eq. 8.2).

This method for computing the energy parameter c_{ES} differs significantly from previous methods, as the parameter (c_{ES}), is dependent on the species involved. Our modification is based on the fact that the size and shape of the molecules should play a role in the energies of contact through their nominal interaction areas. We also note that this model does not take into account strong interactions related to hydrogen bonds. The incorporation of this may be added in future work. With these modifications our model has three universal fitting parameters (f_{d} in Eq. 8.3, b_{es} in Eq. 8.13, and R_3). The latter parameter is used to obtain a_{eff}. In contrast, the COSMOSAC (2010) method has four universal parameters (A_{ES}, B_{ES}, a_{eff}, and f_{d}).

8.3 Numerical Procedure

The contact areas were computed using COSMO surfaces generated by the Amsterdam Density Functional Theory code (ADF release 2010.01) [200, 56, 213]. These surfaces are available in a database which comes with the COSMO-RS implementation [200, 182]. Our area preprocessor was used to compute the optimal b_{es} for a discrete set of probe radii in the range $R_3 \in [0.1, 2.0]$ Å. In the calculation of the contact area we ignored the occurrence of overlapping molecules. This is a systematic error, which also exists in COSMOSAC. In fact, like COSMOSAC, all surface segments can make contact through the use of σ-profiles. This systematic error seems to be of small magnitude, since the obtained phase diagrams are in close agreement with the experimental results. This phenomenon of overlapping molecules is also ignored in the UNIFAC models [60]. Also in here, all UNIFAC groups are assumed to be accessible, although some interaction might be impossible due to the structure of the interacting molecules.

The three general parameters for the new model ($R_3 = 0.5$ Å, $b_{es} = 1.41$, $f_d = 1.73$) were fitted to experimental vapor-liquid equilibria data for the binary mixtures of the training set (see Table 2). The value of f_{pol} is obtained directly from the used QM-model, i.e. ADF. Note that the element-specific parameters (Atom Dispersion Constants) were not modified, keeping the default values of the ADF implementation [200]. The objective function used for the minimization is

$$\min\left\{\left[\frac{1}{N_{\exp}}\sum_{k=1}^{N_{\exp}}\left(\frac{P_k^{\exp}-P_k^{pred}}{P_k^{\exp}}\right)^2\right]^{1/2} + \left[\sum_{k=1}^{N_{\exp}}\frac{\left(y_k^{\exp}-y_k^{pred}\right)^2}{N_{\exp}}\right]^{1/2}\right\}, \quad (8.14)$$

where $P_k^{\exp}$ is the experimental vapor pressure, P_k^{pred} the predicted vapor pressure, $y_k^{\exp}$ the experimental vapor fractions, y_k^{pred} the predicted vapor fraction, and $N_{\exp} = 598$ is the total number of experimental data points in the training set. All calculations for COSMOSAC are also obtained using the ADF code [182] corresponding to the COSMOSAC model of Hsieh *et al.* [94]. The contact areas for the molecules used in this work are listed in Table 1.

8.4 Results and Discussion

Prediction of thermodynamic properties of mixtures (vapor-liquid equilibria (VLE) and activity coefficients) are obtained using our new COSMO-3D method. The predictions are compared with those obtained using COSMOSAC (2010) [94]. Here we note that we did not refit the COSMOSAC parameters to the training set. All experimental data used for the training and the test set are summarized in table 8.2 in appendix S.

Mixture	T (°C)	Ref.
1-hexene/ethyl acetate	60, **40**	[66, 68]
2-butanone/heptane	45	[210]
2-methoxy-2-methylpropane	80, 90, 100	[135]
/methyl acetate		
2-methylpropanal/heptane	45,**61.85**	[49, 68]
acetone diisopropyl/ether	70.11, 80.12, 90.14	[136]
acetone/heptane	**0**,40,**50**	[122, 68]
benzene/cyclohexane	**10**, **24.91**, **25**, **38.5**,	[154, 242, 68]
	39.90, 40, **50**, **55**, **60**, **70**	
benzene/heptane	25, **35**, 40, **55**, **60**, **80**	[72, 155, 68]
benzene/hexane	**25**, 30, 40,50, 60	[139, 193, 68]
cyclohexane/heptane	25, 40	[242, 193]
ethyl acetate/methylcyclohexane	56.85, 76.85	[233]
ethylbenzene/benzaldehyde	75, **85**, 95	[107, 68]
heptane/3-pentanone	**26**, 65, 80, 95	[65, 68]
heptane/pentanal	75	[212]
hexane/1,4-dioxane	35	[24]
oxolane/benzene	**30**, **40**, **50**, 60	[64, 68]
oxolane/cyclohexane	**30**,40, **50**, 60	[237, 68]
oxolane/ethyl acetate	**40**, 60	[237, 68]
oxolane/hexane	**30**, **40**, **50**, 60	[237, 68]
pentane/hexane	25.55, 30.55, 35.55	[188]
pentane/propanal	40	[49]
propanal/cyclohexane	45	[163]

Table 8.2: Mixtures and temperatures of the training and test (bold) set. Experimental data at the listed temperatures were taken from references in the third column.

A first overall comparison between activity coefficients calculated by COSMOSAC for non-polar/non-polar and polar/non-polar mixtures is shown in Fig. 8.2. As can
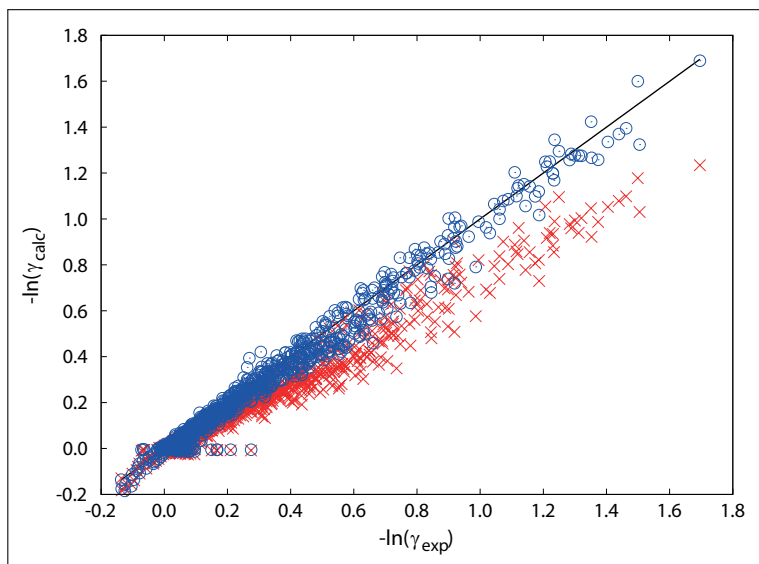
Figure 8.2: Parity plot of experimental and computed activity coefficients of the training set for COSMO-3D (blue circles) and standard COSMOSAC (red crosses). The solid line represents ideal agreement.

be observed in this figure the COSMOSAC model gives for the training set under-estimated values. The COSMO-3D model with optimized parameters gives better results than the COSMOSAC model.

Two examples of the test set, the binaries acetone/heptane and cyclohexane/heptane are shown in Fig. 8.3, where the activity coefficients are plotted as a function of the mole fraction of component 1 in the liquid mixture. The left panel of fig. 8.3 shows that COSMO-3D slightly overestimates the activity coefficients for the mixture acetone/heptane at 40°C. Nevertheless, the results are in better agreement with experimental data than those predicted by COSMOSAC, which clearly underestimates experimental data. Accurate correlations are obtained for all the mixtures, which were used in the training set, except for the systems pentane/hexane and cyclohexane/heptane. The computed activity coefficients for the system cyclohexane/heptane are shown in right panel of Fig. 8.3. This binary system shows experimentally a positive deviation from Raoult's law, but COSMOSAC and COSMO-3D show a negative deviation. The reason for this negative deviation is the combinatorial term (Eq. 8.7); the σ-profiles of these compounds, which show a single sharp peak centered around zero, give a residual activity which is very close to unity. A way to improve this is by introducing a dispersive interaction term in the model [239]. The fact that PC-SAFT [78] predicts a correct activity behavior supports this hypothesis. Activity coefficients

a: acetone/heptane at 40 °C. **b:** cyclohexane/heptane at 25 °C.
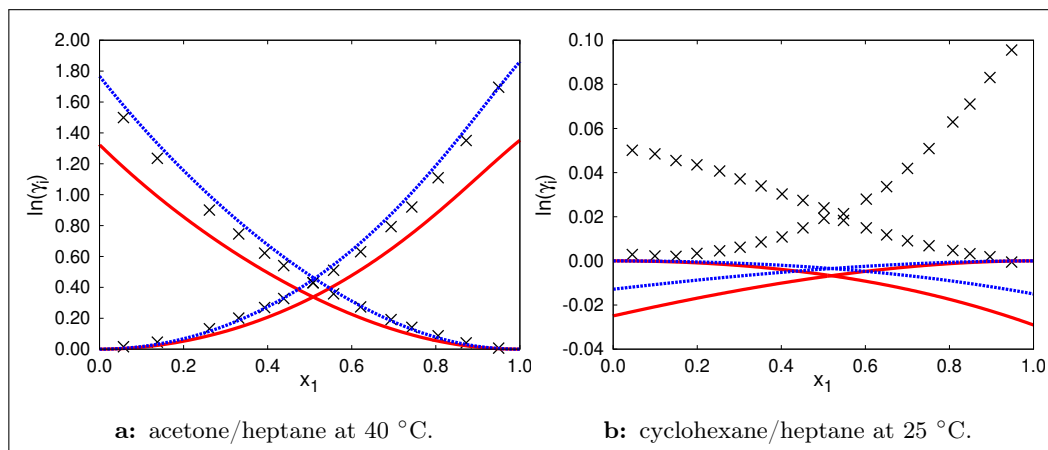
Figure 8.3: Activity coefficients obtained with COSMOSAC (red line) and COSMO-3D (blue line) as a function of the mole fraction in the liquid phase of the first component. Experimental data (black crosses) from [122, 193].

for all the systems can be found in the Supporting Information of [82]. The accurate description of the training set by the COSMO-3D model can also be observed in the isothermal vapor-liquid equilibria. In the left panel of Fig. 8.4 the agreement of VLE calculations for the mixture 2-methylpropanal/heptane at 45°C can be observed. In the right panel of Fig. 8.4 the VLE of the binary 1-hexene/ethyl acetate at 60°C is plotted. Note that COSMO-3D gives an azeotrope, while COSMOSAC gives a zeotropic result. In many cases COSMO-3D provides a better azeotrope description than the COSMOSAC model (acetone/diisopropylether, acetone/heptane, benzene/cyclohexane, ethyl acetate/methylcyclohexane, heptane/3-pentanone, heptane/pentanal, oxolane/hexane, pentane/propanal, and propanal/cyclohexane). These VLE diagrams are included in the Supporting Information of [82]. A concrete measure of the model quality can be obtained using the average deviations calculated from experimental thermodynamic data over the training and test sets. For the pressure this deviation is defined by

$$P^{\mathrm{dev}} = \frac{1}{N_{\mathrm{exp}}} \sum_{i=1}^{N_{\mathrm{exp}}} \frac{|P_i^{\mathrm{exp}} - P_i^{\mathrm{pred}}|}{P_i^{\mathrm{exp}}} \times 100 \qquad (8.15)$$

For the deviation in vapor fraction the average absolute deviation is used, which is equated as:

$$y^{\mathrm{dev}} = \frac{1}{N_{\mathrm{exp}}} \sum_{i=1}^{N_{\mathrm{exp}}} |y_i^{\mathrm{exp}} - y_i^{\mathrm{pred}}| \times 100. \qquad (8.16)$$

a: 2-methylpropanal/heptane at 45 °C.   **b:** 1-hexene/ethyl acetate at 60 °C.
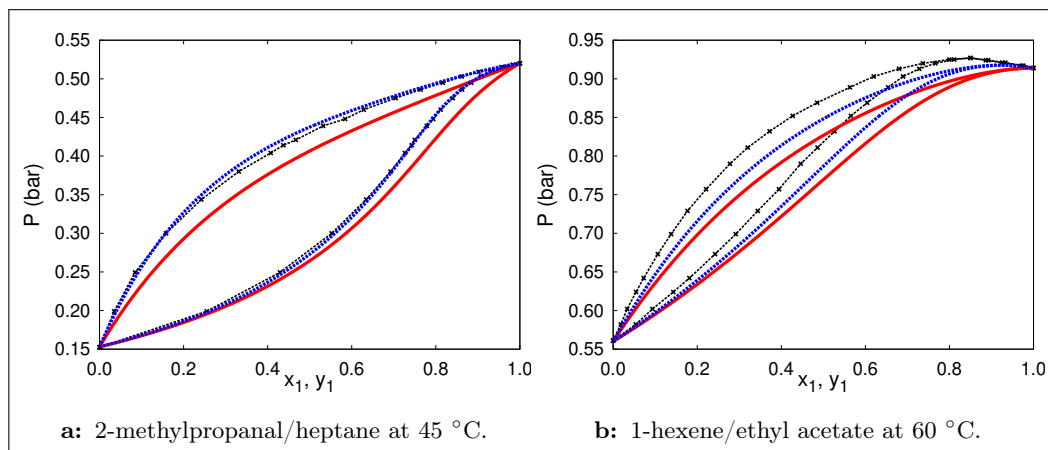
Figure 8.4: VLE curves obtained with COSMOSAC (red line) and COSMO-3D (blue line). Experimental data (black crosses) [66, 49].

For the deviation in activity coefficients the root mean square deviation is applied:

$$\gamma^{\text{dev}} = \frac{1}{N_{\exp}} \sqrt{\sum_{i=1}^{N_{\exp}} \left(\gamma_i^{\exp} - \gamma_i^{\text{pred}} \right)^2}. \tag{8.17}$$

Table 8.3 lists the values of these deviations of the mixtures obtained with COSMOSAC [94] and COSMO-3D models, for the training set. Detailed information for each binary mixture is given in appendix S. The training set contained 598 experi-

Mixture	P^{dev}(Eq. 8.15)			y^{dev}(Eq. 8.16)			γ^{dev}(Eq. 8.17)		
	SAC	REF	3D	SAC	REF	3D	SAC	REF	3D
Training set	3.89	2.36	1.56	1.41	0.89	0.63	0.24	0.14	0.08
Test set	3.14	2.16	1.66	1.08	0.84	0.69	0.31	0.23	0.16

Table 8.3: Overall results of the training and test set for the COSMOSAC [94] (SAC), the refined COSMOSAC (REF), and the COSMO-3D (3D) models. Details in appendix S
.

mental and computed data points. The overall deviations for the whole set of points are lower using our new method (we removed one anomalous experimental data point from the training set in the system oxolane/cyclohexane at 60 °C, (see appendix T, Fig. S32). This also happens if we compute the three deviations defined earlier for each system. As can be observed in Table 8.3, the new model has significantly better performance. As mentioned before COSMOSAC (2010) underestimates the
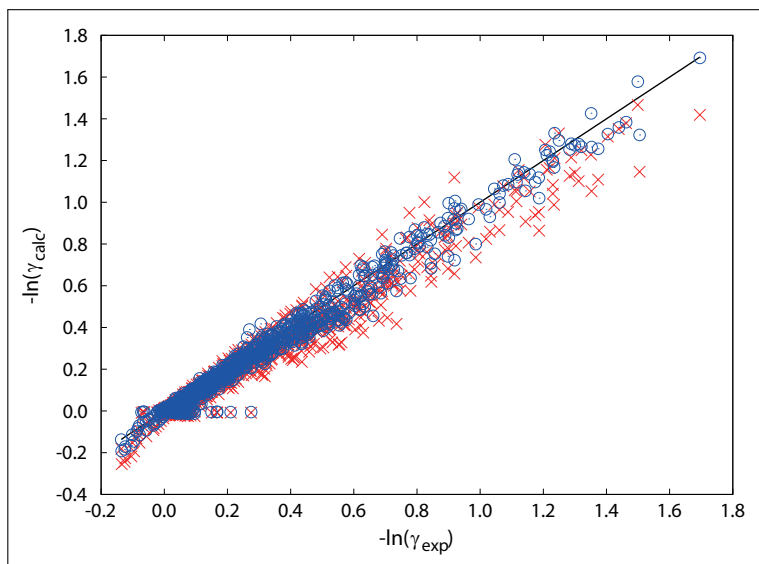
Figure 8.5: Parity plot of experimental and computed activity coefficients for the training set using COSMO-3D method (blue circles) and the refined COSMOSAC (red crosses). The solid line represents ideal agreement.

activity coefficients of our training set. A reason for this could be that the COSMOSAC misfit energy parameters is an average for a larger data-set, or that the presence of associating systems in the training set of the COSMOSAC(2010) version has biased the energy parameter. For that reason we have adjusted the energy parameter to get a better agreement in the training set. Instead of doing a refitting of the COSMOSAC(2010) energy parameters it is sufficient to set the c_{ES} parameter of the COSMOSAC model (Eq. 8.11) equal to the c_{ES} of the COSMO-3D model (Eq. 8.13) by using the average value of the effective area of the compounds used in the training set ($a_{eff} = 5.82$ Å^2). Effectively this means that the energy parameter of the COSMOSAC(2010) was scaled down by the effective contact area from 7.25 to 5.82 Å^2. We denote this modification as the refined COSMOSAC model. As we can see from the VLE diagrams of the mixture (Appendix T) and the statics in table S.1 this gives a significant improvement for our training set (Fig. 8.5). These results show also that the effective contact area is not a universal constant as was assumed in the COSMOSAC model, but a system dependent constant.

The differences between the three models can be explained by comparing the systems hexane/1,4-dioxane and pentane/propanal (Fig. 8.6). The left panel of Fig. 8.6 shows that COSMO-3D and the refined COSMOSAC provides a more accurate description of the phase envelope of hexane/1,4-dioxane than the original

COSMOSAC model, while in the right panel of Fig. 8.6 COSMO-3D clearly outperforms the other two models. The better description of COSMO-3D is a result of the system dependent effective area, which scales the interaction energy (Eq. 8.13). In the COSMOSAC, where $A_{ES} = 6525.69$ (kcal/mol)·($\text{Å}^4/\text{e}^2$), and the refined COSMOSAC model, where $A_{ES} = 7850.69$ (kcal/mol)·($\text{Å}^4/\text{e}^2$), the effective areas, 7.25 Å^2 and 5.82 Å^2, respectively, are assumed to be a constant for all systems. For the COSMO-3D model this is not the case. The binary system hexane/1,4-dioxane in the left panel of Fig. 8.6) has a system effective area of $(5.78 + 5.81)/2 = 5.80$ Å^2, while the binary system in the right panel of Fig. 8.6, pentane/propanal, has $(5.68 + 5.67)/2 = 5.68$ Å^2. Since the effective area of the COSMO-3D and the re-
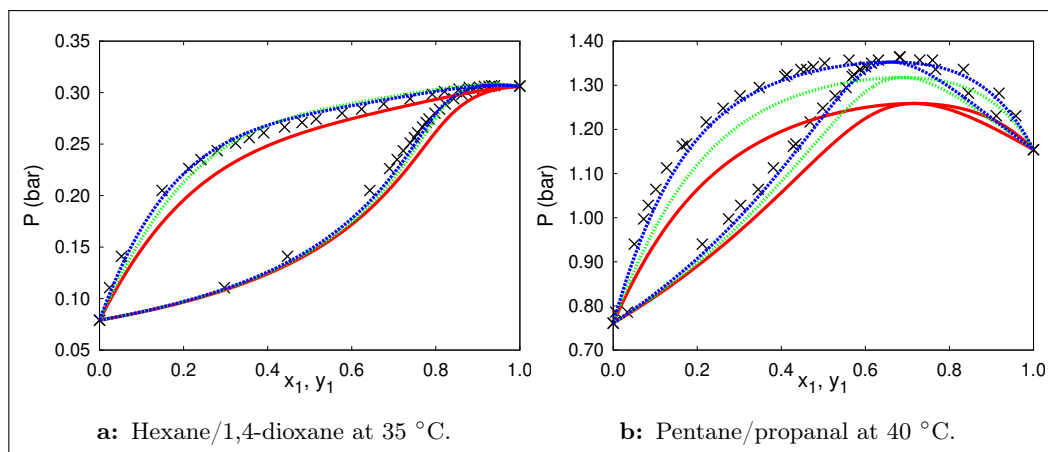
Figure 8.6: VLE curves obtained with original COSMOSAC 2010 in red, refined COSMOSAC in green, and COSMO-3D in blue. Experimental data (black crosses) taken from [24, 49].

fined COSMOSAC model are nearly the same in case of the first binary system the calculated activity coefficients and phase envelop are also nearly the same. However, in case of the second binary system there exists a significant difference in the effective areas between the COSMO-3D and the refined COSMOSAC model. This gives a significant difference in the activity coefficients. As a result the COSMO-3D model shows a much better description of the phase envelope. The above shows that the COSMO-3D model, due to the system dependent effective area, gives a better phase equilibrium description than the two COSMOSAC models, which treat this as a universal constant.

A second validation of our COSMO-3D model was carried out by performing calculations on a test set of 455 data points, consisting of mixtures with molecules of the training set but at different temperatures. The mixtures, temperatures, and experimental data references are listed in Table 8.3. Detailed information on the
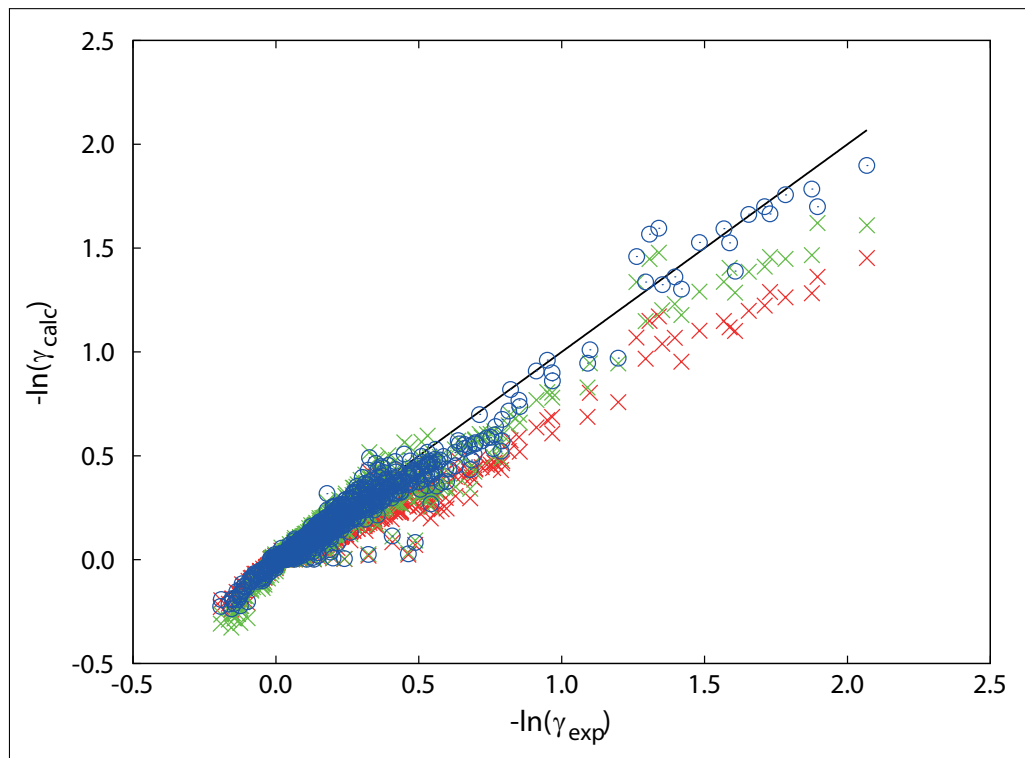
Figure 8.7: Parity plot of experimental and predicted activity coefficients for the test set using COSMO-3D method (blue circles), and the standard (red crosses) and refined COSMOSAC (green crosses). The solid line represents ideal agreement.

test set is given in appendix S. An overall comparison of the activity coefficients predicted using the three models is shown in Fig. 8.7. Activity coefficients predicted using COSMO-3D are closer to the experimental results than those predicted using original COSMOSAC and refined COSMOSAC. In all mixtures our new method provides better agreement with experimental data except for the mixture benzene/-cyclohexane, where at some temperatures, COSMOSAC is more accurate. This can be observed from the activity coefficient plots in appendix T and deviations in Table 8.2 of Appendix S. Again, overall deviations and deviations of each system are lower for our model (we removed one anomalous experimental data point from the training set in the system acetone/heptane at 50 °C, see appendix S, Fig. S58). Using the refined COSMOSAC model we get a more accurate description than using the original COSMOSAC model. This support the idea that COSMOSAC can be improved by adjusting just one parameter (A_{ES}), resulting in a major improvement for the mixtures of this work. Nevertheless, our COSMO-3D approach, though using one fitting parameter less, produces a further improvement.

8.5   Conclusions

We have developed a variation of the COSMOSAC model, COSMO-3D, for computing activity coefficients for non-polar and slightly polar systems (hydrocarbons, ethers, ketons, esters, and aldehydes). We emphasize that the development has not been carried out for systems that include hydrogen bonding or molecules that strongly associate, since our objective was to explore a way to incorporate the 3-dimensional structure of a molecule in the COSMOSAC model. For this purpose COSMO-3D is based on a new descriptor called molecule effective contact area a_{i} (Eq.1), which is unique for each species. We have designed a method to calculate this descriptor taking into account the 3-dimensional structure of a molecule. This effective contact area needs only to be computed once for each species and can be stored for later use. The molecule effective contact area has been incorporated into the COSMOSAC formalism in such a way that the effective area (a_{eff}) (Eq. 8.2) and the segment-segment interaction energy ($E(\sigma, \sigma')$) depend on it (Eq. 8.13. COSMO-3D requires 3 parameters (R_3, b_{es}, and f_{d}), to describe systems between 10 and 100 °C, this is one less than the COSMOSAC model (A_{ES}, B_{ES}, a_{eff}, and f_{d})

At this point of the development it is unclear whether COSMO-3D requires a temperature dependent parameter to make the model applicable for a larger temperature range. A way to investigate this further would be to check the prediction of the excess enthalpy. These parameters have been fitted using a training set of binary mixtures containing non-polar and polar compounds. The fitted parameter values are in a physically plausible range. The accuracy of the method has been tested

by comparing with experimental data and with the COSMOSAC (2010) model. We show (1) that COSMOSAC can be improved by adjusting the parameter A_{ES}, and (2) that the COSMO-3D model, though with one parameter less, gives a further improvement. COSMO-3D is computationally two-three times more expensive than the traditional COSMOSAC, but the fitting parameters of the COSMO-3D model have a more direct physical interpretation, suggesting that COSMO-3D will be a better starting point for further improving COSMO approaches in the future. The results here show that the COSMO screening models are sensitive to the shape of the screening surfaces and for the temperature range studied, the screening energy for a mixture can be computed directly from the molecule area.

8.6 Acknowledgement

8.7 Note

To ensure that the method is thermodynamically consistent we have checked that COSMO-3D obeys the Gibbs-Duhem equation, just like COSMOSAC. All the results obtained have passed the 'integration' test and 'differential' test [103] described in appendix R.2.In appendix S a comparison of activity coefficients and VLE diagrams obtained with COSMO-3D and COSMOSAC for the training and the test sets is given.

Chapter 9

Main conclusions and outlook

9.1 Main conclusions

In this thesis we have discussed and evaluated various activity coefficient contributions, which were derived in different ways. In retrospect we summarize the following main conclusions:

- A generalized equation has been derived to replace the Staverman-Guggenheim combinatorial activity coefficient (Eq. 3.36).

- It was revealed that the Donohue-Prausnitz exponent (Eq. 2.23) to modify the volume fraction in the combinatorial Flory-Huggins equation is physically incorrect, implying that UNIFAC models with this modification are subject to systematic errors. In fact Table 2.1 shows that the UNIFAC(Do) model is not self-consistent in view of the defined coordination number $z = 10$, while the relative surface area and volume yield negative values. The original UNIFAC model is less sensitive to it, although it would have been better to use equation the general expression for the number of nearest neighbors.

- The Huyskens model (Eq. 2.31) and the original UNIFAC model, with or without Donahue-Prausnitz modification, are athermal models for alkane mixtures, which conflicts with the experimentally observation that the mixing of alkanes is endothermal. Models to describe the endothermal mixing of alkanes need, therefore, at least a temperature-dependent activity coefficient, of which the dispersion contribution is the most logical choice.

- We have shown that the original Oishi-Prausnitz free-volume model is inconsistent. This inconsistency might have its roots in an incorrect entropy expression proposed by Flory. A more consistent equation has been derived, but it

is advised not to use this model to capture free-volume contributions in activity coefficient models (ACMs), since the compression factor, and thereby the activity coefficient expression, deviates more than 10-20% from the Carnahan-Starling (CS) model, while this model is generally accepted as the most accurate model for liquid mixtures of hard-spheres for packing fractions below 0.5. The CS model is widely used in hard-sphere chain equation of state models. Therefore, the free-volume ACMs derived from a hard-sphere chain equation of state models is a better option (Eq. 2.62) than the Oishi-Prasnitz model.

- We have demonstrated that dispersive interaction between the unequal sized hard spheres leads always to positive deviation from Raoult's law (Eq. 2.56), while linking spheres into a chain invokes negative deviation from ideal solubility (Eq. 2.62).

- A generalized expression for the number of nearest neighbors has been derived (Eq. 4.3), which converts lattice models into off-lattice equivalents and eliminates the requirements of defining a lattice coordination number and a reference surface area.

- We have demonstrated that the aforementioned generalized expression for the number of nearest neighbors and the generalized Guggenheim expression for the combinatorial activity coefficient in combination with the COSMOSPACE model gives excellent description for the phase behavior of alkane - alcohol systems (Chapter 4).

- We have improved the derivation for the residual part of the UNIQUAC model by removing an inconsistency in the definition of interaction energy (Appendix C).

- We have revealed that all cEoS yield a van Laar type of expression for the dispersion contribution of the activity coefficient, which differ only in a cEoS specific constant, and that the expression do not match experimental results due to the neglect of intramolecular interactions and the use of a too large covolume. Therefore, these equations are applicable for the descriptions of phase equilibria, rather than applicable for predictions.

- We have made clear that the PC-SAFT equation of state is only applicable for mixtures and pure compounds, where the packing fraction of the liquid is below 0.46. For denser liquids the first perturbation integral gives nonphysical results . In fact, the original expression for the first perturbation integral does not have this issue and has a correct starting point at the ideal gas state. It might

explain convergence problems of the PC-SAFT model at high pressures, and the existence of multiple roots, as was pointed by Privat *et al.* [181]. It puts also a question mark to the solubility of solids, which were calculated with the PC-SAFT equation of state, since the solute probably has a packing fraction exceeding 0.46. It is recommended to optimize the universal constants again, using constraints to prevent that multiple chain lengths have the same first perturbation integral result at the same packing fraction. It is also recommended to include statistical information for the obtained universal constants.

- We have improved the PC-SAFT based activity coefficient expression for the dispersion contribution by introducing the concept of molecular topology. We have shown that the Zagreb number and our proposed derivative, the dispersion number, are topology numbers which are useful to define the effective number of spheres that interact with a central sphere.

- Our dispersion number appears to be a good parameter to predict the normal boiling point of alkanes within 3 K. Current methods are at a level of accuracy of 6 K.

- We have extended the concept of topology into a group contribution method, and derived interaction parameters for the energy between methyl groups. (Chapter 7)

- A new way to determine the effective contact area in the COSMO-RS family of models was introduced. Rotational hindrance between two molecules and differences between atoms sizes determine an effective area, by which electrostatic surface charge density are averaged. (Chapter 8). It is an improvement over the earlier COSMO-RS(ADF) model version, but it requires that each multicomponent systems has an average area based on the number of compound types, in order to make the model Gibbs-Duhem proof. The concept of accessibility might also be solved by the topology design number, which we introduced later in the development of the PC-SAFT based activity model for dispersion.

9.2 Outlook

The aforementioned conclusions and comments give rise to new ideas to improve the current available equation of state and activity models. It also give seeds to other areas like physical chemistry, chemical engineering, polymer and pharmaceutical sciences. Here are some of these ideas.

- The applied way of topology by means of the dispersion number, D, introduced by us, is in fact a way to quantify how many segments from the surrounding account for interaction, denoted by the parameter Z_k in Eq. 6.37, and 6.38. This concept has been applied with success in the COSMOSPACE methodology (Chapter 7). The COSMOSPACE model is a particular form of the Ising model. Since the Zagreb number or our dispersion number is very similar to a Google engine, where search words are linked to other words, it seems worthwhile to investigate whether the new ideas in information technology are also applicable to the description of interaction between molecules.

- The successful introduction of the dispersion number for the normal boiling point demonstrates that QSAR methodology can be made more accurate by applying non-linear relations with positive curvature.

- The applied topology concept is comparable with the COSMO3D concept, where the effective contact area is defined by rotating a probing sphere around molecules. It seems that these two techniques are equivalent. A proof for this would be useful, since topology numbers are easy definable. The COSMO3D method, might help to clarify the use topology number D instead of the first Zagreb number $Z1$ in case of strongly branched structures and structures containing rings.

- The χ-parameter in the Flory-Huggins model is applied in many models to quantify not only solubility, but also surface tension, and the computation of self-organizing molecules by Dissipative Particle Dynamics (DPD) or Self-Consistent Field (SCF)-theory. Our study shows that interaction can not be fully characterized by a simple number. The interaction is influenced by intramolecular bonds. In other words, the χ parameter obtained from phase equilibria, should be adjusted by a topology index number, before it can be applied in molecular models. One could say the χ-parameter obtained from phase equilibria studies is a fluid-based interaction parameter value, while DPD and SCF require a molecular based value.

- The activity coefficient model of dispersion, introduced in this work, has been applied to hydrocarbon systems, in particular to linear and branched alkanes.

This can be extended to systems that have permanent dipoles and associating groups. These interactions are also present in the dipole-dipole and hydrogen bonding interaction, since the EM-field of the electrons in the permanent dipole or the associative bond will influence each other. In other words, dispersive interaction always occurs between atoms that come into contact. Of course, the level of interaction is smaller than the interaction between permanent dipoles and hydrogen bonds, but by considering dispersive interaction as base to which the other interactions can be added sets another way of creating activity models. Thus, on top of a dispersive interaction one builds the dipole-dipole and hydrogen bonding contribution, instead of separating it as has been done in the Flory-Huggins and UNIFAC family of models. This approach is already incorporated in the PC-SAFT models that include permanent dipole interaction and associations schemes. PC-SAFT requires first a more solid parameterization of the universal constants.

Bibliography

[1] Vahid A. *An efficient molecular theory and simulation methodology for explicit treatment of polarity.* PhD thesis, University of Akron, May 2012. 29

[2] D.S. Abrams and J.M. Prausnitz. Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems. *AIChE J.*, 21(1):116–127, 1975. 6, 7, 20, 45, 48, 49, 98, 208

[3] J. Ahlers and J. Gmehling. Development of an universal group contribution equation of state I. Prediction of liquid densities for pure compounds with a volume translated Peng-Robinson equation of state. *Fluid Phase Equil.*, 191(1-2):177–188, 2001. 73

[4] M.J. Anselme and A.S. Gude, M.and Teja. The critical temperature and densities of the n-alkanes from pentane to octadecane. *Fluid Phase Equil.*, 57:317–326, 1990. 83

[5] G.L. Aranovich and M.D. Donohue. A new model for lattice systems. *J. Chem. Phys.*, 105(16):7059–7063, 1996. 6, 7

[6] A.J. Ashworth. Activity coefficients of C5 to C8 hydrocarbons in squalane and dinonyl phthalate at 30C determined by a vacuum microbalance technique. *J. Chem. Soc. Faraday Trans.1*, 69:459–466, 1973. xvi, xviii, xviii, 91, 93, 116, 118, 120, 121, 122

[7] AspenTech inc., AspenPlus v8.2. 141

[8] R. A. Aziz and H.H. Chen. An accurate intermolecular potential for argon. *J. Chem. Phys.*, 67:5719–5726, 1977. 217

[9] T. Banerjee, R.K. Sahoo, S.S. Rath, R. Kumar, and A. Khanna. Multicomponent liquid-liquid equilibria prediction for aromatic extraction systems using cosmo-rs. *Ind. Chem. Eng. Res.*, 46(4):1292–1304, 2007. 141

[10] A.F.M. Barton. *CRC Handbook of solubility parameters and other cohesion parameters*. CRC Press Inc., 2nd edition, 1991. 7, 82

[11] S. Batsanov. Van der Waals radii of elements. *Inorganic Materials*, 37(9):871, 2001. 53, 55, 138

[12] A. Ben-Naim. *Solvation Thermodynamics*. Springer US, 1987. ISBN: 978-1-4757-6550-2. 15, 16

[13] C.H. Bennett. Serially deposited amorphous aggregates of hard spheres. *J. Appl. Phys.*, 43(6):2727–2734, 1972. 106

[14] S. Beret and J.M. Prausnitz. Perturbed hard-chain theory: An equation of state for fluids containing small or large molecules. *AIChE J.*, 21:1123–1132, 1975. 99

[15] C. Berro, M. Rogalski, and A. Peneloux. Excess Gibbs energies and excess volumes of 1-butanol - n-hexane and 2-methyl-1-propanol - n-hexane binary systems. *J. Chem. Eng. Data*, 27(3):352–355, 1982. xv, 61, 66

[16] C. Berro, M. Rogalski, and A. Peneloux. A new ebulliometric technique. vapour-liquid equilibria in the binary systems ethanol-n-heptane and ethanol-n-nonane . *Fluid Phase Equil.*, 8(1):55–73, 1982. 61

[17] D. Berthelot. Sur le mélange des gaz. *Comptes Rendus Acad. Sci.*, 126:1703–1855, 1898. 75, 113

[18] A. Bondi. *Physical Properties of Molecular Crystals, Liquids, and Glasses*. John Wiley, 1968. xiv, 52, 53, 54, 76, 120, 207

[19] C. Boukouvalas, N. Spiliotis, P. Coutsikos, N. Tzouvaras, and D. P. Tassios. Prediction of vapor-liquid equilibrium with the LCVM model. a linear combination of the Vidal and Michelsen mixing rules coupled with the original UNIFAC and the t-mPR equation of state. *Fluid Phase Equil.*, 92:75, 1994. 73

[20] A-J Briard, M. Bouroukba, D. Petitjean, N. Hubert, and M. Dirand. Experimental enthalpy increments from the solid phases to the liquid phase of homologous n-alkane series (c18 to c38 and c41, c44, c46, c50, c54, and c60). *J. Chem. Eng. Data*, 48:497–513, 2003. xxv, 245

[21] R. Bronneberg and A. Pfennig. Improvement of the UNIQUAC combinatorial-entropy term by adjusting the standard segment. *J. Chem. Eng. Data*, 56(4):1305–1310, 2011. 49, 141, 149

[22] R. Bronneberg and A. Pfennig. MOQUAC, a new expression for the excess Gibbs energy based on molecular orientations. *Fluid Phase Equil.*, 338:63 – 77, 2013. 6, 19, 22, 45, 49, 143

[23] BYU DIPPR Thermophysical Properties Laboratory. Design Institute for Physical Properties, 2018. xiii, xxi, xxv, 29, 83, 85, 105, 120, 228, 244, 245

[24] E. Calvo, M. Artal, J.M. Embid, I. Velasco, and S. Otin. Isothermal vapor-liquid equilibria of 1,3-dioxolane or 1,4-dioxane plus hexane or plus cyclohexane or plus ethanol mixtures at T= 308.15 K. *J. Chem. Eng. Data*, 44(2):193–196, 1999. xx, xxi, 152, 157, 259

[25] W. Cao, A. Fredenslund, and P. Rsmussen. Statistical thermodynamic model for the vicosity of pure and liquid mixtures. *Ind.Eng.Chem.Res.*, 31:2603–2619, 1992. 9

[26] N.F. Carnahan and K.E. Starling. Equation of state for nonattracting rigid spheres. *J. Chem. Phys.*, 51(2):635–636, 1969. 26

[27] W.G. Chapman, K.E. Gubbins, G. Jackson, and M. Radosz. Saft: Equation-of-state solution model for associating fluids. *Fluid Phase Equil.*, 52:31 – 38, 1989. 27

[28] W.G. Chapman, K.E. Gubbins, G. Jackson, and M. Radosz. Phase equilibria of associating fluids: Chain molecules with multiple bonding sites. *Ind. Eng. Chem. Res.*, 29:1709–1720, 1990. 72, 99

[29] W.G. Chapman, G. Jackson, and K.E. Gubbins. New reference equation of state for associating liquids. *Mol. Phys.*, 65:1057–1079, 1988. 27, 72, 99

[30] Y.C. Chiew. Percus-Yevick integral equation theory for athermal hard-sphere chains.: II. average intermolecular correlation functions. *Molecular Phys.*, 73(2):359–373, 1991. 107, 231, 232, 235

[31] J.J. Christensen, R.M. Izatt, B.D. Stitt, R.W. Hanks, and K.D.J. Williamson. Excess enthalpies of seven n-nonane + alcohol mixtures at 298.15 K. *J. Chem. Thermodynamics*, 11(11):1029–1034, 1979. xv, 62, 67

[32] M. Cismondi and J. Mollerup. Development and application of a three-parameter RK-PR equation of state. *Fluid Phase Equil.*, 232:74–89, 2005. 74, 87

[33] M. Cismondi-Duarte, M.V. Galdo, M.J. Gomez, N.G. Tassin, and M. Yanes. High pressure phase behavior modeling of asymmetric alkane + alkane binary systems with the RKPR EOS. *Fluid Phase Equil.*, 362:125–135, 2014. 87

[34] L. Constantinou and R. Gani. New group contribution method for estimating properties of pure compounds. *AIChE J.*, 40(10):1697–1710, 1994. 141

[35] L. Constantinou, R. Gani, and J.P. O'Connell. Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method. *Fluid Phase Equil.*, 103(1):11–22, 1995. 141

[36] C.J. Cramer. *Essentials of computational chemistry: theories and models.* John Wiley & Sons Ltd, England, 2nd edition, 2004. 142

[37] K.C. Das, I. Gutman, and M.J. Nadjafi-Arani. Relations between distance-based and degree-based topological indices. *Appl. Math. Comp.*, 270:142–147, 2015. 107

[38] Dortmund Data Bank GmbH software package, 2018. xvi, xvii, xviii, xviii, xix, 90, 91, 117, 119, 120, 121, 122, 137

[39] E.L. Derr and C.H. Deal. Predicted compositions during mixed solvent evaporation from resin solutions using the analytical solutions of groups method. *Amer. Chem. Soc. Div. Org. Coatings Plast. Chem.*, 31(2):1–3, 1971. 6, 7

[40] M. Diaz-Pena and D.R. Cheda. Excess volume of the system containing ethanol and n-alkanes. *Anales de Quimica*, 66:747–755, 1970. xiv, xv, 61, 64, 66

[41] A. Diedrichs, J Rarey, and J. Gmehling. Prediction of liquid heat capacities by the group contribution equation of state vtpr. *Fluid Phase Equil.*, 248(1):56–69, 2006. 141

[42] L.R. Dodd and S.I. Sandler. Practical equation of state and activity coefficient models based on the exp-6 fluid. *Fluid Phase Equil.*, 69:99–139, 1991. 27

[43] M.D. Donohue and J.M. Prausnitz. Combinatorial entropy of mixing molecules that differ in size and shape. a simple approximation for binary and multicomponent mixtures. *Can. J. Chem.*, 53:1586–1592, 1975. 6, 7, 17, 45, 72, 98

[44] M.D. Donohue and J.M. Prausnitz. Perturbed hard chain theory for fluid mixtures: Thermodynamic properties for mixtures in natural gas and petroleum technology. *AIChE J.*, 24:849–860, 1978. 99

[45] DWSIM 3.2. 141

[46] K. Egner, J. Gaube, and A. Pfennig. GEQUAC, an excess Gibbs energy model describing associating and non-associating liquid mixtures by a new model concept for functional groups. *Fluid Phase Equil.*, 158:381–389, 1999. 50

[47] H.S. Elbro, A. Fredenslund, and P. Rasmussen. New simple equation for the prediction of solvent activities in polymer solutions. *Macromolecules*, 23(21):4707–4714, 1990. 8, 23, 24

[48] J.R.E Elliott, J. Richard Jr., and R.N. Natarajan. Extension of the Elliott-Suresh-Donohue equation of state to polymer solutions. *Ind. Chem. Eng. Sci.*, 41:1043–1050, 2002. 29

[49] R. Eng and S.I. Sandler. Vapor-liquid equilibria for three aldehyde/hydrocarbon mixtures. *J. Chem. Eng. Data*, 29(2):156–161, 1984. xx, xx, 152, 155, 157

[50] Fernando A. Escobedo and Juan J. de Pablo. Extended continuum configurational bias Monte Carlo methods for simulation of flexible molecules. *J. Chem. Phys.*, 102:2636–2652, 1995. xiii, 30

[51] G.B. Flores, P.B. Staudt, and R. de P. Soares. Including dispersive interactions in the f-sac model. *Fluid Phase Equil.*, 426(Special Issue: 10th Iberoamerican Conference on Phase Equilibria and Fluid Properties for Process Design):56 – 64, 2016. 22

[52] P.J. Flory. Molecular size distribution in three dimensional polymers. *J. Am. Chem. Soc.*, 63:3083–3090, 1941. 37, 72

[53] P.J. Flory. Thermodynamics of high polymer solutions. *J. Chem. Phys.*, 10(1):51–61, 1942. 6, 149

[54] P.J. Flory. Statistical thermodynamics liquid mixtures. *J. Am. Chem. Soc.*, 87(9):1833–1838, 1965. 204

[55] P.J. Flory. Fifteenth Spiers memorial lecture. thermodynamics of polymer solutions. *Discuss. Faraday Soc.*, 49:7–29, 1970. 25

[56] C. Fonseca Guerra, J.G. Snijders, G. te Velde, and E.J. Baerends. Towards an order-n dft method. *Theoretical Chemistry Accounts*, 99(6):391–403, 1998. 151

[57] R.H. Fowler and G.S. Rushbrooke. An attempt to extend the statistical theory of perfect solutions. *Transactions of the Faraday Society*, 33:1272, 1937. 6

[58] A. Fredenslund. UNIFAC and related group-contribution models for phase equilibria. *Fluid Phase Equil.*, 52(C):135–150, 1989. 23

[59] A. Fredenslund, R.L. Jones, and J.M. Prausnitz. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChI J.*, 21(6):1086, 1975. 2, 6, 7, 45, 71, 95, 98, 129

[60] A. Fredenslund, P. Rasmussen, and J. Gmehling. *Vapor-Liquid Equilibria Using UNIFAC: A Group-Contribution Method.* Elsevier, 1977. 141, 151

[61] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, San Diego, 2nd edition, 2002. 141

[62] J.R. Fried, J.S. Jiang, and E. Yeh. Group-contribution methods in polymer thermodynamics. *Comp. Polymer Sci.*, 2:95–113, 1992. 25

[63] J. Garces. The configurational entropy of mixing of interstitials solid solutions. *Applied Phys. Letters*, 96(16):161904, 2010. 18

[64] K Geier and H.J. Bittrich. Thermodynamics of liquid-vapor-equilibrium of binary-systems, hexane-benzene, cyclohexane-benzene and tetrahydrofuran with hexane, cyclohexane, benzene and dimethyl-formamide. *Zeitschrift fur Physikalische Chemie-Leipzig*, 260(4):705–718, 1979. 152

[65] G. Geiseler and H. Kohler. Thermodynamic behavior of methyl ethyl ketoxime/n-heptane diethyl ketone/n-heptane and methyl ketoxime/diethyl ketone mixed systems. *Berichte der Bunsen-Gesellschaft für Physikalische Chemie*, 72(6):697, 1968. 152

[66] J. Gmehling. Isothermal vapor-liquid equilibriums in binary systems formed by esters with alkenes. *J. Chem. Eng. Data*, 28(1):27–30, 1983. xx, 152, 155

[67] J. Gmehling, B. Kolbe, M. Kleiber, and J Rarey. *Chemical Thermodynamics for Process Simulation.* Wiley-VCH, 2012. 33

[68] J Gmehling, U Onken, W Arlt, P Gronzheuser, U Weldlich, B Kolbe, and J Rarey. *Vapor-liquid equilibrium data collection*, volume 1. DECHEMA Chemistry Data Series, 1977. 5, 152

[69] J. Gmehling and P. Rasmussen. Flash points of flammable liquid mixtures using UNIFAC. *J. Am. Chem. Soc.*, 21(2):186–188, 1982. 9

[70] S.S. Godavarthy, R.L. Robinson Jr., and K.A.M. Gasem. SVRC-QSPR model for predicting saturated vapor pressures of pure fluids. *Fluid Phase Equil.*, 246(1-2):39–51, 2006. 141

[71] R. Goldman. Understanding quaternions. *Graphical Models /graphical Models and Image Processing /computer Vision, Graphics, and Image Processing*, 73:21–49, 2011. 146

[72] M. Goral. Vapour-liquid equilibria in non-polar mixtures. iii. binary mixtures of alkylbenzenes and n-alkanes at 313.15 K. *Fluid Phase Equil.*, 102(2):275–286, 1994. 152

[73] A.S. Gow and R.B. Kelly. Twenty-one new theoretically based cubic equations of state for athermal hard-sphere chain pure fluids and mixtures. *AIChI J.*, 61(5):1677–1690, 2015. 26

[74] H. Grensemann and J. Gmehling. Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. *Ind. Chem. Eng. Res.*, 44(5):1610–1624, 2005. 142

[75] R.D. Groot and P.B. Warrren. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.*, 107:4423–4435, 1997. 9

[76] J. Gross. *Entwicklung einer Zustandsgleichung für einfache, assoziierende und makromolekulare Stoffe*. PhD thesis, Technische Universität Berlin, 2001. 102

[77] 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 Equil.*, 168:183–199, 2000. 99, 101, 231, 233

[78] 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, 2001. xiii, xvii, xviii, xxiv, 2, 27, 29, 31, 72, 99, 102, 105, 106, 107, 116, 117, 119, 129, 153, 201, 241

[79] 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, 2002. 31

[80] E.A. Guggenheim. Statistical thermodynamics of mixtures with zero energies of mixing. *Proc. Roy. Soc. A*, 183:203–212, 1944. 37, 39

[81] E.A. Guggenheim. *Mixtures.* Clarendon Press Oxford, 1952. 6, 7, 19, 131

[82] J.J. Gutièrrez-Sevillano, T.J.H. Vlugt, K. Leonhard, J.P.J.M. v.d. Eerden, and G.J.P. Krooshof. COSMO-3D: Incorporating three-dimensional contact information into the COSMO-SAC model. *Ind. Chem. Eng. Res.*, 54(7):2214–2226, 2015. v, 6, 8, 19, 154

[83] I. Gutman, B. Ruscic, N. Trinajstic, and C.F. Wilcox. Graph theory and molecular orbitals. xii. acyclic polyenes. *J. Chem. Phys.*, 62(9):3399–3405, 1975. 107, 130

[84] C. Hansen. *Thesis: The Three Dimensional Solubility Parameter and Solvent Diffusion Coefficient and Their Importance in Surface Coating Formulation.* Copenhagen: Danish Technical Press, 1967. 6, 7, 45, 82

[85] H. Hansen-Goos. Accurate prediction of hard-sphere virial coefficients B6 to B12 from a compressibility-based equation of state. *J. Chem. Phys.*, 144(16):1 – 4, 2016. 26

[86] J.H. Hildebrand. Theory of regular solutions. *J. Am. Chem. Soc.*, 51:66–80, 1929. 105, 237

[87] J.H. Hildebrand. The entropy of solution of molecules of different size. *J. Chem. Phys.*, 15:225–228, 1947. 6, 38, 88

[88] J.H. Hildebrand. *The Solubility of Non-Electrolytes.* Dover, 1964. 82, 237

[89] J.H. Hildebrand. An improvement in the theory of regular solutions. *Proc. Nat. Acad. Sci. USA*, 76:1640–1641, 1979. 88

[90] T. Hofman, B. Barbes, and C. Casanova. Thermodynamic properties of n-alcohol–n-alkane mixtures. A comparative study of some group contribution theories. *J. Chem. Soc., Faraday Transactions*, 92(19):3565, 1996. 62

[91] T. Holderbaum and J. Gmehling. Psrk: A group contribution equation of state based on UNIFAC. *Fluid Phase Equil.*, 70(2-3):251–265, 1991. 2, 141

[92] S. Horstmann, A. Jabloniec, J. Krafczyk, K. Fischer, and J. Gmehling. PSRK group contribution equation of state: Comprehensive revision and extension IV, including critical constants and a-function parameters for 1000 components. *Fluid Phase Equil.*, 227(2):157–164, 2005. 71

[93] C.-M. Hsieh, S.-T. Lin, and J. Vrabec. Considering the dispersive interactions in the COSMO-SAC model for more accurate predictions of fluid phase behavior. *Fluid Phase Equil.*, 367:109 – 116, 2014. 98

[94] C.-M. Hsieh, S.I. Sandler, and H.M. Lin. Improvements of cosmo-sac for vapor-liquid and liquid-liquid equilibrium predictions. *Fluid Phase Equil.*, 297(1):90–97, 2010. xxv, xxv, xxv, 8, 142, 143, 148, 149, 151, 152, 155, 257, 258

[95] M.L. Huggins. Solutions of long chain compounds. *J. Chem. Phys.*, 9(5):440, 1941. 6, 37

[96] M.L. Huggins. Some properties of solutions of long-chain compounds. *J. Phys. Chem.*, 46(1):151–158, 1942. 149

[97] M.J. Huron and J. Vidal. New mixing rules in simple equations of state for representing vapor-liquid equilibria of strongly non-ideal mixtures. *Fluid Phase Equil.*, 3:255–263, 1979. 73, 98, 101

[98] P.L. Huyskens. Mobile and static molecular disorder in liquids. *J. Mol. Structure*, 274:223 – 246, 1992. 6, 18

[99] P.L. Huyskens and M.C. Haulait-Pirson. A new expression for the combinatorial entropy of mixing in liquid mixtures. *J. Mol. Liquids*, 31(3):135–151, 1985. 19, 148

[100] A. Iguchi. Vapor-liquid equilibrium at 25°C for binary systems between alcohols and hydrocarbons. *Kagaku Sochi*, 20:66–68, 1978. xiv, xv, 61, 64, 66

[101] D. Ivanizki. A generalization of the carnahan–starling approach with applications to four-and five-dimensional hard spheres. *Phys.Lett.A*, 382:1745–1751, 2018. 26

[102] J.D. Jackson. *Classical electrodynamics.* Wiley, New York, 3rd edition, 1999. 150

[103] P.L. Jackson and R.A. Wilsak. Thermodynamic consistency tests based on the Gibbs-duhem equation applied to isothermal, binary vapor-liquid equilibrium data: data evaluation and model testing. *Fluid Phase Equil.*, 103(1):155–197, 1995. 160, 201, 254

[104] W.L. Jorgensen. Optimized intermolecular potential functions for liquid alcohols. *J. Phys. Chem.*, 90(7):1276–1284, 1986. 142

[105] J. Jover, A. J. Haslam, A. Galindo, G. Jackson, and E. A. Müller. Pseudo hard-sphere potential for use in continuous molecular-dynamics simulation of spherical and chain molecules. *J. Chem. Phys.*, 137:144505–144513, 2012. xiii, 30

[106] C.F.F. Karney. Quaternions in molecular modeling. *J. Mol. Graphics & Modelling*, 25:595–604, 2007. 146

[107] K.-D. Kassmann and H. Knapp. Vapor-liquid equilibria for binary mixtures of benzaldehyde with ethylbenzene and styrene: new UNIFAC parameters. *Fluid Phase Equil.*, 33(1-2):125–136, 1987. 152

[108] A.R. Katritzky, S.H. Slavov, D.A. Dobchev, and M. Karelson. Rapid QSPR model development technique for prediction of vapor pressure of organic compounds. *Comp. & Chem. Eng.*, 31(9):1123–1130, 2007. 141

[109] H.V. Kehiaian. Thermodynamik flüssiger Mischungen von Kohlenwasserstoffen mit verwandte Substanzen. *Ber. der Bunsengesellschaft für physikalische Chemie*, 81(10):908–921, 1977. 62

[110] H.V. Kehiaian. Thermodynamics of binary liquid organic mixtures. *Pure and Applied Chemistry*, 57(1):15–30, 1985. 62

[111] J. Kestin, K. Knierim, E.A. Mason, B. Najafi, S.T. Ro, and M. Waldman. Equilibrium and transport properties of the noble gases and their mixtures at low density. *J. Phys. Chem. Ref. Data*, 13:229–303, 1984. 76, 217, 218

[112] I. Kikic, P. Alessi, P. Rasmussen, and A. Fredenslund. On the combinatorial part of the UNIFAC and UNIQUAC models. *Can. J. Chem. Eng.*, 58(2):253, 1980. 8, 22, 48, 141

[113] J.G. Kirkwood and F.P. Buff. The statistical mechanical theory of solutions. i. *J. Chem. Phys.*, 19(6):774–777, 1951. 141

[114] A. Klamt. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.*, 99(7):2224–2235, 1995. 2, 6, 8, 141, 142, 144

[115] A. Klamt. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*. Elsevier B.V., Amsterdam, 1st edition, 2005. 14, 36, 142

[116] A. Klamt, V. Jonas, T. Burger, and J.C.W. Lohrenz. Refinement and parametrization of COSMO-RS. *J. Chem. Phys. A*, 102:5074–5085, 1998. v, 71, 98, 129, 135, 141, 142

[117] A. Klamt, G.J.P. Krooshof, and R. Taylor. COSMOSPACE: Alternative to conventional activity-coefficient models. *AIChE J.*, 48(10):2332–2349, 2002. 6, 45, 50, 60, 129, 131, 134, 211

[118] A. Klamt and G. Schuurmann. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc.-Perkin Trans. 2*, 5:799–805, 1993. 142

[119] B.A. Klumov, Y. Jin, and H.A. Makse. Structural properties of dense hard sphere packings. *J. Phys. Chem. B*, 118(36):10761–10766, 2014. xvii, 103, 106, 107

[120] K. Kniaź. Influence of size and shape effects on the solubility of hydrocarbons: the role of the combinatorial entropy. *Fluid Phase Equil.*, 68:35–46, 1991. xviii, xxv, 121, 122, 247

[121] J. Kolafa, S. Labik, and A. Malijevsky. Accurate equation of state of the hard sphere fluid in stable and metastable regions. *Phys. Chem. Chem. Phys.*, 6(9):2335–2340, 2004. xiii, 30

[122] G Kolasinska, M. Goral, and J. Giza. Vapor-liquid-equilibria and excess Gibbs free-energy in binary-syste of acetone with aliphatic and aromatic-hydrocarbons at 313.15 k. *Zeitschrift fur Physikalische Chemie-Leipzig*, 263(1):151–160, 1982. xx, 152, 154

[123] C.L. Kong. Combining rules for intermolecular potential parameters II. Rules for the Lennard-Jones (12-6) potential and the morse potential. *J. Chem. Phys.*, 59:2464–2467, 1973. 75, 113, 114

[124] G.M. Kontogeorgis and G.K. Folas. *Thermodynamic Models for Industrial Applications : From Classical and Advanced Mixing Rules to Association Theories*. Wiley, 2010. 72, 82, 211

[125] G.M. Kontogeorgis, A. Fredenslund, P. Coutsikos, and D. Tassios. Improved models for the prediction of activity coefficients in nearly athermal mixtures. part i. empirical modifications of free-volume models. *Fluid Phase Equil.*, 92(C):35–66, 1994. 32

[126] G.M. Kontogeorgis, A. Fredenslund, and D.P. Tassios. Simple activity coefficient model for the prediction of solvent activities in polymer solutions. *Ind. Eng. Chem. Res.*, 32(2):362–372, 1993. 23, 48

[127] M. Kontogeorgis, G and P. Coutsikos. Thirty years with EOS/GE models - What have we learned? *Ind. Eng. Chem. Res.*, 51:4119–4142, 2012. 73

[128] G.J.P. Krooshof. Ten ways to improve UNIFAC, 2011. UNIFAC Consortium meeting. v

[129] G.J.P. Krooshof, R. Tuinier, and G. de With. Generalization of Guggenheims combinatorial activity coefficient equation. *J. Mol. Liq.*, 266:467–471, 2018. 70, 73, 80, 89, 98, 114, 129, 135

[130] G.J.P. Krooshof, R. Tuinier, and G. de With. Dispersion activity coefficient models. P art 1: Cubic equations of state. *Fluid Phase Equil.*, 2019. to be published. 113, 116, 129, 130, 131

[131] P. Krueger, S.K. Schnell, D. Bedeaux, S. Kjelstrup, T.J.H. Vlugt, and J.-M. Simon. Kirkwood-buff integrals for finite volumes. *J. Physical Chemistry Letters*, 4(2):235–238, 2013. 141

[132] J.P. Laib and D.M. Mittleman. Temperature-dependent terahertz spectroscopy of liquid n-alkanes. *J Infrared Milli Terahz Waves*, 31(9):1015–1021, 2010. xvii, 112

[133] B.L. Larsen, P. Rasmussen, and A. Fredenslund. Modified UNIFAC group-contribution model for prediction of phase equilibria and heats of mixing. *Ind. Eng. Chem. Res.*, 26(11):2274–2286, 1987. 8, 141

[134] K-H Lee and S.I. Sandler. The generalized van der Waals partition function - IV. local composition models for mixtures of unequal size molecule. *Fluid Phase Equil.*, 34:113 – 147, 1987. 98

[135] M.J. Lee, C.C. Hsiao, and H.M. Lin. Isothermal vapor-liquid equilibria for mixtures of methyl tert-butyl ether, methyl acetate, and ethyl acetate. *Fluid Phase Equil.*, 137(1-2):193–207, 1997. 152

[136] M.J. Lee and C. H. Hu. Isothermal vapor-liquid equilibria for mixtures of ethanol, acetone, and diisopropyl ether. *Fluid Phase Equil.*, 109(1):83 – 98, 1995. 152

[137] G.N. Lewis. The law of physico-chemical changes. *Proceedings of the American Academy of Arts and Sciences*, 37, 1901. 5

[138] G.N. Lewis. Outlines of a new system of thermodynamic chemistry. *Proceedings of the American Academy of Arts and Sciences*, 43, 1907. 5

[139] Ida P.C. Li, Y.-W. Wong, S.-D. Chang, and B.C.Y. Lu. Vapor-liquid equilibriums in systems n-hexane-benzene and n-pentane-toluene. *J. Chem. Eng. data*, 17(4):492–498, 1972. 152

[140] X. Liang and G.M. Kontogeorgis. New variant of the universal constants in the perturbed chain-statistical associating fluid theory equation of state. *Ind. Eng. Chem. Res.*, 54(4):1373–1384, 2015. 102

[141] S.-T. Lin, J. Chang, S. Wang, W.A. Goddard, and S.I. Sandler. Prediction of vapor pressures and enthalpies of vaporization using a COSMO solvation model. *J. Phys. Chem. A*, 108(36):7429–7439, 2004. 8, 142, 143, 150

[142] S.-T. Lin and S. I. Sandler. Henry's law constant of organic compounds in water from a group contribution model with multipole corrections. *Chem. Eng. Science*, 57(14):2727–2733, 2002. 8, 141, 142

[143] S.-T. Lin and S.I. Sandler. Infinite dilution activity coefficients from ab initio solvation calculations. *AIChE J.*, 45(12):2606–2618, 1999. 148

[144] S.-T. Lin and S.I. Sandler. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind. Eng. Chem. Res.*, 41(5):899–913, 2002. 6, 45, 141, 142, 144, 147, 148

[145] S.-T. Lin and S.I. Sandler. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind. Eng. Chem. Res.*, 41(5):899–913, 2002. 50, 71, 98, 129, 135, 143

[146] S.I. Lin, S.-T.and Sandler. Prediction of octanol-water partition coefficients using a group contribution solvation model. *Ind. Chem. Eng. Res.*, 38(10):4081–4091, 1999. 141, 148

[147] S.I. Lin, S.-T.and Sandler. Multipole corrections to account for structure and proximity effects in group contribution methods: Octanol-water partition coefficients. *J. Physical Chemistry A*, 104(30):7099–7105, 2000. 141

[148] H. Liu and Y. Hu. Molecular thermodynamic theory for polymer systems part II. Equation of state for chain fluids. *Fluid Phase Equil.*, 122(1):75 – 97, 1996. 101, 110

[149] X. Liu, A. Martin-Calvo, E. McGarrity, Sondre K. Schnell, S. Calero, J.-M. Simon, D. Bedeaux, S. Kjelstrup, A. Bardow, and T.J.H. Vlugt. Fick diffusion coefficients in ternary liquid systems from equilibrium molecular dynamics simulations. *Ind. Chem. Eng. Res.*, 51(30):10247–10258, 2012. 141

[150] X. Liu, S.K. Schnell, J.-M. Simon, P. Krueger, D. Bedeaux, S. Kjelstrup, A. Bardow, and T.J.H. Vlugt. Diffusion coefficients from molecular dynamics simulations in binary and ternary mixtures. *International J. Thermophysics*, 34(7):1169–1196, 2013. 141

[151] F. London. The general theory of molecular forces. *Trans. Faraday Soc.*, 33:8b–26, 1937. 111

[152] H.C. Longuet-Higgins. Solutions of chain molecules - a new statistical theory. *Discuss. Faraday Soc.*, 15:73–80, 1953. 16, 17

[153] H.A. Lorentz. Über die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys.*, 248:127–136, 1881. 75, 113

[154] L.M. Lozano, E.A. Montero, M.C. Martin, and M.A. Villama nan. Vapor-liquid equilibria of binary mixtures containing methyl tert-butyl ether (MTBE) and/or substitution hydrocarbons at 298.15 K and 313.15 K. *Fluid Phase Equil.*, 110(1-2):219–230, 1995. 152

[155] L.M. Lozano, E.A. Montero, M.C. Martin, and M.A. Villamañán. Isothermal vapor-liquid equilibria of binary mixtures containing methyl tert-butyl ether (mtbe) and / or substitution hydrocarbons. *Fluid Phase Equil.*, 133(1-2):155–162, 1997. International Workshop on Vapour-Liquid Equilibria and Related Properties in Binary and Ternary Mixtures of Ethers, Alkanes and Alkanols. 152

[156] H. E. Lundager Madsen and R. Boistelle. Solubility of long-chain n-paraffins in pentane and heptane. *J. Chem. Soc., Faraday Transactions 1: Phys. Chem .in Condensed Phases*, 72:1078–1081, 1976. xxi, xxv, 121, 245, 246, 247, 248

[157] Hans E. Lundager Madsen and R. Boistelle. Solubility of octacosane and hexatriacontane in different n-alkane solvents. *J. Chem. Soc., Faraday Transactions 1: Phys. Chem .in Condensed Phases*, 75:1254–1258, 1979. xxi, xxv, 121, 245, 246, 247

[158] T. Magnussen, P. Rasmussen, and A. Fredenslund. UNIFAC parameter table for prediction of liquid-liquid equilibria. *Ind. Eng. Chem. Process Des. Dev.*, 20(2):331–339, 1981. 48

[159] G.A. Mansoori, N.F. Carnahan, K.E. Starling, and T.W. Leland Jr. Equilibrium thermodynamic properties of the mixture of hard spheres. *J. Chem. Phys.*, 54(4):1523–1526, 1971. 26

[160] M. Margules. Über die Zusammensetzung der gesättigten Dämpfe von Mischungen. *Sitz.Akad.Wiss.Wien, Math.Naturw.Klasse, II*, 104:1243, 1895. 6

[161] J. Marrero and R. Gani. Group-contribution based estimation of pure component properties. *Fluid Phase Equil.*, 183-184(0):183–208, 2001. Proceedings of the fourteenth symposium on thermophysical properties. 83, 86, 141

[162] K.N. Marsh and P.P. Organ. Excess molar enthalpies and excess molar volumes for three- and four-component n-alkane mixtures simulating (n-hexane + n-hexadecane). *J. Chem. Thermodyn.*, 17:835–841, 1985. xiii, xvi, 32, 93

[163] I. Matsunaga and T. Katayama. Vapor-liquid equilibria for three binary mixtures of propionaldehyde. *J. Chem. Eng. of Japan*, 6(5):397–401, 1974. 152

[164] G. Maurer and J.M. Prausnitz. On the derivation and extension of the UNIQUAC equation. *Fluid Phase Equil.*, 2(2):91 – 99, 1978. 211

[165] C. McDermott and N. Ashton. Note on the definition of local composition. *Fluid Phase Equil.*, 1(1):33 – 35, 1977. 211

[166] M.L. Michelsen. A method for incorporating excess Gibbs energy models in equations of state. *Fluid Phase Equil*, 60:47–58, 1990. 73

[167] R.C. Miller and A.G. Williamson. Excess molar enthalpies for (n-hexane + n-hexadecane) and for three- and four-component alkane mixtures simulating this binary mixture. *J. Chem. Thermodyn.*, 16:793–799, 1984. xiii, xvi, xix, xix, 32, 33, 93, 124, 138

[168] J. Mollerop. A note on excess Gibbs energy models, equations of state and the local composition concept. *Fluid Phase Equil.*, 7:121–138, 1981. 73

[169] Daniel V. Nickel, Alejandro J. Garza, Gustavo E. Scuseria, and Daniel M. Mittleman. The isotropic molecular polarizabilities of single methyl-branched alkanes in the terahertz range. *Chemical Physics Letters*, 592:292 – 296, 2014. xvii, 112

[170] A. Nicolas, B. Fourcade, and A. Halperin. The Flory mixing free energy: The effects of solute topology and monomer size. *Macromolecular Theory and Simulations*, 9(9):755–758, 2000. 17

[171] B. Nienhaus, U. Limbeck, R. Bölts, A.B. de Haan, S-.H. Niemann, and J. Gmehling. Vapor-liquid equilibria at 413.65 K and excess enthalpies at 323.15, 363.15, and 413.15 K for mixtures of benzene, toluene, phenol, and benzaldehyde. *J. Chem. Eng. data*, 43(6):941–948, 1998. xiii, 33, 35

[172] T. Oishi and J.M. Prausnitz. Estimation of solvent activities in polymer solutions using a group-contribution method. *Ind. Eng. Chem. Process Des. Dev.*, 17(3):333–339, 1978. 23, 25, 30, 202, 203

[173] S.J. O'Shea and R.H. Stokes. Activity coefficients and excess partial molar enthalpies for (ethanol + hexane) from 283 to 318 K. *J. Chem. Thermodynamics*, 18(7):691–696, 1986. xiv, xiv, xv, xv, 61, 62, 64, 65, 66, 67

[174] J.B. Ott, K.N. Marsh, and R.H. Stokes. Excess enthalpies, excess volumes, and excess Gibbs free energies for (n-hexane + n-dodecane) at 298.15 and 308.15 K. *J. Chem. Thermodynamics*, 13:371 – 376, 1981. xiii, xix, xix, 32, 33, 124, 138

[175] G. Parisi and F. Zamponi. Mean-field theory of hard sphere glasses and jamming. *Rev. Mod. Phys.*, 82:789–845, 2010. 106

[176] Linus Pauling. *The Nature of the Chemical Bond*. Ithaca: Cornell Univ., 3rd. edition, 1960. Bond lengths used in this paper: Lhc=1.10, Lcc=1.54, Loc=1.43, Loh=0.96. 53, 76

[177] D.-Y. Peng and D.B. Robinson. New two-constant equation of state. *Ind. Eng. Chem. Fund.*, 15:59–64, 1976. 2, 73, 74

[178] B.E. Poling, J.M. Prausnitz, J.P. O'Connell, and R.C. Reid. *The properties of gases and liquids*. McGraw-Hill, New York, 5th edition, 2004. 141

[179] I. Polishuk. About the numerical pitfalls characteristic for saft eos models. *Fluid Phase Equil.*, 298(1):67 – 74, 2010. 102

[180] J.M. Prausnitz, R.N. Lichtenthaler, and E.G. de Azevedo. *Molecular Thermodynamics of Fluid-Phase Equilibria*. Prentice Hall PTR, New Jersey, 3rd edition, 1999. 141

[181] R. Privat, R. Gani, and J.-N. Jaubert. Are safe results obtained when the pc-saft equation of state is applied to ordinary pure chemicals?. *Fluid Phase Equil.*, 295(1):76 – 92, 2010. 102, 163

[182] C.C. Pye, T. Ziegler, E. van Lenthe, and J.N. Louwen. An implementation of the conductor-like screening model of solvation within the amsterdam density functional package - part ii. COSMO for real solvents. *Can. J. Chemistry*, 87(7):790–797, 2009. 151

[183] D. Ravindranath, B.J. Neely, R.L. Robinson Jr., and K. A.M. Gasem. QSPR generalization of activity coefficient models for predicting vapor-liquid equilibrium behavior. *Fluid Phase Equil.*, 257(1):53–62, 2007. 141

[184] O. Redlich and J.N.S. Kwong. On the thermodynamics of solutions. *Chem. Review*, 44:233–24, 1949. 73, 74

[185] F.H. Ree and W.G. Hoover. Fifth and sixth virial coefficients for hard spheres and hard disks. *J. Chem. Phys.*, 40:939–950, 1964. 26

[186] H. Renon and J.M. Prausnitz. Local compositions in thermodynamic excess functions for liquid mixtures. *AIChI J.*, 14(1):135, 1968. xiii, xiii, 1, 5, 6, 7, 32, 35

[187] F.A.de.L. Ribeiro and M.M.C. Ferreira. QSPR models of boiling point, octanol-water partition coefficient and retention time index of polycyclic aromatic hydrocarbons. *J. Mol. Structure: THEOCHEM*, 663(1-3):109–126, 2003. 141

[188] P. Rice and A. El-Nikheli. Isothermal vapour-liquid equilibrium data for the systems n-pentane with n-hexane, n-octane and n-decane. *Fluid Phase Equil.*, 107(2):257–267, 1995. 152

[189] M. Ronc and G.R. Ratcliff. Measurement of vapor-liquid equilibria using a semi-continuous total pressure static equilibrium still. *Can. J. Chem. Eng.*, 54(4):326, 1976. xiv, 61, 64

[190] R.S. Rowland and R. Taylor. Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals radii. *J. Phys. Chem.*, 100(18):7384–7391, 1996. 47, 53

[191] R.L. Rowley. A local composition model for multicomponent liquid mixture thermal conductivities. *Chem. Eng. Science*, 37(6):897 – 904, 1982. 9

[192] O. Sackur. Die anwendung der kinetischen theorie der gase auf chemische probleme. *Annalen der Physik*, 341(15):958, 1911. 15

[193] C. Saez, A. Compostizo, R.G. Rubio, A. Crespo Colin, and M. Díaz Peña. p, t, x, y data of benzene + n-hexane and cyclohexane + n-heptane systems. *Fluid Phase Equil.*, 24(3):241–258, 1985. xx, 152, 154

[194] S.I. Sandler. *Chemical, Biochemical, and Engineering Thermodynamics*. John Wiley & Sons, Inc., New Jersey, 4th edition, 2006. 141

[195] S.G. Sayegh and J.H. Vera. Lattice-model expressions for the combinatorial entropy of liquid mixtures: A critical discussion. *Chem. Eng. Journal*, 19(An InternationalJ. Research and Development):1 – 10, 1980. 52, 207

[196] J.M.H.M. Scheutjens and G.J. Fleer. New mixing rules in simple equations of state for representing vapor-liquid equilibria of strongly non-ideal mixtures. *J. Phys. Chem.*, 83:1619, 1979. 9

[197] S.K. Schnell, P. Englebienne, J.-M. Simon, P. Krueger, S.P. Balaji, S. Kjelstrup, D. Bedeaux, A. Bardow, and T.J.H. Vlugt. How to apply the Kirkwood-Buff theory to individual species in salt solutions. *Chem. Phys. Letters*, 582(0):154–157, 2013. 141

[198] S.K. Schnell, X. Liu, J.-M. Simon, A. Bardow, S.P., S. Kjelstrup, D. Bedeaux, A. Bardow, and T.J.H. Vlugt. Calculating thermodynamic properties from fluctuations at small scales. *J. Phys. Chem. B*, 115(37):10911–10918, 2011. 141

[199] P.D. Schreiner. Overcoming lability of extremely long alkane carbon–carbon bonds through dispersion forces. *Nature*, 477:308–311, 2011. 138

[200] Scientific computational modelling, SCM(ADF) 2011.03 COSMO-RS. 149, 150, 151

[201] G.D. Scott. Radial distribution of the random close packing of equal spheres. *Nature*, 194:956–957, 1962. 106, 110

[202] L.C. Scott, D. Ramjugermath, and J.D. Raal. Vapour-liquid equilibrium of carboxylic acid system: Propionic acid + valeric acid and isobutyric acid + valeric acid. *Fluid Phase Equil.*, 237(1):89–99, 2005. 2

[203] J.P. Shatas, M.M. Abbott, and H.C. van Ness. Excess thermodynamic functions for ternary systems ii chloroform-ethanol-n-heptane at 50 ^{circ}C. *J. Chem. Eng. Data*, 20:406–409, 1975. xv, 62, 67

[204] P.S. Snyder and J.F. Thomas. Solute activity coefficients at infinite dilution via gas-liquid chromatography. *J. Chem. Eng. Data*, 13(4):527–529, 1968. 226

[205] R. de P. Soares. The combinatorial term for COSMO-based activity coefficient models. *Ind. Chem. Eng. Res.*, 50(5):3060 – 3063, 2011. 22, 49

[206] G. Soave. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chem. Eng. Sci.*, 27:1197–1203, 1972. 2, 73, 74

[207] D. Sola, A. Ferri, M. Banchero, L. Manna, and S. Sicardi. QSPR prediction of n-boiling point and critical properties of organic compounds and comparison with a group-contribution method. *Fluid Phase Equil.*, 263(1):33–42, 2008. 141

[208] F.B. Sprow and J.M. Prausnitz. Surface thermodynamics of liquid mixtures. *Can. J. Eng. Sci.*, 45(2):26–28, 1967. 9

[209] A.J. Staverman. The entropy of high polymer solutions. generalization of formulae. *Rec. Trav. Chim. Pays-Bas*, 69(2):163–174, 1950. 6, 19, 38

[210] M. Takeo, K. Nishi, T. Nitta, and T. Katayama. Isothermal vapor-liquid equilibria for two binary mixtures of heptane with 2-butanone and 4-4-methyl-2-pentanone measured by a dynamic still with a pressure regulation. *Fluid Phase Equil.*, 3(2-3):123–131, 1979. 152

[211] Y. Tang and B. C.-Y. Lu. Direct calculation of radial distribution function for hard-sphere chains. *J. Chem. Phys*, 105:8262–8265, 1996. 232

[212] R.R. Tarakad and W.A. Scheller. Isothermal vapor-liquid equilibrium data for the system n-heptane-n-valeraldehyde at 75 and 90°c. *J. Chem. Eng. data*, 24(2):119–120, 1979. 152

[213] G. te Velde, F.M. Bickelhaupt, E.J. Baerends, C. Fonseca Guerra, S.J.A. van Gisbergen, J.G. Snijders, and T. Ziegler. Chemistry with ADF. *J. Comp. Chemistry*, 22(9):931–967, 2001. 151

[214] E.R. Thomas and C.A. Eckert. Predicting of limiting activity coefficients by modified cohesive energy density model and UNIFAC. *Ind. Eng. Chem. Process Design Dev.*, 23(2):194–210, 1984. 6, 45, 72, 98

[215] A. Tihic, G.M. Kontogeorgis, N. von Solms, and M.L. Michelsen. Applications of the simplified perturbed-chain SAFT equation of state using an extended parameter table. *Fluid Phase Equil.*, 248:29–43, 2006. xvii, xviii, xxiv, 29, 105, 106, 116, 117, 119, 120

[216] A. Tihic, N. von Solms, M.L. Michelsen, G.M. Kontogeorgis, and L. Constantinou. Analysis and applications of a group contribution sPC-SAFT equation of state. *Fluid Phase Equil.*, 281:60–69, 2009. 29

[217] H. Tompa. *Polymer Solutions*. Butterworths Scientific Publications, 1956. 38, 39

[218] L. Tonks. The complete equation of state of one, two and three-dimensional gases of hard elastic spheres. *Phys. Review*, 50:955–963, 1936. 25, 202

[219] J.D. van der Waals. Paragraph 35 of chapter 6 in [221]. 72, 237

[220] J.D. van der Waals. Paragraph 24 of chapter 4 in [221]. 72

[221] J.D. van der Waals. *Over de continuïteit van den gas- en vloeistoftoestand.* PhD thesis, University of Amsterdam, 1873. 72, 74, 186

[222] T. van Duijnen, P and M. Swart. Molecular and atomic polarizabilities Thole's model revisited. *J. Phys. Chem. A*, 102(14):2399–2407, 1998. xvii, 112

[223] J.J. van Laar. Die Dampfdruck. *Zeitschrift für Physikalische Chemie*, 72, 1910. 6, 72, 105

[224] J.H. Vera, S.G. Sayegh, and G.A. Ratcliff. A quasi lattice-local composition model for the excess Gibbs free energy of liquid mixtures. *Fluid Phase Equil.*, 1:113 – 135, 1977. 45, 49, 52, 207

[225] J. Vidal. Mixing rules and excess properties in cubic equations of state. *Chem. Eng. Science*, 33:797–791, 1978. 73, 80, 98, 101

[226] A.P.C.M. Vinhal, W. Yan, and G.M. Kontogeorgis. Application of a crossover equation of state to describe phase equilibrium and critical properties of n-alkanes and methane n-alkane mixtures. *J. Chem. Eng. Data*, 63(4):981–993, 2018. 82

[227] I.V. Vorobyov, V.M. Anisimov, and A.D. MacKerell. Polarizable empirical force field for alkanes based on the classical drude oscillator model. *The Journal of Physical Chemistry B*, 109(40):18988–18999, 2005. xvii, 112

[228] M. Waldman and A.T. Hagler. New combining rules for rare gas van der Waals parameters. *J. Comp. Chem.*, 14:1077–1084, 1993. 75, 113, 114

[229] S. Wang, S.I. Sandler, and C.-C. Chen. Refinement of cosmo-sac and the applications. *Ind. Chem. Eng. Res.*, 46(22):7275–7288, 2007. 8, 49, 142, 143, 146, 149, 150

[230] S.-J. Wang, C.-C. Yu, and H.-P. Huang. Plant-wide design and control of dmc synthesis process via reactive distillation and thermally coupled extractive distillation. *Computers & Chem. Eng.*, 34(3):361–373, 2010. 141

[231] U. Weidlich and J. Gmehling. A modified UNIFAC model. 1. prediction of vle, h^E, and γ^∞. *Ind. Eng. Chem. Res.*, 26(7):1372–1381, 1987. 8, 98, 141

[232] M.S. Wertheim. Thermodynamic perturbation theory of polymerization. *J. Chem. Phys.*, 87(12):7323–7331, 1987. 2

[233] I Wichterle and J Linek. Isothermal vapor-liquid equilibria in binary mixtures of cyclohexane or methylcyclohexane with n-alkyl (ethyl, propyl, or butyl) ethanoates (acetates) at temperatures from 335 to 370 k. *ELDATA Int. Electron. J. Phys.-Chem. Data*, 2:67–78, 1996. 152

[234] G.M. Wilson. Vapor-liquid equilibrium. xi. a new expression for the excess free energy of mixing. *J. Am. Chem. Soc.*, 86(2):127–130, 1964. 6, 7, 208

[235] H. Wolff and A. Shadiakhy. The vapor pressure behavior and association of mixtures of 1-hexanol and n-hexane between 293 and 373 K. *Fluid Phase Equil.*, 7(3-4):309–325, 1981. xv, 61, 66

[236] D-W. Wu, Y. Cui, and M.D. Donohue. Local composition models for lattice mixtures. *Ind. Eng. Chem. Res.*, 37(8):2936–2946, 1998. 7

[237] H.S. Wu and S.I. Sandler. Vapor-liquid equilibria of tetrahydrofuran systems. *J. Chem. Eng. data*, 33(2):157–162, 1988. 152

[238] A.K. Wyczalkowowska, M.A. Anisimov, and J.V. Sengers. Global crossover equation of state of a van der Waals fluid. *Fluid Phase Equil.*, 158-160:523–535, 1999. 82

[239] Xiong, S.I. Sandler, and R.I. Burnett. An improvement to COSMO-SAC for predicting thermodynamic properties. *Ind. Chem. Eng. Res.*, 53(19):8265–8278, 2014. 8, 142, 143, 144, 153

[240] Z. Xue, T. Mu, and J. Gmehling. Comparison of the a priori cosmo-rs models and group contribution methods: Original unifac, modified unifac(do), and modified unifac(do) consortium. *Industrial & Engineering Chemistry Research*, 51(36):11809–11817, 2012. 121

[241] L. Yelash, M. Müller, W. Paul, and K. Binder. Artificial multiple criticality and phase equilibria: an investigation of the pc-saft approach. *Phys. Chem. Chem. Phys.*, 7(21):3728, 2005. 102

[242] K.L. Young, R.A. Mentzer, R.A. Greenkorn, and K.C. Chao. Vapor-liquid equilibrium in mixtures of cyclohexane + benzene, + octene-1, + m-xylene, and + n-heptane. *J. Chem. Thermodynamics*, 9(10):979–985, 1977. 152

[243] S. Zhang, T. Hiaki, M. Hongo, and K. Kojima. Prediction of infinite dilution activity coefficients in aqueous solutions by group contribution models. a critical evaluation. *Fluid Phase Equil.*, 144(1-2):97–112, 1998. 141

[244] J.M. Zielinski and J.L. Duda. Predicting polymer/solvent diffusion coefficients using free-volume theory. *AIChE J.*, 38:405–414, 1992. 9

List of symbols

Symbols in Chapter 1

f_j	fugacity of compound j [bar]
f_j^0	fugacity of pure compound j at T [bar]
x_j	mole fractions of compound j in the liquid at T [-]
m	the number of spheres in a chain in PC-SAFT model [-]
y_j	mole fractions of compound j in the gas phase at T [-]

P	pressure [bar]
P_j^{sat}	saturation pressure of compound j [bar]
R	universal gas constant [J/mol.K]
T	absolute temperature [K]

γ	activity coefficient [-]
γ_{H}	enthalpy part of total activity coefficient [-]
γ_{S}	entropy part of total activity coefficient [-]
γ_{comb}	combinatorial part of activity coefficient [-]
γ_{res}	residual part of activity coefficient [-]
γ_{fv}	free-volume part of activity coefficient [-]
γ_{disp}	dispersion part of activity coefficient [-]
γ_{pol}	polar part of activity coefficient [-]
$\gamma_{\text{asso/hb}}$	association/hydrogen bonding part of activity coefficient [-]
ϵ	the dispersion energy of a sphere in PC-SAFT model [K]
ϵ_{AB}	the energy of association, i.c. hydrogen bond in PC-SAFT model [K]
κ_{AB}	the covolume of the association bond in PC-SAFT model
μ_j^0	ideal gas chemical potential [J/mol]
μ_j^{gas}	chemical potential of component j in gas phase [J/mol]
μ_j^{liq}	chemical potential of component j in liquid phase [J/mol]
μ_j^{mix}	chemical potential of component j in mixture (gas/liquid) [J/mol]
ϕ_j	fugacity coefficient of compound j in a fluid phase [-]

| σ | the average diameter of spheres in a chain in PC-SAFT model [Å] |
| χ | interaction parameter in Flory-Huggins model |

Symbols in Chapter 2

a_j	power to tune volume fraction, value between 0 and 1 [-]
d_j	diameter of a hard-sphere or chain segment [Å^3]
h	Plank constant 6.62606×10^{34} [J.s]
$\bar{h}_j^{\mathrm{E}}(x_j)$	partial molar excess enthalpy of compound j at concentration x_j [J/mol]
$h^{\mathrm{E,disp}}(x_j)$	molar excess enthalpy of mixture [J/mol]
k_{B}	Boltzmann constant $1.3806488 \times 10^{-23}$ [J/K]
m_j	molecular mass of molecule j [kg/mol]
m_k	number of segments in hard-sphere chain[-]
p_j	probability to find a particular system configuration j [-]
q_j	relative area size of molecule j [-]
r_j	relative size of molecule j [-]
x_j	mole fractions of compound j in the liquid at T [-]
v_0	reference size [Å^3]
v_j	molecular liquid volume [Å^3]
z	coordination number, value between 2 and 12 [-]

N_j	number of molecules of type j [-]
N_t	total number of molecules in mixture [-]
P_j	Gibbs probability associated with energy E_j [-]
R	universal gas constant 8.3143 [J/mol.K]
S_j	molecular entropy of molecule of type j [J/K]
S^{cEoS}	molecular entropy for cubic equation of state j [J/K]
$S^{\mathrm{FV,cEoS}}$	free-volume part of S^{cEoS} [J/K]
S^{IG}	molecular entropy of ideal gas [J/K]
T	absolute temperature [K]
T_j	absolute temperature at situation j [K]
V_j	volume of container holding compound j
W_j	free-volume fraction of compound j
Z^{cEoS}	compression factor of a cubic equation of state [-]
Z^{HS}	compression factor of Carnahan-Starling equation of state [-]
Z^{HC}	compression factor of hard-sphere chain equation of state [-]

α	mobile order parameter [-]
γ_j	activity coefficient of compound j [-]
γ_{FH}	combinatorial activity coefficient of Flory-Huggins model [-]
γ_{mFH}	modified combinatorial Flory-Huggins activity coefficient [-]
γ_{SG}	combinatorial activity coefficient of Staverman-Guggenheim model[-]
γ_{comb}	combinatorial part of activity coefficient [-]
γ_{res}	residual part of activity coefficient [-]

γ^{disp}	dispersion part of activity coefficient [-]
$\gamma^{\mathrm{FV,T}}$	free-volume part of the activity coefficient from Tonks' model [-]
$\gamma^{\mathrm{FV,cEoS}}$	free-volume part of the activity coefficient model from cEoS [-]
$\gamma^{\mathrm{FV,HS}}$	free-volume part of the activity coefficient model from hard-sphere EoS [-]
$\gamma^{\mathrm{FV,HC}}$	free-volume part of the activity coefficient from hard-sphere chain EoS [-]
$\bar{\eta}$	packing fraction of mixture [-]
η_j	packing fraction of pure component j [-]
$\tilde{\eta}$	normalized packing fraction [-]
η_∞	maximum packing fraction [-]
ϕ_j	volume fraction of compound j in a fluid phase [-]
ψ_j	fractional volume fraction of compound j in a fluid phase [-]
ρ	number density in mixture
ρ_j	number density of compound j
θ_j	surface area fraction of compound j in a fluid phase [-]

ΔG^{E}	molecular excess Gibbs energy [J]
ΔH^{E}	molecular excess enthalpy [J]
ΔS^{E}	molecular excess entropy [J/K]
$\Delta S^{\mathrm{E}}_{\mathrm{FH}}$	molecular excess entropy in Flory-Huggins model[J/K]
$\Delta S^{\mathrm{E}}_{\mathrm{mFH}}$	molecular excess entropy in Donahue-Prausnitz model[J/K]
$\Delta S^{\mathrm{E}}_{\mathrm{Gibbs}}$	molecular excess entropy in Gibbs probability model[J/K]
$\Delta S^{\mathrm{E}}_{\mathrm{Huy}}$	molecular excess entropy in Huyskens model [J/K]
$\Delta S^{\mathrm{E}}_{\mathrm{HS}}$	molecular excess entropy in hard-sphere model [J/K]
$\Delta S^{\mathrm{E}}_{\mathrm{HC}}$	molecular excess entropy in hard-sphere chain model [J/K]
ΔU_{disp}	molecular dispersion energy [J]
ΔU_{pol}	molecular dipolar energy [J]
ΔU_{AB}	molecular association energy [J]
ΔV_{FV}	molecular free-volume change [m^3]

Symbols in Chapter 3

f_{lat}	ratio in nearest neighbors for a lattice
f_{fluid}	ratio in contact area for a fluid
$r_{\text{A}}, r_{\text{B}}$	number of lattice cells of type A and B, respectively [-]
r_{tot}	number of lattice cells [-]
p_{XY}	probability that molecule segment X is neighbor of molecule segment Y
P_j	partial pressure [bar]
q	relative surface area of the largest molecule in a binary [-]
q_j	relative area size of molecule j [-]
r	relative size of the largest molecule in a binary [-]
r_j	relative size of molecule j [-]
x_j	mole fractions of compound j in the liquid at T [-]
z	coordination number of a lattice [-]
$z_{\text{A}}, z_{\text{B}}$	coordination number for segments of molecule A and B, respectively [-]

$K_{\text{gas}}, K_{\text{liq}}$	geometric constants [-]
$N_{\text{A}}, N_{\text{B}}$	number of molecules of type A and B [-]
$V_{\text{A}}, V_{\text{B}}$	molecular volume of molecule A and B, respectively [Å^3]

$\alpha_{\text{gas}}, \alpha_{\text{liq}}$	ratio in probabilities to fill a gas and liquid volume elements with only compound A or B segments [-]
γ_j	combinatorial activity coefficient [-]
γ_j^{∞}	combinatorial activity coefficient of solute j at infinite dilution [-]
$\tilde{\mu}_{\text{A}}, \tilde{\mu}_{\text{B}}$	reduced chemical potential of molecule A and B respectively [-]
$\phi_{\text{A}}, \phi_{\text{B}}$	volume fraction of compound A and B, respectively, in a fluid phase [-]
$\theta_{\text{A}}, \theta_{\text{B}}$	surface area fraction of compound A and B, respectively, in a fluid phase [-]

$\Delta\tilde{\mu}_{\text{A}}, \Delta\tilde{\mu}_{\text{B}}$	reduced chemical potential of compound A and B, respectively [-]
$\Delta\tilde{G}$	total reduced Gibbs energy [-]
$\Gamma_{\text{A}}, \Gamma_{\text{B}}$	half number of nearest neighbors [-]

Symbols in Chapter 4

n_k	total number of surface segments of molecule k in COSMOSPACE model[-]
$z.q_k$	number of nearest neighbors of molecule k in UNIQUAC model[-]
n_{OH}	number of nearest neighbors of hydroxyl group [-]
q_j	relative area size of molecule j [-]
r_j	relative size of molecule j [-]
x_j	mole fractions of compound j in the liquid at T [-]
z	coordination number, value between 2 and 12 [-]

A_{vdW}	molecular surface area [Å^2]
$H^{\text{E}}(x_j)$	excess enthalpy of mixture [J/mol]
N_{C}	carbon number of alkane or alcohol [-]
Q_j	half number of nearest neighbors of molecule j [-]
T	absolute temperature [K]
A_{vdW}	molecular surface area [Å^2]
V_{vdW}	molecular volume [Å^3]
V_j	volume of container holding compound j

γ_j^{comb}	combinatorial activity coefficient of compound j [-]
$\gamma_{\text{A},\bar{x}}$	surface coefficient of type A at concentration $\bar{x}$ [-]
$\gamma_{\text{A},k}$	surface coefficient of type A of pure compound k [-]
ϕ_j	volume fraction of compound j [-]
τ	interaction parameter in COSMOSPACE model
τ_{ij}	interaction parameter pair in UNIQUAC model [-]
θ_j	surface area fraction of compound j [-]
ω	interaction parameter in COSMOSPACE model [-]
$\zeta_j^{\text{A}}, \zeta_j^{\text{B}}$	fraction of surface of type A and B on compound j [-]

ΔE_{298}	interaction energy at 298.15 K [J/mol]
$\Delta U_{i,j}$	interaction energy pair in UNIQUAC model[J/mol]

Symbols in Chapters 5

9.3   List of symbols

β	empirical coefficients in linear relation for covolume [-]
γ	activity coefficient [-]
ϵ_{min}	Lennard-Jones potential energy [J/mol]]
η	packing fraction [-]
κ	constant in cEoS [-]
λ_1, λ_2	cEoS parameters [-]
ϕ_j, ψ_j	volume fractions [-]
σ_{ij}	diameter of segment pair i and j [Å]
θ_j	surface area fractions [-]
Δ_j	dimensionless Hildebrand parameter [Å]
Ω_a, Ω_b	cEoS constants [-]

a, a_j, a_{ij}	attractive cEoS term, pure compound j compound pair i, j [Jm^3/mol^2]
b	covolume of system [m^3/mol]
b_{ij}	covolume of molecule pair i, j [m^3]
$b_{\mathrm{C}}, b_{\mathrm{H}}$	partial covolume calculated by carbon and hydrogen atom [Å^3]
h_j^{E}	excess enthalpy of compound j [J/mol]
H_{mix}	enthalpy change on mixing [J/mol]
k_{jk}, l_{jk}	correction parameter in size and energy mixing rule[-]
m_j	number of repeating units in compound j given by
	ratio of critical volume of compound and methane [-]
$\overline{m}$	molar average number of units in mixture [-]
n_j	number of molecules of type j [-]
n_{tot}	total number of molecules in fluid mixture [-]
p_j	number of primary groups in molecule [-]
P_C, P_H	external contact probability of carbon and hydrogen atoms in a fluid[-]
r	ratio in size [-]
s_j	number of secondary groups in molecule [-]
t_1, t_2	emperical GC parameters [-]
x_j	mole fraction of compound j in mixture [-]

K	parameter in activity coefficient eq. and cEoS [-]
N_{C}	carbon (alkane) number [-]
P	system pressure [Pa]
v	system molar volume [m^3/mol]
T	system temperature [K]
T_0	reference temperature 298.15 [K]
$P_{\mathrm{c},j}$	critical pressure of compound j [Pa]
R	universal gas constant: 8.3143 [J/mol.K]

$V_{c,j}$ critical volume of compound j [m^3/mol]
$T_{c,j}$ critical temperature of compound j [K]
$Z_{c,j}$ critical compression factor of compound j [-]

Symbols in Chapters 6

α	dispersion coefficient [Å^3]
γ	activity coefficient [-]
ϵ_0, ϵ_1	energy parameters of Eq. 6.28 [K]
ϵ_{CHx}	energy increment of CHx group [K]
ϵ_{ij}	interaction energy between segments i and j [K]
$\tilde{\epsilon}_{\text{ij}}$	reduced energy (Eq. 6.13) [-]
η	packing fraction [-]
ν_0	characteristic frequency [1/s]
ϕ_j	volume fractions based on σ_j used in PC-SAFT model[-]
$\hat{\phi}_j$	volume fractions based on d_j used in IPC model [-]
ν_{CHx}	number of CHx fragment in molecule [-]
ρ	mumber density [Å^{-3}]
σ_{ij}	diameter of segment pair i and j [Å]
θ_j	surface area fractions based on σ_j (PC-SAFT) or d_j (IPC) [-]
$\hat{a}_j$	chain-length dependent parameter of I_1 [-]
$\hat{a}_{0,\text{j}}, \hat{a}_{1,\text{j}}, \hat{a}_{2,\text{j}}$	constants in parameter function $\hat{a}_j$ [-]
$\tilde{a}_1, \tilde{a}_2, \tilde{a}_3$	PC-SAFT perturbation integral I_1 constants [-]
a^{disp}	Helmholtz energy [J.mol^{-1}]
$\tilde{a}_1, \tilde{a}_2, \tilde{a}_2$	PC-SAFT activity model constant [-]
d_j	diameter of segment j [Å]
h	Planck constant, $6.62607015 \times 10^{-34}$ [J.s]
h_j^{E}	excess enthalpy of compound j [J.mol^{-1}]
$\Delta H_{\text{mix}}^{\text{E}}$	enthalpy of mixing [J.mol^{-1}]
k_{B}	Boltzmann constant, 1.380649×10^{-23} [J/mol]
k_{jk}	correction parameter in size combining rule [-]
l_{jk}	correction parameter in energy combining rule[-]
m_j	number of repeating units in compound j [-]
$m_{\text{v,j}}$	number of units with j connections in compound j [-]
$\overline{m}$	molar average number of units in mixture [-]
n_j	number of molecules of type j [-]
n_{tot}	total number of molecules in fluid mixture [-]
x_j	mole fraction of compound j in mixture [-]
C_1	PC-SAFT EoS function used in I_2 [-]
D	dispersion topology descriptor (Eq. 6.22) [-]
I_1, I_2	first and second perturbation integral [-]
J_{QH}	quadratic hydrogen number [-]
N_{C}	carbon (alkane) number [-]
N_{s1}	number of spheres in direct contact with central sphere [-]
N_{s2}	number non-contacting spheres within range of interaction[-]
P	system pressure [Pa]

R_0	distance between polarizable systems
R	universal gas constant: 8.3143 [J.mol^{-1}K^{-1}]
T	system temperature [K]
T_o	reference temperature 298.15 [K]
R	universal gas constant: 8.3143 [J.mol^{-1}K^{-1}]
U	London dispersion energy [J/mol]
V	system molar volume [m^3.mol^{-1}]
ZM_1	topological descriptor: first Zagreb index [-]
Z	total compression factor in PC-SAFT [-]
Z^{hc}	hard-sphere chain compression factor in PC-SAFT [-]
Z_0	coordination number for sphere $m = 1$ [-]
Z_j	coordination number of compound j [-]
Z_∞	coordination number for infinite chains $m >> 1$ [-]

Symbols in Chapter 7

γ_j	activity coefficient of a component in a mixture [-]
$\hat{\gamma}_j$	activity coefficient of a group in a pure compound or a mixture [-]
ϵ_0, ϵ_1	energy parameters of IPC activity model [K]
ϵ_{CHx}	energy increment of CHx group [K]
ϵ_{ij}	interaction energy between segments i and j [K]
$\tilde{\epsilon}_{\mathrm{ij}}$	reduced energy (Eq. 6.13)[-]
ϕ_j, ψ_j	volume fractions [-]
$\hat{\phi}_j$	volume fractions [-]
ψ_j	group fractions [-]
ν_{CHx}	number of CHx fragment in molecule [-]
d_j	diameter of segment j [Å]
h_j^{E}	excess enthalpy of compound j [J/mol]
$h_{\mathrm{mix}}^{\mathrm{E}}$	enthalpy of mixing [J/mol]
m_j	number of repeating units in compound j [-]
$m_{\mathrm{v,j}}$	number of units with j connections in compound j [-]
x_j	mole fraction of compound j in mixture [-]
D	topology number of as molecule (Eq. 7.3) [-]
D_j	topology number of an atom group [-]
J_{QH}	quadratic hydrogen number [-]
N_{C}	carbon (alkane) number [-]
N_{s1}	number of spheres in direct contact with central sphere [-]
P	system pressure [Pa]
Q_j	average number of local interacting spheres of compound j [-]
T	system temperature [K]
T_{o}	reference temperature 298.15 [K]
R	universal gas constant: 8.3143 [J^{-1}molK^{-}1]
ZM_1	first Zagreb index of a molecule [-]

Symbols in Chapter 8

$\Gamma_{i/S}(\sigma_{\mathrm{m}})$	segment activity coefficients for segments with charge density σ_{m} [-]
λ_{uv}	distance between segments u and v [Å]
ϕ_j	volume fraction of compound j [-]
σ_{m}	smoothed surface charge density of segment m [e/Å^2]
σ_v^*	raw COSMO charge densities [e/Å^2]
θ	angle of contact [rad]
A_j	total inner surface area of COSMO cavity embedding compound j [Å^2]
a_{eff}	effective (model reference) contact area [Å^2]
a_v	raw COSMO area of surface segment v [Å^2]
a_{uv}^{θ}	contact area for segment pair (u, v) at angle θ [Å^2]
A_{ES}	parameters [J.Å^2/mol.e/)]
B_{ES}	parameters [J.Å^2/mol.K.e/)]
b_{es}	COSMO-3D parameter (=1.41) [-]
c_{ES}	COSMOSAC/COSMO-3D universal energy parameter [J/mol/(e/Å^2)]
f_{pol}	COSMO-3D polarization factor (=1.73) [-]
f_{d}	decay parameter of filter function [-]
$E(\sigma, \sigma')$	electrostatic interaction energy [J/mol]
N_{θ}	number of different configurations [-]
N_{S}	number of segments [-]
$P(\sigma_{\mathrm{m}})$	sigma-profile [-]
r_{ij}	distance between atom centers [Å]
R_j	radius of cavity around atom j [Å]
R_3	probe radius (0.5 Å) in COSMO-3D procedure [Å]
x_j	mole fraction of compound j [-]

Appendix A

Free-volume model notes

Here we make some notes related to free-volume models and the use in activity coefficient models.

A.1 Note on the derivative of the packing fractions

The derivative of the packing fraction of a mixture towards the mole fraction of a component is an important factor in the derivation of the free-volume contribution. The packing fraction of a mixture, η, can be equated as:

$$\eta = \rho \sum_{j=1}^{M} x_j b_j = \frac{\sum_{j=1}^{M} x_j b_j}{\sum_{j=1}^{M} x_j v_j} \tag{A.1}$$

where ρ is the average number density (mol/m^3), b_j is the hard-core volume (m^3/mol) of molecule j, and v_j the liquid volume (m^3/mol). By differentiation of Eq. A.1 towards the mole fraction we get:

$$\left(\frac{\partial \eta}{\partial x_k} \right)_{T,P,x_j \neq x_k} = \rho b_k - \eta \frac{\phi_k}{x_k} = \frac{\phi_k}{x_k} \left(\eta_k - \eta \right) \tag{A.2}$$

In the SAFT model [103] and the PC-SAFT [78] we have for the packing fraction

---

201

of a mixture and a pure component consisting of equal sized spheres:

$$\eta = \frac{\pi d^3}{6} \frac{\sum_{j=1}^{M} x_j m_j}{\sum_{j=1}^{M} \frac{x_j}{\rho_j}} = \frac{\pi d^3}{6} \rho \bar{m}$$

$$\eta_k = \frac{\pi d^3}{6} \rho_k m_k, \tag{A.3}$$

where d is the diameter of the spheres, m_j is the number of spheres of a molecule of type j, x_j is the liquid mole fraction in the mixture, ρ_j is the number density of the pure compound and $\bar{m}$ the average number of segments in the chains. The ratio of the two equation gives the useful relation:

$$\frac{\eta_k}{\eta} = \frac{\rho_k}{\rho} \frac{m_k}{\bar{m}}. \tag{A.4}$$

Differentiation of the average packing factor with respect to the mole fraction of compound k yields:

$$\frac{\partial \eta}{\partial x_k} = \frac{\pi d^3}{6} \rho m_k - \eta \frac{\rho}{\rho_k} = \eta \left(\frac{m_k}{\bar{m}} - \frac{\rho}{\rho_k} \right) = (\eta_k - \eta) \frac{\phi_k}{x_k}, \tag{A.5}$$

where the liquid volume fraction is defined by

$$\phi_k = \frac{x_k v_k}{\sum_{j=1}^{M} x_j v_j} = \frac{\rho}{\rho_k} x_k. \tag{A.6}$$

A.2 Tonks EoS

The compressibility factor of the Tonks (T) equation of state [218], which holds for spherical molecules, is:

$$Z^T = \frac{1}{1 - \left(\frac{\eta}{\eta_\infty} \right)^{\frac{1}{3}}} = \frac{1}{1 - \tilde{\eta}^{\frac{1}{3}}}, \tag{A.7}$$

where we used the relative packing factor $\tilde{\eta} = \eta/\eta_\infty$, so that the derived activity coefficient expression can be compared later to the expression of Oishi and Prausnitz [172], which was used unaltered in further developments of polymer-solvent activity coefficient models. Since $Z^{IG} = 1$, we get for a pure compound

$$\frac{S^{FV,T}}{N_t k_B} = - \int_0^{\tilde{\eta}} \frac{1}{\tilde{\eta}} \left[\frac{\tilde{\eta}^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}} \right] d\tilde{\eta} = 3 \ln \left(1 - \tilde{\eta}^{\frac{1}{3}} \right). \tag{A.8}$$

We note that this result is in line what we obtained for the van der Waals volume. With this result we obtain for the entropy change upon mixing

$$\frac{\Delta S_T^{\mathrm{E}}}{k_{\mathrm{B}}} = -\sum_{j=1}^{M} 3N_j \ln \frac{\left(1 - \tilde{\eta}_j^{\frac{1}{3}}\right)}{\left(1 - \tilde{\eta}^{\frac{1}{3}}\right)}. \tag{A.9}$$

The activity coefficient follows from Eq. C.19

$$\ln \gamma_k^T = -\left(\frac{\partial \frac{\Delta S^E}{k_B}}{\partial N_k}\right)_{P,T,N_j \neq N_k} = 3 \ln \left(\frac{1 - \tilde{\eta}_k^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}}\right) + \frac{\tilde{\eta}^{-\frac{2}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}} \sum_{j=1}^{M} x_j \frac{\partial \tilde{\eta}}{\partial x_k}$$

$$= 3 \ln \left(\frac{1 - \tilde{\eta}_k^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}}\right) + \frac{\tilde{\eta}^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}} \frac{\tilde{\eta}_k}{\tilde{\eta}}. \tag{A.10}$$

The Tonk activity coefficient is only valid for spherical molecules. For poly-atomic molecules the free-volume contribution for each compound j is adjusted by a factor c_j. In Eqs. A.8,A.9 and Eq. A.10 this means that the numerical factor 3 is replaced by $3c_j$ and $3c_k$, respectively. The product $3c_k$ reflects the number of freedom that molecule k has. A molecule that moves in the fluid by translation in 3 independent directions has $c_k = 1$, while a molecule that also rotates around 3 independent molecular axis has $c_k = 2$. Vibration of the molecule in the pure liquid could be different from that in a mixture, but, in general, this is considered to be negligible in the free-volume contribution. In the model of Oishi and Prausnitz [172] the value is set to 1.1 for all molecules. When we replace in Eq. A.8 the factor 3 into $3c_j$ we get for the entropy:

$$\frac{\Delta S_T^{\mathrm{E}}}{k_{\mathrm{B}}} = -3 \sum_{j=1}^{M} c_j N_j \ln \frac{\left(1 - \tilde{\eta}_j^{\frac{1}{3}}\right)}{\left(1 - \tilde{\eta}^{\frac{1}{3}}\right)}, \tag{A.11}$$

where we assumed that the degrees of freedom of a molecule stay the same upon mixing. The free-volume contribution to the activity coefficient of a poly-atomic molecule is therefore

$$\ln \gamma_k^{\mathrm{TF}} = 3c_k \ln \left(\frac{1 - \tilde{\eta}_k^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}}\right) + \sum_{j=1}^{M} c_j x_j \frac{\tilde{\eta}^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}} \left(\frac{\eta_k}{\tilde{\eta}} - 1\right) \frac{\phi_k}{x_k}$$

$$= 3c_k \ln \left(\frac{1 - \tilde{\eta}_k^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}}\right) + \bar{c} \frac{\tilde{\eta}^{\frac{1}{3}}}{1 - \tilde{\eta}^{\frac{1}{3}}} \left(\frac{\eta_k}{\tilde{\eta}} - 1\right) \frac{\phi_k}{x_k}, \tag{A.12}$$

where $\bar{c}$ is the mole average degrees of freedom. This equation differs from the published equation of Oishi-Prausnitz:

$$\ln \gamma_k^{\mathrm{OP}} = 3c_k \ln \left[\frac{\eta_k^{-\frac{1}{3}} - 1}{\bar{\eta}^{-\frac{1}{3}} - 1} \right] - c_k \left[\frac{\frac{\bar{\eta}}{\bar{\eta}_k} - 1}{1 - \eta_k^{\frac{1}{3}}} \right]. \tag{A.13}$$

We see for example that there is no average degrees of freedom in the second term. Oishi and Prausnitz refer to a work of Flory [54], who uses Eq. A.7, and obtains for the entropy change:

$$\frac{\Delta S_F^{\mathrm{E}}}{k_{\mathrm{B}}} = -3 \sum_{j=1}^{M} c_j N_j \ln \left[\frac{1 - \tilde{\eta}_j^{-\frac{1}{3}}}{1 - \tilde{\eta}^{-\frac{1}{3}}} \right]. \tag{A.14}$$

This implies that the entropy of a pure compound j is:

$$\frac{S^{FV,F}}{k_{\mathrm{B}}} = 3c_j \ln \left[1 - \tilde{\eta}_j^{-\frac{1}{3}} \right], \tag{A.15}$$

but this equation has singular point at $\eta = 0$, while it should go to zero in the ideal gas limit ($\eta = 0$). It seems that already in the Flory entropy definition exists an error.

The example calculation of 10 vol% benzene in isobutylene gives $\tilde{\eta}_{\mathrm{benzene}} = 0.69$ and $\tilde{\eta} = 0.84$. This gives:

$$\ln \gamma_{\mathrm{Benzene}}^{\mathrm{OP}} = 2.55 - 2.02 = 0.53 \tag{A.16}$$

Applying Eq. A.10 and using $c_1 = c_2 = 1.1$ gives

$$\ln \gamma_{\mathrm{Benzene}}^{\mathrm{T}} = 2.83 + 0.33 = 2.67 \tag{A.17}$$

This is a big difference.

Appendix B

The number of nearest neighbors

In this appendix we demonstrate the usefulness of Eq. 4.3 by applying it to particular geometric cases.

Chains of repeating units in a lattice

Lattices are in theory made out of space-filling structures. Table B.1 summarizes the volume and area of some space-filling structures as function of the number of units (N) in the chain structure. We also included the case of a chain of touching spheres. In these expressions L is the length of a side of the regular polyhedron, and D the

Table B.1: Volume and surface area of a chain of structures.

Repeating unit	Volume	Area
Tetrahedron	$N\frac{6}{\sqrt{2}}L^3$	$(N+1)\frac{\sqrt{3}}{2}L^2$
Cube	NL^3	$(2N+1)2L^2$
Dodecahedron	$N\frac{15+7\sqrt{5}}{4}L^3$	$(5N+1)\frac{3}{6}\sqrt{25+10\sqrt{5}}L^2$
Touching spheres	$N\frac{\pi}{6}D^3$	$N\pi D^2$

sphere diameter. The mole, area and volume fractions of mixtures of chains of one type of structure have the general forms

$$x_{\mathrm{j}} = \frac{N_{\mathrm{j}}}{\sum\limits_{\mathrm{j}=1}^{M} N_{\mathrm{j}}}, \tag{B.1}$$

$$\theta_j = \frac{sN_j^2 + N_j}{\sum\limits_{j=1}^{M}(sN_j^2 + N_j)}, \tag{B.2}$$

and

$$\phi_j = \frac{N_j^2}{\sum\limits_{j=1}^{M} N_j^2}, \tag{B.3}$$

where the parameter s, which denotes the quadratic contribution of the surface area, is 1, 2, and 5 for respectively, a tetrahedron, a cube, and a dodecahedron. Substitution of B.1, B.2, B.3 into Eq. 4.3 gives for Q

$$Q(N_k) = 1 + sN_k, \tag{B.4}$$

which is concentration-independent. Eq. B.4 proofs that for the calculation of the combinatorial activity coefficient of a mixture of space-filling chains, of which the beads are made of tetrahedra, cubes, or dodecahedra only, the Staverman-Guggenheim correction term is needed. For a mixture of chains of touching hard-spheres, the situation is different, as in this case the ratio of volume and surface fraction is always unity; ($\theta_j/\phi_j = 1$). This implies that in Eq. 4.1 the Staverman-Guggenheim correction term becomes zero, and that only the Flory-Huggins term survives.

Linear and branched molecules

Eq. 4.3 was derived by using the connectivity relation of linear and branched structures. Here we show how Eq. 4.3 leads to expressions for the number of nearest neighbors as function of the number of repeating units. We start with a general formula for the volume and the area of a molecule containing N_k repeating groups and an end group.

$$V_k(N_k) = V_0(1 + \alpha N_k), \tag{B.5}$$

$$A_k(N_k) = A_0(1 + \beta N_k). \tag{B.6}$$

The constants, A_0 and V_0 define the size of the end group, while α and β define the ratio of the volumes of a repeating group and the chain end group. Substitution into the equations for the volume and area fraction yield for mixtures that consists of molecules of one class

$$\phi_k = \frac{x_k(1 + \alpha N_k)}{\sum\limits_{j=1}^{M} x_j(1 + \alpha N_j)}, \tag{B.7}$$

$$\theta_k = \frac{x_k(1 + \beta N_k)}{\sum\limits_{j=1}^{M} x_j(1 + \beta N_j)}. \tag{B.8}$$

As a result we obtain

$$Q_k = \frac{1 - \frac{\phi_k}{x_k}}{1 - \frac{\phi_k}{\theta_k}} = \frac{1 - \frac{(1+\alpha N_k)}{\sum\limits_{j=1}^{M} x_j(1+\alpha N_j)}}{1 - \frac{\frac{(1+\alpha N_k)}{\sum\limits_{j=1}^{M} x_j(1+\alpha N_k)}}{\frac{(1+\beta N_k)}{\sum\limits_{j=1}^{M} x_j(1+\beta N_j)}}} = \frac{\alpha}{\alpha - \beta}(1 + \beta N_k). \tag{B.9}$$

This result explains why Sayegh and Vera [195, 224] found a linear relation between the van der Waals volume of molecules and the number of nearest neighbors.

As example we evaluate the binary propane - tetradecane the calculation of the number of nearest neighbors using Eq. 4.3. According to Bondi's table [18] propane has $V_{vdW} = 37.57$ cm^3/mol and $A_{vdW} = 5.59$ cm^2/nmol, and tetradecane $V_{vdW} = 150.10$ cm^3/mol and $A_{vdW} = 20.44$ cm^2/nmol. With these values we can calculate the volume and surface area fractions, and subsequently the number of nearest neighbors for each alkane by applying Eq. 4.3. The step-by-step results are given in Table B.2.

Table B.2: Step-by-step calculation of the number of nearest neighbors of propane(1) and tetradecane(2) using Eq. 4.3.

x_1	ϕ_1	θ_1	ϕ_1/x_1	ϕ_1/θ_1	ϕ_2/x_2	ϕ_2/θ_2	Q_1	Q_2	zq_1	zq_2
0.50	0.200	0.2145	0.4004	0.9322	1.600	1.019	8.844	32.34	17.69	64.68
0.75	0.077	0.0835	0.3080	0.9218	1.231	1.231	8.844	32.34	17.69	64.68

The results of the other linear alkanes are shown in Fig. 4.1 of this work. Already with the two values from the above example we can write for the n-alkanes

$$zq_k = 64.68 + \frac{64.68 - 17.69}{14 - 3}(N_C - 14) = 4.87 + 4.27 N_C \tag{B.10}$$

This is equation 4.17 in this work. The above demonstrates that for a group of linear molecules there is a linear relation between the number of repeating groups N_k and the number of nearest neighbors Q_k, which can be derived without the definition of lattice coordination number, a reference area and volume.

Appendix C

Corrected UNIQUAC model

The original UNIQUAC model contains some inconsistencies. In order to understand this we need to go back to the concept of Wilson [234], who stated that molecules in a fluid are not distributed randomly, but tend to cluster more around molecules for which they have a higher affinity. This leads to the concept of local composition. Within this concept the first shell around a central molecule has a different composition than the average concentration of the bulk. The local mole fractions of compounds 1 and 2 around molecule 1 and those around molecule 2 are defined by the quantities x_{11}, x_{21}, x_{12} and x_{22}, respectively. A general expression for the local mole fractions x_{ij} follows from the situation where molecules are distributed randomly. This is certainly the case when the temperature is very high. We denote the temperature of the random case by T_{ra}. In the random configuration all molecules are interacting with same, "random" interaction energy U_{ra}. When we move molecule j from the random system to a system where it has a local configuration and interacts with central molecule i and energy U_{ji}, the ratio of the two mole fractions can be quantified by the energy difference as

$$x_{ji} = x_j \exp\left(-\left[\frac{U_{ji}}{RT} - \frac{U_{ra}}{RT_{ra}}\right]\right). \tag{C.1}$$

Subsequently, the ratio of local mole fraction of molecules 1 and 2 around central molecule 1 can be defined as

$$\frac{x_{21}}{x_{11}} = \frac{x_2 \exp\left(-\left[\frac{U_{21}}{RT} - \frac{U_{ra}}{RT_{ra}}\right]\right)}{x_1 \exp\left(-\left[\frac{U_{11}}{RT} - \frac{U_{ra}}{RT_{ra}}\right]\right)} = \frac{x_2}{x_1} \exp\left(-\frac{U_{21} - U_{11}}{RT}\right). \tag{C.2}$$

While Wilson implicitly assumed that all molecules have the same size, Abrams and Prausnitz [2] later took into account that the molecules are different in size and they

used surface fractions instead of mole fractions. Hence the excess energy of mixing two pure compounds was defined as

$$U^{\mathrm{E}} = \frac{z}{2} \left[N_1 q_1 \Theta_{21} (U_{21} - U_{11}) + N_2 q_2 \Theta_{21} (U_{12} - U_{22}) \right], \tag{C.3}$$

where Θ_{ji} is the local surface fraction of compound j around central molecule i in the mixture. We note that the factor $z q_{\mathrm{j}}$, i.e. the number of nearest neighbors, indicates that the quantity U_{ij} is the interaction energy between one side of compound i and one side of compound j. Further we observe that the lattice coordination number is identical for all types of molecules; $z_1 = z_2 = z$, which is set to 10 in the original UNIQUAC model. However, this approximation made by the developers of UNIQUAC is not neccesary for the residual part, as we will demonstrate below. In fact the approximation z = 10 made for the original UNIQUAC, is one of the root causes for systematic deviation in systems where molecules have a lower or higher coordination number. However, the approximation of $z = 10$ is not necessary, as we will demonstrate below.

Let us first recall the quantity that denotes half the number of nearest neighbors of molecule j. This is identical to Eq. 4.2, but now we also include the possibility that the lattice coordination number is different for each compound

$$Q_{\mathrm{j}} = \frac{z_{\mathrm{j}} q_{\mathrm{j}}}{2}. \tag{C.4}$$

We mention that summation of this quantity equals to the total number of interacting pairs. With Eq. C.4, Eq. C.3 can be written as

$$U^{E} = Q_1 N_1 \Theta_{21} (U_{21} - U_{11}) + Q_2 N_2 \Theta_{12} (U_{12} - U_{22}). \tag{C.5}$$

Here we implicitly assumed that the lattice coordination number of the pure compound is not affected by the mixture. Subsequently we define the local surface fractions. In line with Wilson's concept it means that we unpair from the random configuration one side of molecule j from one side of molecule i, this involves energy U_{ra}, and pair the two in the local configuration where molecule i is the central molecule. This local configuration requires energy U_{ji}. Applying the same steps as in the Wilson model the ratio of the local and random surface fractions can be written as

$$\frac{\Theta_{ji}}{\Theta_{ii}} = \frac{\theta_{\mathrm{j}}}{\theta_{\mathrm{i}}} \exp \left[-\frac{U_{\mathrm{ji}} - U_{ii}}{k_{\mathrm{B}} T} \right]. \tag{C.6}$$

This equation differs from the original UNIQUAC equation, where the factor $z/2$ was used in the exponent. The quantity z can not be present in the exponent, because we place one side of a lattice cell from a molecule in random configuration in contact

with one side of a lattice cell of the central molecule in the local configuration. The other cell sides of the molecule, which are also taken from the random configuration, can be placed to any other cell side, and do not take part in this calculation step! For convenience we define the interaction energy difference of unequal pairs and equal pairs on a temperature scale by

$$T_{jk} = \frac{U_{jk} - U_{kk}}{k_B}.$$
(C.7)

This gives the following two relations

$$\frac{\Theta_{21}}{\Theta_{11}} = \frac{\theta_2}{\theta_1} \exp\left[-\frac{T_{21}}{T}\right],$$
(C.8)

$$\frac{\Theta_{12}}{\Theta_{22}} = \frac{\theta_1}{\theta_2} \exp\left[-\frac{T_{12}}{T}\right].$$
(C.9)

The sum of the local surface fractions around a central molecule are unity, as imposed by the full occupancy boundary condition

$$\Theta_{21} + \Theta_{11} = 1,$$
(C.10)

and

$$\Theta_{12} + \Theta_{22} = 1,$$
(C.11)

from which follows that

$$\Theta_{21} = \frac{\theta_2 \exp\left[-\frac{T_{21}}{T}\right]}{\theta_1 + \theta_2 \exp\left[-\frac{T_{21}}{T}\right]},$$
(C.12)

and

$$\Theta_{12} = \frac{\theta_1 \exp\left[-\frac{T_{12}}{T}\right]}{\theta_2 + \theta_1 \exp\left[-\frac{T_{12}}{T}\right]}.$$
(C.13)

Substitution of Eqs. C.7, C.12 and C.13 into Eq. C.5 yields

$$\frac{U^E}{k_B} = Q_1 N_1 T_{21} \frac{\theta_2 \exp\left[-\frac{T_{21}}{T}\right]}{\theta_1 + \theta_2 \exp\left[-\frac{T_{21}}{T}\right]}$$

$$+ Q_2 N_2 T_{12} \frac{\theta_1 \exp\left[-\frac{T_{12}}{T}\right]}{\theta_2 + \theta_1 \exp\left[-\frac{T_{12}}{T}\right]}.$$
(C.14)

The Helmholtz energy is obtained from the integral

$$\frac{A^E}{k_B T} = -\int_0^{\frac{1}{T}} \frac{U^E}{k_B} d\frac{1}{T}.$$
(C.15)

Substitution of Eq. C.14 into Eq. C.15 gives

$$\frac{A^{\mathrm{E}}}{k_{\mathrm{B}}T} = -Q_1 N_1 \ln \left[\theta_1 + \theta_2 \exp \left[-\frac{T_{21}}{T} \right] \right]$$
$$-Q_2 N_2 \ln \left[\theta_2 + \theta_1 \exp \left[-\frac{T_{12}}{T} \right] \right]. \tag{C.16}$$

Generalization of this binary case to a multicomponent system results in

$$\frac{A^{\mathrm{E}}}{k_{\mathrm{B}}T} = -\sum_{j=1}^{M} Q_j N_j \ln \left[\sum_{l=1}^{M} \theta_l \tau_{lj} \right], \tag{C.17}$$

where we defined

$$\tau_{lj} = \exp \left[-\frac{U_{lj} - U_{jj}}{k_{\mathrm{B}}T} \right]. \tag{C.18}$$

Using

$$\ln \gamma_k = \frac{\partial \frac{A^{\mathrm{E}}}{k_{\mathrm{B}}T}}{\partial N_k}, \tag{C.19}$$

yields the correct residual term for the UNIQUAC equation

$$\ln(\gamma_k^{\mathrm{res,UQ}}) = Q_k \left[1 - \ln(\sum_{j=1}^{M} \theta_j \tau_{jk}) - \sum_{j=1}^{M} \frac{\theta_j \tau_{kj}}{\sum_{i=1}^{M} \theta_i \tau_{ij}} \right] \tag{C.20}$$

The crucial point in the original UNIQUAC derivation is the incorrect use of the lattice coordination number to define the ratio of the local and random surface fractions. Both Kontogeorgis and Folas [124] and Klamt *et al.* [117] mentioned that the UNIQUAC equation was inconsistent and that the exponent required $z = 2$. This value is unphysical, because in a 3-dimensional system the lowest value is $z = 4$ (i.e. a lattice made of tetrahedrons). In a reply on the comment made by McDermott and Ashton [165] also Maurer and Prausnitz [164] elaborated on the UNIQUAC inconsistency. Their solution was to add a constant of proportionality c into the denominator of the argument of the exponential in Eq. C.18. The value for c is mixture dependent, and has an average value $c = 0.27$. This implies that the factor $z/2$ in the Boltzmann factor of the original UNIQUAC model was reduced from 5 to 1.4. The deviation from unity could be caused by the choice of $z = 10$ as well as the use of a fixed reference area A_{ref} in the UNIQUAC model. But, as we have

shown here, the origin for the aforementioned inconsistency and the need to use of a proportionality factor to obtain better results lies in the incorrect definition of the ratio of the local and random surface fractions. The correct derivation leads to Eq. C.6. This also eliminates the use of a single value of $z = 10$.

Appendix D

The dimensionless solubility constant

Here we show how Eq. 5.30 can be derived. The cEoS parameters a and b can be expressed in reduced in the critical constants of the compounds by:

$$a_j = \Omega_a \frac{R^2 T_{c,j}^2}{P_{c,j}} \text{ and } b_j = \Omega_b \frac{RT_{c,j}}{P_{c,j}}, \tag{D.1}$$

where Ω_a and Ω_b are cEoS specific constants (see Table 5.3), yielding:

$$\ln \gamma^{\text{disp},\infty} = \kappa \frac{b_j}{RT} \left[\frac{\sqrt{a_j}}{b_j} - \frac{\sqrt{a_k}}{b_k} \right]^2 = \kappa \frac{b_j}{RT} \frac{\Omega_a}{\Omega_b^2} \left[\sqrt{P_{c,j}} - \sqrt{P_{c,k}} \right]^2 \tag{D.2}$$

The covolume is proportional to the critical volume by cEoS constant y:

$$b_{c,j} = yV_{c,j}. \tag{D.3}$$

We can relate the critical volumes of compound j to that of methane by:

$$V_{c,j} = m_j V_{c,1}, \tag{D.4}$$

where m_j is a proportionality constant, specific for each compound, and which can be interpreted as the number of spheres that represent the polyatomic molecule. Substitution of Eq. D.3, and D.4 into Eq. D.2 gives:

$$\kappa \frac{b_j}{RT} \left[\frac{\sqrt{a_j}}{b_j} - \frac{\sqrt{a_k}}{b_k} \right]^2 = \kappa \frac{y\Omega_a}{\Omega_b^2} \, m_j \left(\frac{T_0}{T} \right) \left[\sqrt{\frac{P_{c,j}V_{c,1}}{RT_0}} - \sqrt{\frac{P_{c,k}V_{c,1}}{RT_0}} \right]^2$$

$$= Km_j \left(\frac{T_0}{T} \right) [\Delta_j - \Delta_k]^2, \tag{D.5}$$

where we defined:

$$K = \kappa \frac{y \Omega_a}{\Omega_b^2} = \kappa \frac{\Omega_a}{\Omega_b Z_c}, \tag{D.6}$$

which is a constant determined by the cEoS parameters λ_1 and λ_2 and

$$\Delta_j = \sqrt{\frac{P_{c,j} V_{c,1}}{R T_0}}. \tag{D.7}$$

Values of Ω_a, Ω_b, κ, y, and K are provided in Table 5.3.

Appendix E

The cEoS constant K

The following general form of the cubic equation of state:

$$P = \frac{RT}{V - b} - \frac{a}{(V + \lambda_1 b)(V + \lambda_2 b)}, \tag{E.1}$$

can be rearranged into:

$$V^3 - V^2 \left[-(\lambda_1 + \lambda_2) b + b + RT/P \right]$$
$$+ V \left[\lambda_1 \lambda_2 b^2 - (\lambda_1 + \lambda_2) b^2 - RT/P (\lambda_1 + \lambda_2) b + a/P \right]$$
$$- \left(\lambda_1 \lambda_2 b^3 + RT/P \lambda_1 \lambda_2 b^2 + ab/P \right) = 0. \tag{E.2}$$

At the critical point this equation should be continuous and the first and second derivatives should vanish, which leads to the following solution at the critical point:

$$(V - V_c)^3 = V^3 - 3V_c V^2 + 3V_c^2 V - V_c^3 = 0. \tag{E.3}$$

Hence we need to solve the following relations:

$$3V_c = \left[(1 - \lambda_1 - \lambda_2) b + RT_c/P_c \right]$$
$$3V_c^2 = \left[(\lambda_1 \lambda_2 - \lambda_1 - \lambda_2) b^2 - RT_c/P_c (\lambda_1 + \lambda_2) b + a/P_c \right]$$
$$V_c^3 = \left(\lambda_1 \lambda_2 b^3 + RT_c/P_c \lambda_1 \lambda_2 b^2 + ab/P \right). \tag{E.4}$$

By introducing $y = b/V_c$ we obtain, after rearrangement, the following relations:

$$\frac{1}{Z_c} = \frac{RT_c}{P_c V_c} = \left[3 - (\lambda_1 + \lambda_2 + 1) y \right]$$
$$\frac{a}{P_c V_c^2} = 3 - (\lambda_1 + \lambda_2 + \lambda_1 \lambda_2) y^2 - \frac{1}{Z_c} (\lambda_1 + \lambda_2) y$$
$$y = \left[1 - \lambda_1 \lambda_2 y^3 - \frac{1}{Z_c} \lambda_1 \lambda_2 y^2 \right] \frac{P_c V_c^2}{a}. \tag{E.5}$$

Given values for the parameters λ_1 and λ_2, we can now solve the parameters y, Z_c, and $a/(P_c V_c^2)$ by iteration, using as starting point $y = 1/3$ in the first equation. For the van der Waals, the SRK and the PR cEoS the values for y are known exactly. These are $\frac{1}{3}$, $\frac{1}{3}\left(\sqrt{8}\sinh\left(\frac{1}{3}\mathrm{asinh}\left(\sqrt{8}\right)\right) - 1\right)$ and $\sqrt[3]{2} - 1$, respectively. The obtained values for these parameters generate the desired dimensionless energy and volume parameter:

$$\Omega_a = Z_c^2 \left(a/P_c V_c^2\right) \quad \Omega_b = yZ_c. \tag{E.6}$$

Together with the parameter κ, the parameter K can be calculated by the equation:

$$K = \kappa y \Omega_a / \Omega_b^2 = \kappa \Omega_a / (\Omega_b Z_c). \tag{E.7}$$

Appendix F

Mixing rules tested to monoatomic molecules

Kestin [111] obtained the size and energy parameter values for the noble gases by an extensive study of phase equilibria and transport properties of pure systems and of binary and multicomponent mixtures between 100 and 1500 K. In his work all the LJ parameters were scaled to those of Argon using the values reported earlier by Aziz and Chen [8]. The reported binary diameter and energy parameters can be compared to the values obtained by applying the WH, Kong and classical mixing rules. The pure component diameters of For He, Ne, Ar, Kr and Xe are 2.610, 2.755, 3.335, 3.571, and 3.885 Å, respectively, while the Lennard-Jones energies are 10.4, 42.0, 141.5, 197.8, and 274 K. Results of the unlike parameters are shown in table F.1.

Table F.1 shows that the WH mixing rules give the highest accuracy in calculating the unlike parameters of noble gases. The mixing rule of Kong is second best, while the classical mixing (BL) rule gives most deviation.

Table F.1: Comparison of Lennard-Jones size and energy mixing rules.

Pair(i,j)	Experiment[a]		Size mixing rule			Energy mixing rule		
	σ_{ij}(Å)	ϵ_{ij}(K)	BL[b]	WH[c]	Kong [d]	BL	WH	Kong
	Absolute value		Relative deviation					
He-Ne	2.755	19.44	0.0	0.1	−0.7	−7.6	−6.2	−3.3
He-Ar	3.335	30.01	2.2	0.2	−1.0	−27.9	0.3	4.7
He-Kr	3.571	31.05	3.1	0.2	−0.8	−46.1	1.0	4.7
He-Xe	3.885	29.77	4.6	0.5	0.3	−79.4	0.4	2.8
Ne-Ar	3.084	64.17	1.4	0.2	1.7	−20.2	−2.8	−6.8
Ne-Kr	3.267	67.32	1.5	−0.7	1.0	−35.4	−2.7	−10.0
Ne-Xe	3.533	67.25	2.1	−1.3	1.3	−59.6	−1.0	−13.2
Ar-Kr	3.464	165.8	0.2	0.0	0.3	−1.0	1.1	0.3
Ar-Xe	3.660	182.6	0.7	−0.1	1.3	−7.9	2.5	−1.6
Kr-Xe	3.753	225.4	0.4	0.2	0.9	−3.3	−0.1	−1.6
AAD(%)[e]			1.6	0.4	0.5	28.8	1.8	4.9

[a] Parameter values of Kestin *et al.* [111].
[b] Berthelot and Lorentz Eq. 5.4 with $k_{ij} = l_{ij} = 0$.
[c] Waldman and Hagler Eq. 5.6.
[d] Kong Eq. 5.5.
[e] Absolute Average Deviation (AAD) = 100%(1 − calc.value/exp.value).

Appendix G

Dispersion parameters of alkanes

Table G.1 shows the parameters of investigated alkanes derived from critical data (exp.) and the results obtained by the groups contribution (GC) method as was depicted in figs. 5.4 and 5.5. The relative deviation (Dev.) is calculated by $100\% \left(1 - \left(exp.value\right)/\left(GC - value\right)\right)$. M denotes Methyl, M_2 denotes dimethyl, M_3 denotes trimethyl, E denotes Ethyl and E_2 denotes diethyl as branch.

Table G.1: n-Alkanes parameters.

Dispersion model parameters						
Compound Name	N_C	m (exp.)	m (GC)	Dev. %	Δ_j (exp.)	Δ_j (GC)	Dev. %
Methane	1	1	1	0.0	0.427	0	0
Ethane	2	1.47	1.45	1.3	0.44	0.439	0.2
Propane	3	2.02	2.03	−0.4	0.411	0.41	0.1
n-Butane	4	2.58	2.61	−1.2	0.388	0.385	0.7
n-Pentane	5	3.17	3.19	−0.7	0.366	0.362	1
n-Hexane	6	3.76	3.77	−0.4	0.346	0.342	1.1
n-Heptane	7	4.34	4.35	−0.4	0.33	0.326	1
n-Octane	8	4.92	4.93	−0.2	0.314	0.311	0.8
n-Nonane	9	5.51	5.51	−0.1	0.301	0.298	0.9
n-Decane	10	6.08	6.09	−0.3	0.289	0.287	0.5
n-Undecane	11	6.68	6.67	0.0	0.278	0.277	0.4
n-Dodecane	12	7.26	7.25	0.0	0.269	0.268	0.3
n-Tridecane	13	7.86	7.83	0.2	0.258	0.259	−0.4
n-Tetradecane	14	8.41	8.41	−0.1	0.249	0.251	−0.9
n-Pentadecane	15	9.01	8.99	0.1	0.242	0.244	−0.8
n-Hexadecane	16	9.57	9.58	−0.1	0.235	0.238	−1
n-Heptadecane	17	10.14	10.16	−0.2	0.23	0.232	−0.6
n-Octadecane	18	10.75	10.74	0.0	0.224	0.226	−0.7
n-Nonadecane	19	11.35	11.32	0.3	0.219	0.22	−0.6
n-Eicosane	20	11.86	11.9	−0.3	0.214	0.216	−0.7
n-Heneicosane	21	12.47	12.48	−0.1	0.21	0.211	−0.6
n-Docosane	22	13.08	13.06	0.0	0.205	0.206	−0.7
n-tricosane	23	13.69	13.64	0.0	0.201	0.202	−0.4
n-tetracosane	24	14.30	14.22	0.0	0.197	0.197	−0.3
n-Pentacosane	25	14.80	14.8	0.0	0.194	0.194	−0.3
n-Hexacosane	26	15.41	15.38	0.0	0.19	0.191	−0.5
n-Heptacosane	27	16.02	15.96	0.0	0.187	0.187	−0.1
n-Octacosane	28	16.53	16.54	−0.0	0.183	0.184	−0.5
n-Nonacosane	29	17.13	17.12	0.0	0.181	0.181	−0.2
i-Butane	4	2.62	2.5	4.8	0.38	0.384	−1.2
i-Pentane	5	3.10	3.08	0.7	0.366	0.368	−0.7
Neopentane	5	3.11	2.96	4.6	0.356	0.373	−4.7
2-methylpentane	6	3.73	3.66	1.9	0.347	0.345	0.5
2,2-M$_2$-butane	6	3.63	3.54	2.2	0.351	0.354	−1.1

Continuation of Table G.1						
Compound Name	N_C	m (exp.)	m (GC)	Dev. %	Δ_j (exp.)	Δ_j (GC)	Dev. %
2,3-M$_2$-butane	6	3.66	3.54	3.1	0.353	0.351	0.8
3-M-pentane	6	3.73	3.66	1.9	0.352	0.345	1.8
2-M-hexane	7	4.26	4.24	0.6	0.33	0.33	−0.2
2,2,3-M$_3$-butane	7	4.03	4.01	0.5	0.342	0.345	−0.8
2,2-M$_2$-pentane	7	4.21	4.22	−0.1	0.331	0.335	−1.3
2,3-M$_2$-pentane	7	3.98	4.12	−3.6	0.34	0.345	−1.5
2,4-M$_2$-pentane	7	4.23	4.12	2.6	0.33	0.333	−0.9
3,3-M$_2$-pentane	7	4.19	4.22	−0.6	0.342	0.336	1.6
3-E-pentane	7	4.21	4.24	−0.5	3.5	0.332	1.7
3-M-hexane	7	4.09	4.24	3.5	0.334	0.338	−1.3
2-M-heptane	8	4.94	4.82	2.6	0.315	0.311	1.0
2,2,3,3-M$_4$-butane	8	4.67	4.48	4.1	0.337	0.327	3.1
2,2,3-M$_3$-pentane	8	4.42	4.59	−3.9	0.329	0.335	−1.9
2,2,4-M$_3$-pentane	8	4.74	4.59	3.2	0.319	0.322	−0.9
2,2-M$_2$-hexane	8	4.84	4.87	−0.6	0.317	0.317	−0.2
2,3,3-M$_2$-pentane	8	4.61	4.59	0.4	0.334	0.327	2.1
2,3,4-M$_2$-pentane	8	4.66	4.59	1.6	0.329	0.324	1.5
2,3-M$_2$-hexane	8	4.74	4.7	0.8	0.323	0.32	0.9
2,4-M$_2$-hexane	8	4.78	4.7	1.7	0.319	0.318	0.0
2,5-M$_2$-hexane	8	4.88	4.7	3.7	0.314	0.314	−0.1
2-M-3-E-pentane	8	4.48	4.7	−5.0	0.327	0.33	−1.0
3,3-M$_2$-hexane	8	4.49	4.71	−4.8	0.324	0.331	−2.3
3,4-M$_2$-hexane	8	4.72	4.7	0.4	0.327	0.321	1.8
3-E-hexane	8	4.61	4.82	−4.5	0.322	0.324	−0.8
3-M-3-E-pentane	8	4.61	4.71	−2.1	0.334	0.326	2.2
3-M-heptane	8	4.7	4.82	−2.5	0.318	0.321	−0.9
4-M-heptane	8	4.82	4.82	0.1	0.317	0.316	0.4
2-M-octane	9	5.48	5.23	4.5	0.303	0.3	0.8
2,2,5-M$_3$-pentane	9	5.26	5.17	1.7	0.304	0.31	−1.9
2,2,3,3-M$_4$-pentane	9	4.84	5.06	−4.4	0.33	0.326	1.0
2,2,3,4-M$_4$-pentane	9	4.96	5.06	−1.8	0.321	0.321	0.1
2,2,4,4-M$_3$-pentane	9	5.11	5.06	0.9	0.314	0.316	−0.8
2,2-M$_2$-3-E-pentane	9	5.18	5.17	0.1	0.319	0.312	2.1
2,2-M$_2$-heptane	9	5.26	5.38	−2.3	0.305	0.309	−1.3
2,3,3,4-M$_4$-pentane	9	5	5.06	−1.2	0.328	0.319	2.7
2,4,4-M$_3$-pentane	9	5.18	5.27	−1.7	0.312	0.313	−0.1

Compound Name	N_C	m (exp.)	m (GC)	Dev. %	Δ_j (exp.)	Δ_j (GC)	Dev. %
2,4-M_2-3-E-pentane	9	5.19	5.17	0.4	0.317	0.311	1.8
2,6-M_2-heptane	9	5.27	5.28	−1.8	0.302	0.308	−1.9
3,3-E_2-pentane	9	4.79	5.29	−10.3	0.326	0.326	−0.1
3-E-heptane	9	5.35	5.4	−0.9	0.308	0.304	1.1
3-M-octane	9	5.36	5.4	−0.7	0.305	0.304	0.2
4-M-octane	9	5.3	5.4	−1.9	0.305	0.306	−0.5
2-M-nonane	10	5.91	5.98	−1.2	0.29	0.293	−1
2,2,3,3-M_4-hexane	10	5.64	5.64	0.1	0.315	0.304	3.6
2,2,5,5-M_4-hexane	10	5.92	5.64	4.7	0.295	0.296	−0.4
2,2-M_2-octane	10	5.82	5.87	−0.8	0.293	0.297	−1.5
2,3-M_2-octane	10	5.83	5.86	−0.6	0.296	0.296	0.1
2,4-M_2-octane	10	5.83	5.86	−0.6	0.293	0.296	−0.8
2,5-M_2-octane	10	5.83	5.86	−0.6	0.293	0.296	−0.8
2,6-M_2-octane	10	5.83	5.86	−0.6	0.293	0.296	−0.8
2,7-M_3-octane	10	5.83	5.86	−0.6	0.291	0.296	−1.8
3,3,5-M_3-heptane	10	5.85	5.75	1.6	0.303	0.296	2.2
3-M-nonane	10	5.91	5.98	−1.2	0.293	0.293	−0.1
4-M-nonane	10	5.91	5.98	−1.2	0.293	0.293	−0.1
5-M-nonane	10	5.91	5.98	−1.2	0.293	0.293	−0.1
3-M-undecane	12	7.02	7.14	−1.6	0.272	0.273	−0.3
2,2,4,4,6,8,8-M_7-nonane	16	8.75	8.78	−0.41	0.249	0.253	−1.3
Squalane	30	16.63	16.73	−0.6	0.185	0.185	0.2
Cyclopropane	3	1.64	1.57	4.0	0.469	0.483	-3.0
Cyclobutane	4	2.12	2.1	1.2	0.445	0.452	-1.8
Cyclopentane	5	2.63	2.62	0.3	0.423	0.424	-0.2
Cyclohexane	6	3.12	3.15	-1.0	0.402	0.401	0.4
Cycloheptane	7	3.58	3.17	12	0.389	0.382	1.8
Cyclooctane	8	4.15	3.68	12	0.376	0.36	4.3
M-cyclopentane	6	3.23	4.2	-30	0.388	0.395	-2.0
1,3,5-M_3-cyclohexane	9	4.78	4.78	0.1	0.336	0.342	-2.0
i-Propylcyclohexane	9	4.7	4.74	-0.8	0.336	0.34	-1.2
n-Butylcyclopentane	9	4.89	4.91	-0.3	0.328	0.33	-0.6
n-Propylcyclohexane	9	4.83	4.85	-0.4	0.334	0.334	-0.3
n-Butylcyclohexane	10	5.41	5.43	-0.4	0.319	0.318	0.5

Continuation of Table G.1

End of Table

Appendix H

Cismondi and Mollerup cEoS

Here we derive the dispersive activity coefficient from the three-parameter cEoS of Cismondi and Mollerup.

The Gibbs energy of mixing at infinite pressure for the three-parameter cEoS of Cismondi and Mollerup is given by:

$$\frac{n_t g^{\mathrm{mix}}}{RT} = \frac{1}{RT}\left[\sum_j n_j \kappa_j \frac{a_j}{b_j} - n_t \bar{\kappa} \frac{\sum_{j,k} n_j n_k a_{jk}}{\sum_{j,k} n_j n_k b_{jk}}\right], \tag{H.1}$$

where κ_j is the parameter κ of compound j, and $\bar{\kappa}$ is the average value given by the expression:

$$\bar{\kappa} = \frac{1}{\bar{\lambda}_1 - \bar{\lambda}_2} \ln\left[\frac{1 + \bar{\lambda}_1}{1 + \bar{\lambda}_2}\right], \tag{H.2}$$

where

$$\bar{\lambda}_1 = \sum_j x_j \lambda_{1,j} \tag{H.3}$$

and

$$\bar{\lambda}_2 = \frac{1 - \bar{\lambda}_1}{1 + \bar{\lambda}_1}. \tag{H.4}$$

In here $\lambda_{1,\mathrm{j}}$ is the λ_1 parameter of compound j. Substitution of Eq. H.1 into Eq. 6.15 yields the following general expression for the activity coefficient of a two-parameter

cEoS:

$$\ln \gamma_j^{\text{disp, CM}} = \frac{1}{RT} \left[\kappa_j \frac{a_j}{b_j} - \bar{\kappa} \frac{2 \sum_k x_k a_{jk}}{\sum_{k,m} x_k x_m b_{km}} \right]$$

$$+ \frac{1}{RT} \left[\bar{\kappa} \frac{\left(2 \sum_k x_k b_{jk} - \sum_{k,m} x_k x_m b_{km} \right) \sum_{k,m} x_k x_m a_{km}}{\left(\sum_{k,m} x_k x_m b_{km} \right)^2} \right]$$

$$+ \frac{1}{RT} \left[\lambda_{1,j} \frac{\partial \bar{\kappa}}{\partial \bar{\lambda}_1} \frac{\sum_{j,k} x_j x_k a_{jk}}{\sum_{j,k} x_j x_k b_{jk}} \right], \tag{H.5}$$

where a_{jk} and b_{jk} are defined by one of the mixing rules given in this main paper, and

$$\frac{\partial \bar{\kappa}}{\partial \bar{\lambda}_1} = \frac{2 \left[(\bar{\lambda}_1)^2 + 2\bar{\lambda}_1 - 1 \right] - \left[(\bar{\lambda}_1 + 2)\bar{\lambda}_1 + 3 \right] \ln(\frac{1+\bar{\lambda}_1}{1+\bar{\lambda}_2})}{((\bar{\lambda}_1)^2 + 2\bar{\lambda}_1 - 1)^2}$$

$$= \frac{2}{(\bar{\lambda}_1 - \bar{\lambda}_2)(1 + \bar{\lambda}_1)} - \frac{\left[(\bar{\lambda}_1 + 2)\bar{\lambda}_1 + 3 \right]}{(\bar{\lambda}_1 - \bar{\lambda}_2)(1 + \bar{\lambda}_1)^2} \bar{\kappa} \tag{H.6}$$

The quantity κ and the derivative $\partial \kappa / \partial \lambda_1$ are depicted in Fig. H.1. We notice that this derivative, which starts at zero at $\lambda_1 = \sqrt{2} - 1$, is nearly constant for $1.4 < \lambda_1 < 2.5$, where it has a value of -0.05.

Eq. H.5 is quite complex. However, for the activity coefficients at infinite dilution of compound 1 dissolved in solvent 2, $x_1 = 0, x_2 = 1$, we get the expression:

$$\ln \gamma_1^{\infty,\text{CM}} = \frac{1}{RT} \left[\kappa_1 \frac{a_1}{b_1} - 2\kappa_2 \frac{\sqrt{a_1 a_2}}{b_2} + \kappa_2 \frac{b_1}{b_2} \frac{a_2}{b_2} + \lambda_{1,1} \frac{\partial \bar{\kappa}}{\partial \bar{\lambda}_1} \Big|_{\lambda_{1,2}} \frac{a_2}{b_2} \right]$$

$$= \frac{\kappa_2 b_1}{RT} \left[\left(\frac{\sqrt{a_1}}{b_1} - \frac{\sqrt{a_2}}{b_2} \right)^2 + \frac{a_1}{b_1^2} \left(\frac{\kappa_1}{\kappa_2} - 1 \right) - \frac{a_2}{b_1 b_2} \frac{\lambda_{1,1}}{\kappa_2} \frac{\partial \bar{\kappa}}{\partial \bar{\lambda}_1} \Big|_{\lambda_{1,2}} \right]. \tag{H.7}$$

The activity coefficient at infinite dilution of solute 2 in solvent 1 is found by interchanging the indices 1 and 2. These equations can also be represented in terms of the dimensionless solubility parameters. We have:

$$\ln \gamma_1^{\infty,\text{CM}} = K_2 m_1 \frac{T_0}{T} \left[(\Delta_1 - \Delta_2)^2 + \Delta_1^2 \left(\frac{\kappa_1}{\kappa_2} - 1 \right) - \Delta_2^2 \frac{m_2}{m_1} \frac{\lambda_{1,1}}{\kappa_2} \frac{\partial \bar{\kappa}}{\partial \bar{\lambda}_1} \Big|_{\lambda_{1,2}} \right], \tag{H.8}$$

and

$$\ln \gamma_2^{\infty,\text{CM}} = K_1 m_2 \frac{T_0}{T} \left[(\Delta_1 - \Delta_2)^2 + \Delta_2^2 \left(\frac{\kappa_1}{\kappa_2} - 1 \right) - \Delta_1^2 \frac{m_1}{m_2} \frac{\lambda_{1,2}}{\kappa_1} \frac{\partial \bar{\kappa}}{\partial \bar{\lambda}_1} \Big|_{\lambda_{1,1}} \right]. \tag{H.9}$$
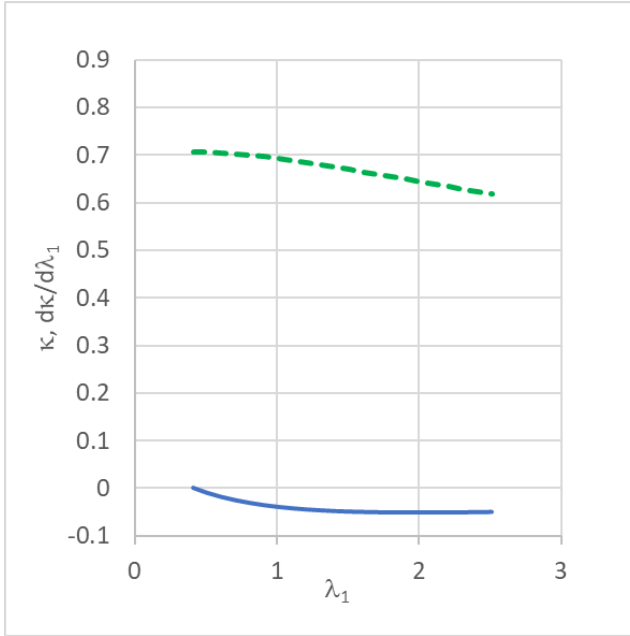
Figure H.1: The CM EoS quantity κ (Eq. 5.15, dashed green curve) and $\partial\kappa/\partial\lambda_1$ (Eq. H.6, solid blue curve).

Eqs. H.7, H.8 and H.9 reduce to the two-parameter dispersion activity coefficient model when $\lambda_{1,2} = \lambda_{1,1}$, and thus $\kappa_2 = \kappa_1$ and $K_2 = K_1$.

Lets consider a very asymmetric system in terms of the $\lambda_{1,j}$ parameters. For instance, the binary system of n-butane (1) and n-hexadecane (2). The corresponding CM parameters are, according to the work of Cismondi *et al.*, $\lambda_{1,1} = 1.40$ and $\lambda_{1,2} = 2.50$, respectively. The infinite dilution activity coefficient (IDAC) results for this case are shown in table H.1.

Table H.1: Combinatorial (FH) and dispersion (CM) IDAC at $T = T_0$.

Comp.	m_j	Δ_j	$\lambda_{1,j}$	$\lambda_{2,j}$	κ_j	K_j	$\frac{\partial\kappa_j}{\partial\lambda_{1,j}}$	$\gamma_j^{\infty,\text{FH}}$	$\gamma_j^{\infty,\text{CM}}$
Eqs.	5.24	5.26	-	5.2	H.2	D.6	H.6	6.36	H.8,H.9
n-C4	2.58	0.388	1.40	-0.167	0.675	10.6	-0.049	0.58	61
n-C16	9.75	0.235	2.50	-0.429	0.619	12.0	-0.051	0.27	24781

When the λ_1 parameters for both molecules are equal, i.e we apply the two-parameter cEoS, than we get the results of table H.2. We observe that the values for the logarithmic dispersive activity coefficient of the 3-parameter CM model are of the

Table H.2: Two-parameter example at $T = T_0$

Comp.	m_j	Δ_j	$\lambda_{1,j}$	$\lambda_{2,j}$	κ_j	K	$\gamma_j^{\infty,\text{FH}}$	$\gamma_j^{\infty,\text{CM}}$
Eqs.	5.24	5.26	-	5.2	5.15	D.6	6.36	5.25
n-C4	2.58	0.388	1.40	-0.167	0.675	10.6	0.58	14
n-C16	9.75	0.235	1.40	-0.167	0.675	10.6	0.27	18981
n-C4	2.58	0.388	2.50	-0.429	0.619	12.0	0.58	19
n-C16	9.75	0.235	2.50	-0.429	0.619	12.0	0.27	67926

same magnitude as the two-parameter cEoS model. It indicates that the dispersion activity coefficient model derived from the CM-model has the same magnitude of error as the two-parameter cEoS. Snyder and Thomas [204] measured the IDAC of butane in n-hexadecane by gas-liquid chromatography at 40, 60 and 70°C. The IDAC values are 0.85, 0.84 and 0.83, respectively. Eq. 5.37 with $K = 4.13$ for the linear alkanes gives the values 0.75, 0.73 and 72, respectively. This is 12% lower than was measured. The values of tables H.1 and H.2 are much higher due to the very high value for K.

Appendix I

Critical parameters: Experimental, mean field and crossover cEoS.

There is a difference between the experimental and theoretical critical properties (P_c, V_c, T_c) due to the fact the experimental quantities are not pure mean field properties, but also contain effects that are a result of density fluctuations. In other words, the experimental determined P_c and T_c are not representative values to define the cEoS mean field model. This explains why a regression of PVT-data at $T \leq T_c$(observed) yields always higher P_c and T_c parameters for the cEoS model values.

Crossover theory, as elaborated by Kostrowicka Wyczalkowska *et al.*, Fluid Phase Eq. 158-160 (1999) 523, is a way to incorporate density fluctuations into the mean-field models. Fig. 1 from this reference, see below as Fig. I.1, shows schematically the phase envelope of a pure compound described by the classical (dashed curve) and the crossover van der Waals equation of state. It shows that the mean field critical point (MFCP) lies above the experimental observed one (CP). Far below the critical point the density fluctuations are negligible and the models coincide. Table 2 of the same paper, see below as table I.1, shows the relative observed T_c and P_c values as function of density fluctuation parameter c_t. When $c_t = 0$ no density fluctuation is happening; the fluid behaves in a classical way and the experimental and mean field critical values would coincide. However, this situation never occurs.

For a compound where $c_t = 1$, deemed by the authors as a typical value, we see that the experimental determined critical temperature is 89.2% of the mean field value (cEoS parameter), and that the T_c/P_c ratio is 30% below the value of the T_c/P_c of the cEoS mean field model. More important to notice, is that the classical
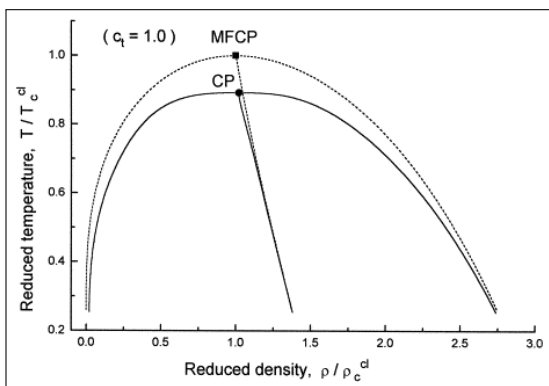
Figure I.1: Coexistence curve and coexistence–curve diameter in the reduced temperature versus reduced density plane. The solid curves represent of the crossover van der Waals equation and the dashed curves the classical van der Waals equation. The square indicates the location of the mean-field critical point (MFCP) and the circle the location of the critical point (CP) of the crossover van der Waals equation. Taken from: Fluid Phase Eq. (1999), Vol. 158-160, p523 - 535.

critical density (i.e. volume) hardly differs, only 2% at $c_t = 1$. In short, the difference between T_c/P_c as measured and as cEoS parameter explains largely the difference in $Z_c(\text{exp.})$ and $Z_c(\text{cEoS})$.

In reference to the present paper, we have 3 options to calculate the repeating unit m:

$$m = \frac{T_c}{P_c} \frac{P_c(CH_4)}{T_c(CH_4)} \text{ with experimental } T_c \text{ and } P_c \text{ (labeled exp.)}$$

$$m = \frac{T_c'}{P_c'} \frac{P_c'(CH_4)}{T_c'(CH_4)} \text{ with mean field } T_c' \text{ and } P_c' \text{ (labeled classical)}$$

$$m = V_c/V_c(CH_4) \text{ with experimental } V_c$$

We plotted these options in figure I.2 using the critical values for alkanes C1-C10 , mean field and experimental. In here we used the critical volume data of DIPPR [23], and the critical temperature and pressure data, mean field and experimental, from table 1 of the work of Vinhal *et al.* J. Chem. Eng. Data 2018, 63, 981-993, see table I.2.

As can be seen from this figure, the V_c approach for m (crosses) is more in line with the classical results (orange dots) and yields a more linear relation than the experimental T_c/P_c method (blue dots). This supports the option to calculate m by Eq. 5.24 in the present paper.

In conclusion, the difference in calculation methods for defining m is small compared to the factor 3 which is needed to bring the cEoS dispersion based activity

Table I.1: Coefficients and critical parameters as a function of c_t. Taken from: Fluid Phase Eq. (1999), Vol. 158-160, p523 - 535.

c_t [a]	$\bar{u}$ [b]	$\xi_0^+/\nu_0^{\frac{1}{3}}$ [c]	T_c $[\%$ of $T_c^{cl}]$	ρ_c $[\%$ of $\rho_c^{cl}]$	P_c $[\%$ of $P_c^{cl}]$	N_G [d]
0.2	0.02	2.2	97.7	100.4	90.9	7×10^{-5}
0.5	0.13	1.4	94.3	100.9	78.6	1×10^{-3}
1.0	0.53	1.0	89.2	101.9	62.4	9×10^{-3}
1.5	1.19	0.8	85.3	102.8	51.3	3×10^{-2}

[a] Crossover parameter.
[b] Crossover coupling constant.
[c] Relative mean-field correlation-length amplitude.
[d] Ginzburg number.
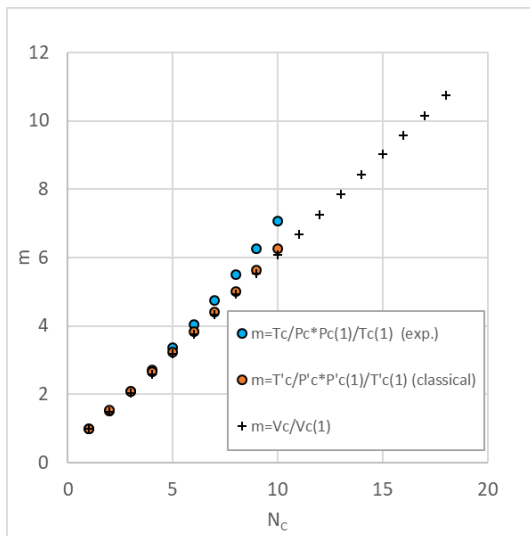
Figure I.2: Repeating unit m according to experimental T_c/P_c (exp.), mean field T_c/P_c (classical), and V_c (this work).

coefficient values in agreement with experimental results. This large difference is explained in the present paper by the cEoS assumption that molecules are spherical, while in reality most molecules are elongated.

Table I.2: Parameters of the Crossover SRK for Methane to n-Decane and Experimental Critical Temperatures, Critical Pressures, and Acentric Factor. Taken from: J. Chem. Eng. Data 2018, 63, p981-993.

substance	Mean Field			Cross-over		Experimental		
	$T_{\mathrm{c}}^{'}(\mathrm{K})$	$P_{\mathrm{c}}^{'}(\mathrm{bar})$	$\omega^{'}(-)^{\mathrm{a}}$	L(Å) [b]	$\phi(-)^{\mathrm{c}}$	$T_{\mathrm{c}}(\mathrm{K})$	$P_{\mathrm{c}}(\mathrm{bar})$	$\omega(-)^{\mathrm{a}}$
C1	197.7	50.39	−0.0578	4.166	2.5	190.6	45.99	0.0115
C2	314.3	52.35	0.0392	5.264	1.2	305.3	48.72	0.0995
C3	382.2	46.74	0.0856	5.913	0.67	369.8	42.48	0.1523
C4	441.7	42.48	0.1204	6.452	0.6	425.2	37.96	0.2002
C5	489.1	38.53	0.1632	6.879	0.6	469.7	33.7	0.2515
C6	526.4	34.98	0.2247	7.207	0.6	507.6	30.22	0.3013
C7	561.4	32.46	0.2647	7.396	0.6	540.2	27.4	0.3485
C8	592.8	30.17	0.3044	7.597	0.6	568.7	24.9	0.3995
C9	620.7	28.09	0.3434	7.809	0.6	594.6	22.9	0.4436
C10	645	26.24	0.3882	8.041	0.6	617.7	21.1	0.4923

[a] Acentric factor.

[b] Cut-off length for classical limit.

[c] Shortest wavelength parameter for density fluctuations.

Appendix J

The low density limit of I1

Here we show that in the infinite low density limit, an analytic expression can be obtained for the first perturbation integral for hard-sphere chains. The first perturbation integral is defined by

$$I_1(\eta, m) = -\int_0^\infty \tilde{u}(x)\, g^{\text{hc}}(m; x, \rho)\, x^2 dx, \qquad (J.1)$$

where x is the reduced radial center to center distance around a central hard sphere ($x = r/\sigma$), and $g^{\text{hc}}(m; x, \rho)$ is the radial distribution function of the reference fluid, which depends upon x the number of spheres m, and the system number density ρ. The reduced potential function $\tilde{u}(x) = u/\epsilon$ for spheres interacting via a square-well potential is defined as:

$$\tilde{u}(x) = \begin{cases} \infty, & x < 1 \\ -1, & 1 \le x < \lambda \\ 0, & x \ge \lambda, \end{cases} \qquad (J.2)$$

where λ is the reduced width of the square-well. In the PHSC model of Gross and Sadowski [77], $\lambda = 1.5$. The radial distribution function value for hard-sphere chains at contact (x=1) in the approximation of averaging the segment concentrations is given by the expression of Chiew [30]:

$$g(m, x = 1, \eta) = \frac{2 + \eta(3m - 2)}{2m(1 - \eta)^2}. \qquad (J.3)$$

In the limit of $\eta = 0$, this reduces to:

$$g(m, x = 1, \eta = 0) = \frac{1}{m}. \qquad (J.4)$$

This result shows that the probability of segment-segment intermolecular contact is reduced by the length of the chain. The change in the probability of finding another segment is linear with x at zero packing, as follows from the computed radial functions for molecules of different length [30]. In order to obtain an analytical expression for the slope of $g(x)$, we start with a recurrence relation for the radial distribution function, as given by Tang and Lu [211]. Their relation holds for chains having no difference between the chain middle en chain end segments. This chain average approach is known as the PY2 approximation and the expression in the $\eta = 0$ limit is given by:

$$xg(x) - x = \frac{m-1}{m}Q(x-1) + \frac{m-1}{m}\int_0^1 (x-t)\left[g(x-t) - 1\right]dt, \qquad \text{(J.5)}$$

where:

$$Q(x-1) = \frac{1}{2m}\left(x^2 - 2x\right) + \frac{m-1}{2m}(x-2). \qquad \text{(J.6)}$$

Differentiation of J.5 with respect to x, and evaluating the result at $x = 1$ and applying the Leibniz integral rule yields:

$$g(1) + g'(1) - 1 = \frac{(m-1)^2}{2m^2} + \frac{m-1}{m}\int_0^1 g(1-t) - 1 + (1-t)g'(1-t)\,dt. \qquad \text{(J.7)}$$

Since $g(r) = g'(r) = 0$ for $r < 1$, we obtain:

$$g(1) + g'(1) - 1 = \frac{(m-1)^2}{2m^2} - \frac{m-1}{m}. \qquad \text{(J.8)}$$

Substitution of Eq. J.4 into J.8 provides the expression

$$g'(1) = \frac{(m-1)^2}{2m^2}. \qquad \text{(J.9)}$$

It follows that $g'(1) = 0$ for $m = 1$, which agrees with the fact that the radial distribution function contains a discontinuity at $x = 1$ for hard spheres. The density at positions beyond the hard sphere overlap is assumed to be uniform, i.e. a mean-field approximation. For longer chains the slope of the radial distribution function increases slowly with increasing chain length to the asymptotic value of 0.5 for $m = \infty$.

The radial function for a hard sphere chain in the $\eta = 0$ limit at relative distances $1 < x < 1.5$ now follows as:

$$g(m, x, \eta = 0) = g(1) + (x-1)g'(1) = \frac{1}{m} + (x-1)\frac{(m-1)^2}{2m^2}, \qquad \text{(J.10)}$$

so that the first perturbation integral can now be written as:

$$I_1(m, \eta = 0) = \int_1^{1.5} \left[g(1) + (x - 1) g'(1) \right] x^2 dx \tag{J.11a}$$

$$= \frac{65}{64} \frac{(m-1)^2}{2m^2} + \frac{19}{24} \frac{(4m - m^2 - 1)}{2m^2} \tag{J.11b}$$

$$= \frac{19}{24} - \frac{565}{768} \frac{(m-1)}{m} + \frac{43}{768} \frac{(m-1)(m-2)}{m^2}. \tag{J.11c}$$

The last equation line (Eq. J.11c) is the form as has been used in the PHSC model, which demonstrates that the model parameters are correlated. In fact, taking $u = 19/24$ and $w = 65/64$, it follows that $\hat{a}_{0,0} = u$, $\hat{a}_{1,0} = (w - 5u)/4$, and $\hat{a}_{2,0} = (w - u)/4$. The published PHSC model parameters, which Gross *et al.* [77] obtained by fitting generated I_1 data, and which yields the $\hat{a}_{0,0}$, $\hat{a}_{1,0}$, and $\hat{a}_{2,0}$ in this paper, are compared to the exact values of Eq. J.11c in Table J.1.

Table J.1: PHSC parameters in the $\eta = 0$ limit.

Parameter	PHSC model [77]	Eq. J.11c parameters
$\hat{a}_{0,0}$	0.791982807	$19/24 \approx 0.7916667$
$\hat{a}_{1,0}$	-0.623115380	$-565/768 \approx -0.7356771$
$\hat{a}_{2,0}$	-0.067775558	$43/768 \approx 0.0559896$

Table J.2 shows the effect on $I_1(\eta = 0)$ for a monomer, dimer, tetramer and decamer. The differences in the parameters in Table J.1 and the differences in the

Table J.2: $I_1(\eta = 0)$ for different molecules.

Molecule	Eq.6.9	Eq. J.11c
monomer	0.7920	0.7917
dimer	0.4804	0.4238
tetramer	0.2992	0.2609
decamer	0.1824	0.1699

values of first perturbation integral for each molecule in Table J.2 are caused by the numerical procedure that was used to solve the perturbation integral constants. In the PHSC model the parameters of the power series are found by regressing a set of computed I_1 values, while we used a theoretical approach. Therefore, it seems that for optimization of the PHSC model parameters, besides the generated data as was used in [77], Eq. J.11c could serve as a boundary condition.

In the PC-SAFT model the square-well potential is modified to

$$\tilde{u}(x) = \begin{cases} \infty, & x < 0.88 \\ 3, & 0.88 \le x < 1 \\ -1, & 1 \le x < 1.5 \\ 0, & x \ge 1.5. \end{cases} \tag{J.12}$$

This makes the above method not directly applicable for PC-SAFT. The equation reflects that the PC-SAFT model no longer has a hard wall at $x = 1$, but has an indent for molecules that have a relative energy higher than 3. The developers opted for a hard contact position at $x = 0.88$. The initial slope of the radial function between 0.88 and 1 becomes lower, because the molecules with less kinetic energy than 3ϵ do not approach the hard wall: they can not approach the central sphere for $x < 1$. Hence, the radial distribution starts at $x = 0.88$ at a lower value and with a lower slope. The net effect is that I_1 is shifted to a higher value. We may write for the starting point:

$$g(0.88) = \frac{1}{m}\mathcal{P}, \tag{J.13}$$

and

$$g'(0.88) = \frac{(m-1)^2}{2m^2}\mathcal{P}, \tag{J.14}$$

where $\mathcal{P}$ is the probability of a molecule to have a kinetic energy larger than 3ϵ, which follows from the equation:

$$\mathcal{P} = \int_{\frac{3\epsilon}{k_\mathrm{B}T}}^{\infty} \exp(-u)\,du = \exp\left(-\frac{3\epsilon}{k_\mathrm{B}T}\right), \tag{J.15}$$

which can be calculated using the energy values given in Table 6.1. For instance, at room temperature methane and n-hexane have $\epsilon/k_\mathrm{B}T = 0.503$ and 0.794, respectively. This yields $\mathcal{P} = 0.22$ for methane and 0.09 for hexane.

The radial distribution function for the PC-SAFT hard-sphere chain at zero packing in the range of $0.88 < x < 1.5$, can now be defined as:

$$g(m, x, \eta = 0) = \begin{cases} 0 & x < 0.88 \\ \left[\frac{1}{m} + (x - 0.88)\frac{(m-1)^2}{2m^2}\right]\exp\left(-\frac{3\epsilon}{k_\mathrm{B}T}\right) & 0.88 \le x < 1 \\ \frac{1}{m} + (x-1)\frac{(m-1)^2}{2m^2} & 1 \le x < 1.5. \end{cases} \tag{J.16}$$

Interestingly, there appears to be a temperature dependency in the radial distribution function, which has been neglected in the parameterization of the perturbation

integral of the PC-SAFT model. As an example, we work out the case of $m = 1$ (methane) and $m = 4$ (nonane):

$$g\left(1, x, \eta = 0\right) = \begin{cases} 0 & x < 0.88 \\ 0.22 & 0.88 \leq x < 1 \\ 1 & 1 \leq x < 1.5. \end{cases}$$

$$g\left(4, x, \eta = 0\right) = \begin{cases} 0 & x < 0.88 \\ 0.09\left(\frac{1}{4} + (x - 0.88)\frac{9}{32}\right) & 0.88 \leq x < 1 \\ \frac{1}{4} + (x - 1)\frac{9}{32} & 1 \leq x < 1.5. \end{cases} \quad \text{(J.17)}$$

Hence, for methane we obtain:

$$I_1\left(1, \eta = 0\right) = 22\% \times 0.106 + \frac{19}{24} = 0.815, \quad \text{(J.18)}$$

which is 3% above the PHSC model. For n-hexane the result is:

$$I_1\left(1, \eta = 0\right) = 9\% \times 0.028 + 0.261 = 0.263, \quad \text{(J.19)}$$

which is only 1% above the PHSC model. When we set $\mathcal{P} = 1$ the maximum difference is obtained; 13% for methane, and 11% for nonane. The results given in Fig. 6.1 show a larger difference than 13% for methane at $\eta = 0$, which can only be explained as a result of the applied parameter optimization, in which this overshoot of I_1 of the PC-SAFT EoS, is compensated by the I_2 of the EoS. As a result, the I_1 integral parameters of PC-SAFT cannot be used to quantify the dispersion contribution in an activity coefficient model, and is one of the two reasons to develop the IPC model. The other reason is the quantification of effects arising from chain ends and cross-links. PC-SAFT does this by tweaking the segment size and the number of segments, because the model uses a Percus-Yevick type 2 approximation [30].

Appendix K

From PC-SAFT based to other dispersion activity coefficient models

We start with the PC-SAFT based activity coefficient equation for a solute at infinite dilution in a solvent (Eq. 6.18):

$$\ln \gamma_k^{\text{disp},\infty} = m_k \frac{Z_k}{2} \left(\frac{T_0}{T}\right) \left(\frac{\sigma_k^3}{d_k^3}\tilde{\epsilon}_k - 2\frac{\sigma_{jk}^3}{d_j^3}\tilde{\epsilon}_{jk} + \frac{d_k^3}{d_j^3}\frac{\sigma_j^3}{d_j^3}\tilde{\epsilon}_j\right)$$
$$+ m_k \frac{Z_\infty}{2} \left(\frac{T_0}{T}\right) \left[\frac{\tilde{a}_2}{\tilde{a}_1 m_j} + \frac{2\tilde{a}_3}{\tilde{a}_1 m_j^2}\right] \left(1 - \frac{m_j}{m_k}\right) \left(\frac{\sigma_j}{d_j}\right)^3 \tilde{\epsilon}_j. \qquad \text{(K.1)}$$

For long chains, $m >> 1$, the second part vanishes and Eq. K.1 reduces to:

$$\ln \gamma_k^{\text{disp},\infty} = m_k \frac{Z_k}{2} \left(\frac{T_0}{T}\right) \left(\frac{\sigma_k^3}{d_k^3}\tilde{\epsilon}_k - 2\frac{\sigma_{jk}^3}{d_j^3}\tilde{\epsilon}_{jk} + \frac{d_k^3}{d_j^3}\frac{\sigma_j^3}{d_j^3}\tilde{\epsilon}_j\right). \qquad \text{(K.2)}$$

When also the segments are temperature independent, $d_j = \sigma_j$ we obtain:

$$\ln \gamma_k^{\text{disp},\infty} = m_k \frac{Z_k}{2} \left(\frac{T_0}{T}\right) \left(\tilde{\epsilon}_k - 2\frac{\sigma_{jk}^3}{\sigma_j^3}\tilde{\epsilon}_{jk} + \frac{\sigma_k^3}{\sigma_j^3}\tilde{\epsilon}_j\right). \qquad \text{(K.3)}$$

In case all chains have the same segments size, $\sigma_j = \sigma_k$, and when either the Lorentz with $l_{jk} = 0$ or the Waldman and Hagler size combining rule is applied, so that $\sigma_{jk} = \sigma_j$, than Eq. K.3, reduces to:

$$\ln \gamma_k^{\text{disp},\infty} = m_k \frac{Z_k}{2} \left(\frac{T_0}{T}\right) (\tilde{\epsilon}_k - 2\tilde{\epsilon}_{jk} + \tilde{\epsilon}_j), \qquad \text{(K.4)}$$

which leads to a van Laar type of equation, when either the Berthelot with $k_{jk} = 0$ or the Waldman and Hagler energy combining rule, so that $\tilde{\epsilon}_{jk} = \sqrt{\tilde{\epsilon}_k \tilde{\epsilon}_k}$, is applied:

$$\ln \gamma_k^{\text{disp},\infty} = m_k \frac{Z_k}{2} \left(\frac{T_0}{T} \right) \left(\sqrt{\tilde{\epsilon}_k} - \sqrt{\tilde{\epsilon}_j} \right)^2 . \tag{K.5}$$

However, for chains with different diameter Eq. K.3 yields under classical mixing rules a nearly van Laar type of equation:

$$\ln \gamma_k^{\text{disp},\infty} = m_k \frac{Z_k}{2} \left(\frac{T_0}{T} \right) \left(\sqrt{\tilde{\epsilon}_k} - \sqrt{\tilde{\epsilon}_j} \right) \left(\sqrt{\tilde{\epsilon}_k} - \frac{\sigma_k^3}{\sigma_j^3} \sqrt{\tilde{\epsilon}_j} \right) . \tag{K.6}$$

The differences between the PC-SAFT based dispersion activity coeffecent contribution and the other equations reflect the assumptions made in the cEoS:
1) Mean field approximation (long chains average the field $m >> 1$)
2) Temperature independent segment volume, size ($d_j = \sigma_j$)
3) Molecules have equal sized segments ($\sigma_j = \sigma$)
4) Classical combing rules can be applied ($2\sigma_{ij} = \sigma_i + \sigma_j$, $\tilde{\epsilon}_{ij} = \sqrt{\tilde{\epsilon}_i \tilde{\epsilon}_j}$)
5) All molecules have same coordination number, $Z_k = Z_0$

The last assumption means that all the molecules are spheres, which is the way van der Waals defined molecules [219]. One could also take the product $m_k Z_k$ as the liquid volume parameter and $\sqrt{\tilde{\epsilon}_k T_0}$ as the solubility parameter. In that way Eq. K.5 becomes identical to the Hildebrand regular solution model [86, 88].

Appendix L

IPC activity model results

Here we show besides the IPC activity model with IC combining rules (Fig. 6.7) also the results of this model with WH and Kong combining rules.
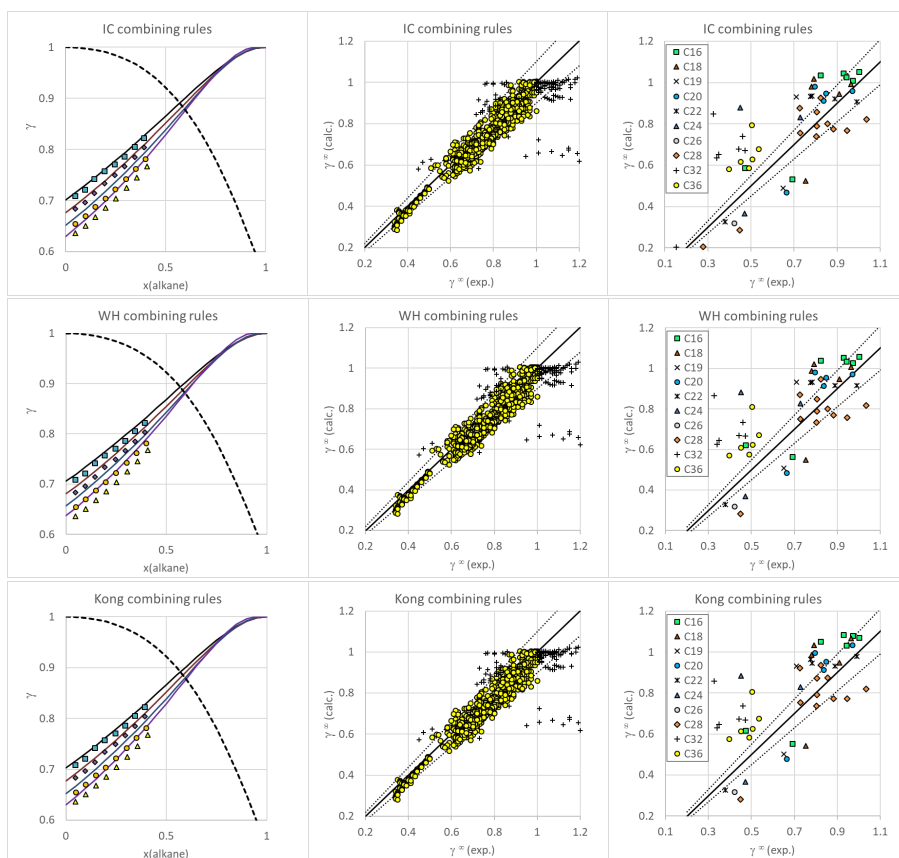
Figure L.1: Results of the IPC model with IC (top), WH (middle) and Kong (bottom) combining rules. Same notation and references as fig. 6.7.

Appendix M

Second perturbation at infinite pressure

Here we show that at infinite pressure the first perturbation term of both the PHSC and the PC-SAFT model becomes dominant over the second term. This can be proven by a theoretical explanation and a practical one based on the equations used in the perturbed chain models.

The theoretical explanation is that at infinite pressure the compressibility factor has become infinite. Since C_1 is inversely proportional to the compressibility factor it has to be zero. This brings I_2 to zero and thus only I_1 is important.

The explanation using model equations is as follows. The ratio of the two terms in the Helmholtz energy for a pure component is given by:

$$\left| \frac{A_2}{A_1} \right| = \left| \frac{\pi \rho m C_1 I_2 m^2 \left(\frac{\epsilon}{k_B T} \right)^2 \sigma^3}{2\pi \rho I_1 m^2 \frac{\epsilon}{k_B T} \sigma^3} \right| = \left| \frac{I_2}{I_1} \right| \frac{m C_1}{2} \left(\frac{\epsilon}{k_B T} \right). \tag{M.1}$$

At infinite pressure we assume that the packing fraction of the liquid $\eta_{\mathrm{LCP}} \simeq 0.637$. With the expression for C_1, Eq. 6.4, we can write for Eq. M.1:

$$\left| \frac{A_2}{A_1} \right| = \left| \frac{I_2}{I_1} \right| \left[\frac{m}{39 + 454m} \right] \left(\frac{\epsilon}{k_B T} \right). \tag{M.2}$$

Fig. M.1 shows the ratio of the two perturbation integrals PHSC and the PC-SAFT model as function of the segment number m at $\eta = \eta_{\mathrm{LCP}} = 0.64$. We see that PHSC model gives a ratio, which is always less than 0.2. The curve starts at 0.19 and runs to the asymptotic value of 0.083. As mentioned in the main text, the PC-SAFT model gives erroneous results beyond the singular point located at
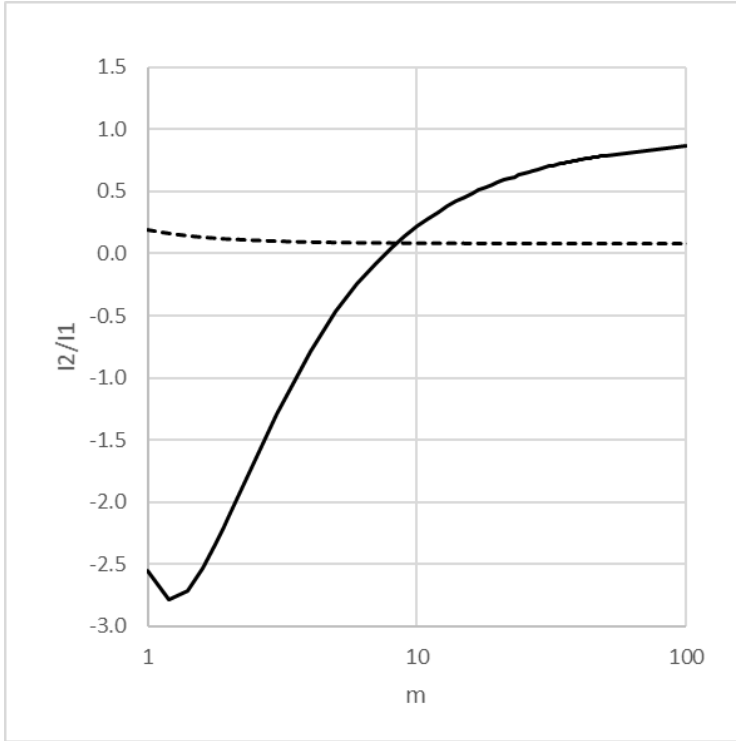
Figure M.1: Ratio in second and first perturbation integral at $\eta = \eta_{\mathrm{LCP}}$ as function of the number of segments, m, according to PHSC (dashed curve) and the PC-SAFT(solid) curve.

$\eta = 0.46$. Therefore, the depicted solid curve in Fig. M.1, which starts at -2.55 and runs to the asymptotic value of 0.94, is of no value. The second term of M.2 is always less than 0.0022. The largest dispersion interaction energy, which has been reported by Gross and Sadowski [78] is about 350 K. The minimum temperature for which the perturbation hard-sphere chain models can be applied is where the de Broglie wavelength is much smaller than the average intermolecular distance. Taking $\sigma = 3$ Å and $\eta = 0.64$ we get for this distance $3 \times 0.36^{1/3} = 2$ Å. When we permit 10% of this distance as acceptable for being a classical fluid, i.e. Maxwell-Boltzmann statistics, than the de Broglie wavelength should not be larger than 0.2 Å. At 298 K the de Broglie wavelength of methane is 0.1 Å. To get a value that not exceeds 0.2 Å, the temperature has to be above 90 K. For larger molecular masses the temperature can be lower. So for alkanes we can take 90 K as the lowest possible temperature at which PHSC and PC-SAFT models are still valid. The largest dispersion interaction energy, which has been reported by Gross and Sadowski [78] is about 350 K. Hence we conclude that at infinite pressure and a temperature above 90 K, according to

the parameters of the PHSC model,

$$\left|\frac{A_2}{A_1}\right| < 0.2 \times 0.0022 \times 360/90 < 0.002. \tag{M.3}$$

For the PC-SAFT model we have:

$$\left|\frac{A_2}{A_1}\right| < 2.55 \times 0.0022 \times 360/90 < 0.025. \tag{M.4}$$

We therefore conclude from theoretical and practical arguments that the second perturbation term can be ignored when the system is set to infinite pressure.

Appendix N

Boiling point estimation

To demonstrate the benefit of the parameter Z_j, we show how well it is capable of predicting the boiling temperature at atmospheric pressure, commonly known as the normal boiling point, $T_{\rm b}$, of alkanes. Initially we had an overestimation of $T_{\rm b}$ for the cyclo-alkanes and a few branched alkanes, because we used the Zagreb index without realizing that rigid rings and the proximity of three or more alkyl groups can prevent intramolecular interaction. After introducing corrections for this, which lead to the definition of D, we obtained the following relation for $T_{\rm b}$:

$$T_{\rm b} = -90.3 + 132.2\sqrt{\frac{m_j Z_j}{2}} - 64.4 \ln \frac{m_j Z_j}{2} - 3.38 \frac{m_j Z_j}{2}. \tag{N.1}$$

Results for the correlation curve and the experimental normal boiling points are plotted in Fig. N.1 for the range from methane to squalane. An excellent description running from methane to squalane is depicted as can also be concluded from the statistical information: coefficient of determination $R^2 = 0.9985$, F-test value $F = 24580$, standard deviation $s = 3.3$ K, degrees of freedom $NF = 109$ and the reduced $\chi^2_{\rm red} = 2.6$.
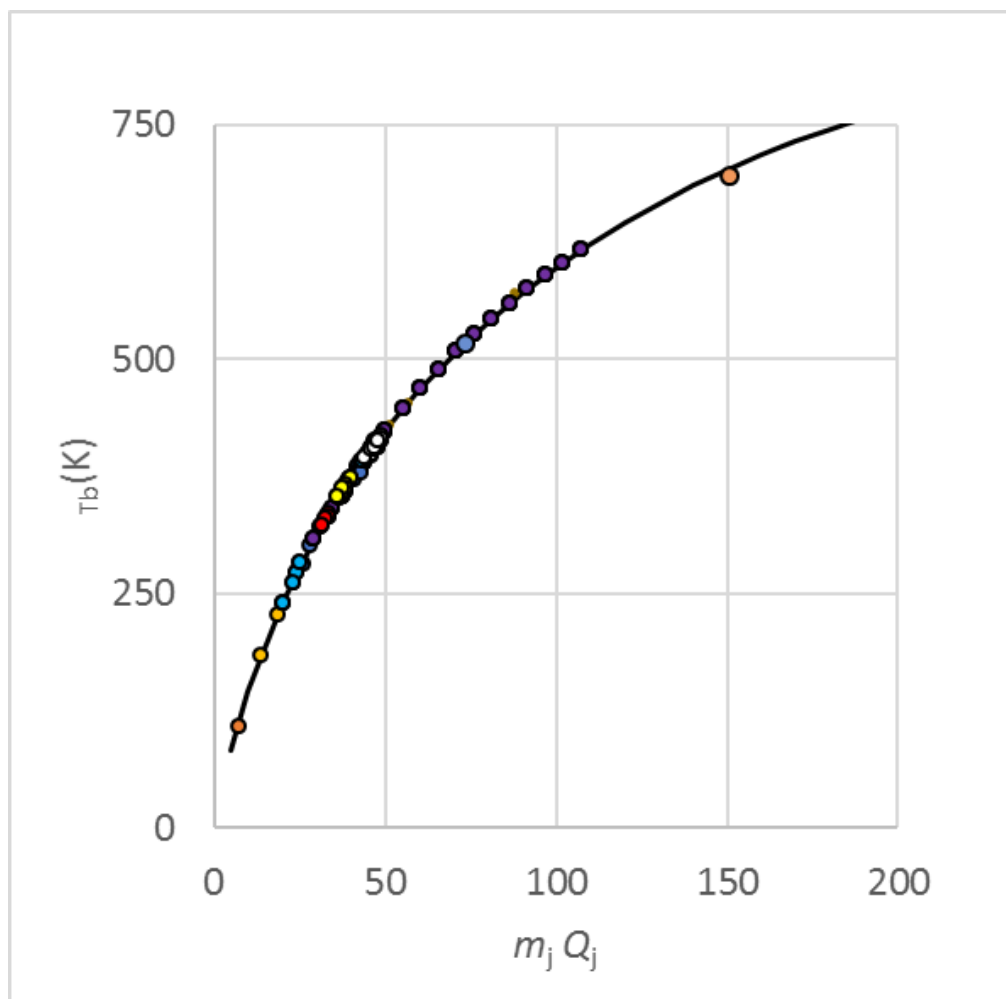
Figure N.1: Normal boiling point, T_b, as function of the compound specific product $m_j Z_j / 2$ for linear, branched and cyclo-alkanes [23], with $1 \leq N_C \leq 30$.

Appendix O

Long n-Alkanes in short alkanes

Here we take a closer look at the experimental solubility data of long alkanes in short n-alkanes, which were studied by Madsen and Boistelle [156, 157], and apply the solubility theory to obtain the activity coefficients of solutes.

The activity of a solute at temperature T is given by the solid-liquid equilibrium

$$\ln\left(a_2\right) = \frac{\Delta H_{\mathrm{m}}}{RT_{\mathrm{m}}}\left(1 - \frac{T_{\mathrm{m}}}{T}\right) + \frac{\Delta H_{\mathrm{tr}}}{RT_{\mathrm{tr}}}\left(1 - \frac{T_{\mathrm{tr}}}{T}\right) + \frac{\Delta C_p}{R}\left[1 - \frac{T_{\mathrm{m}}}{T} + \ln\left(\frac{T_{\mathrm{m}}}{T}\right)\right] \quad \text{(O.1)}$$

where ΔH_{m} is the enthalpy of fusion, ΔH_{tr} is the enthalpy of transition between two solid states, ΔC_p is the average heat capacity difference between pure liquid and solid state. Briard *et al.* studied accurately the solid-liquid state transition of n-alkanes. Table O.1 shows for octacosane, dotriacontane and hexatriacontane the transition point, the melting point, the enthalpy of solid order-disorder transition, the enthalpy of fusion, as reported by Briard, and also the heat capacity of solid and liquid at 300 K retrieved from DIPPR [23]. The last column is the difference in heat capacity between pure solid and liquid state at 300 K, which we take as an estimate for the average change in heat capacity over the range of integration between system temperature and melting point.

| Property | T_{tr} | T_{m} | ΔH_{tr} | ΔH_{m} | $C_{p,\mathrm{solid}}$ | $C_{p,\mathrm{liq.}}$ | ΔC_{p} |
Compound	K	K	kJ/mol	kJ/mol	J/mol.K	J/mol.K	J/mol.K
nC28	330.5	334.2	35.1	65.1	840	882	42
nC32	338.0	341.9	44.2	75.8	886	964	77
nC36	346.1	348.9	31.1	87.7	831	1143	311

Table O.1: Thermodynamic constants for pure n-alkanes [20] and C_p from DIPPR [23] taken at 300K

Fig. O.1 shows the experimental solubility data of octacosane, dotriacontane and hexatriacontane in small alkanes measured by Madsen and Boistelle [156, 157]. The infinite dilution activity coefficient (IDAC) of the long alkanes was calculated by Eq. O.1 using the parameters of table O.1. We see that the solubility data run nearly
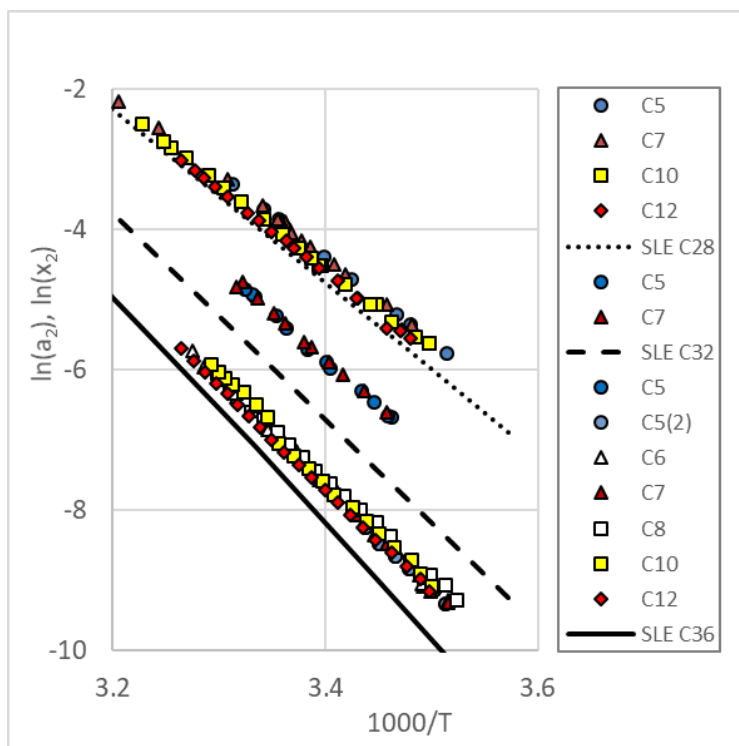
Figure O.1: Logarithmic solubility data of [156, 157] and logarithmic activity curves (Eq. O.1) as function of the reciprocal temperature. Enthalpy and entropy of solvation of the experintal data are summarized in table O.2.

parallel to the calculated activity coefficient. The solubility data can be expressed in terms of the enthalpy, ΔH_{sol} and entropy, ΔS_{sol}, of solvation:

$$\ln\left(x_2\right) = -\frac{\Delta H_{\mathrm{sol}}}{RT} + \frac{\Delta S_{\mathrm{sol}}}{R}. \tag{O.2}$$

By fitting the solubility data series of Madsen and Boistelle, we get the enthalpy and entropy of solvation. These are summarized in Table O.2.

Fig. O.1 shows a gap between the experimental solubility data and the corresponding activity curve. The larger the gap, the lower the activity coefficient as can

Lit.	System	ΔH_{sol} [kJ/mol]	ΔS_{sol} [J/mol.K]	R^2 [-]	γ_2^{∞} (300) Eq. O.3,	γ_2^{∞} (300) IPC	γ_2^{∞} (T_{exp}) Kniaz [120]
[156]	28/5	99.4 ± 0.8	301 ± 3	0.9995	0.72	0.79	0.82
[156]	28/7	97.3 ± 1.0	294 ± 4	0.9986	0.72	0.69	0.80
[157]	28/10	97.4 ± 0.7	293 ± 2	0.9991	0.84	0.75	0.88
[157]	28/12	99.0 ± 0.9	298 ± 3	0.9986	0.90	0.81	1.04
[156]	32/5	111.1 ± 1.7	329 ± 6	0.9976	0.48	0.73	0.33
[156]	32/7	109.4 ± 2.1	323 ± 7	0.9968	0.47	0.61	0.35
[156]	36/5	125.7 ± 6.3	364 ± 21	0.9777	0.61	0.66	**
[157]	36/5	114.1 ± 2.6	324 ± 9	0.9979	0.78	0.66	0.50
[157]	36/6	120.2 ± 11.1	344 ± 39	0.9668	0.68	0.55	0.51
[156]	36/7	120.3 ± 1.6	346 ± 5	0.9980	0.63	0.53	**
[157]	36/7	122.7 ± 5.3	354 ± 18	0.9927	0.63	0.53	0.49
[157]	36/8	118.4 ± 0.9	340 ± 3	0.9980	0.60	0.54	0.40
[157]	36/10	126.3 ± 1.9	367 ± 6	0.9963	0.59	0.59	0.45
[157]	36/12	122.2 ± 0.5	351 ± 2	0.9980	0.70	0.66	0.53

Table O.2: Thermodynamic constants for solvation of long n-alkanes in solvents by analysis of experimental data of [156, 157]. IDAC of the paraffin in solvent calculated by Eq. O.3 and the IPC model (Eq. 6.38) at 300 K, and as reported by Kniaz [120]. (** = not reported).

be understood by the difference of Eq. O.1 and Eq. O.2:

$$\ln\left(\gamma_2\right) = \ln\left(a_2\right) - \ln\left(x_2\right) \tag{O.3}$$

In the fore last and last column of table O.2 we give the values found by Eq. O.3 at 300 K and those reported by Kniaz [120], who used Eq. O.1 without heat capacity correction and varying temperature. We take $T = 300$ K, $1000/T = 3.33$, because it lies far enough below the transition temperature and within the experimental data range. In Fig. O.2 the activity coefficients at 300 K and the results of the IPC model (Eq. 6.38) are shown. We observe that the IPC model gives good description for activity coefficient at 300 K of octacosane and hexatriacontane in small normal alkanes. The trend of an upswing near hexane, as predicted by the IPC model is noticeable in the experimental data. The IDAC of dotriacontane lie outside the range of octacosane and hexatriacontane. This could imply that in the analytic procedure the point of dissolution of dotriacontane crystals in pentane and heptane was not accurately observed. The solubility data points of dotriacontane in Fig. O.1 fall on a curve more to the right than the gap between activity and solubility of hexatriacontane. This could be an indication that the used dotriacontane contained an impurity with a lower melting point, or that the sample contained an amount of residue solvent, which was not completely removed from the final purified sample.
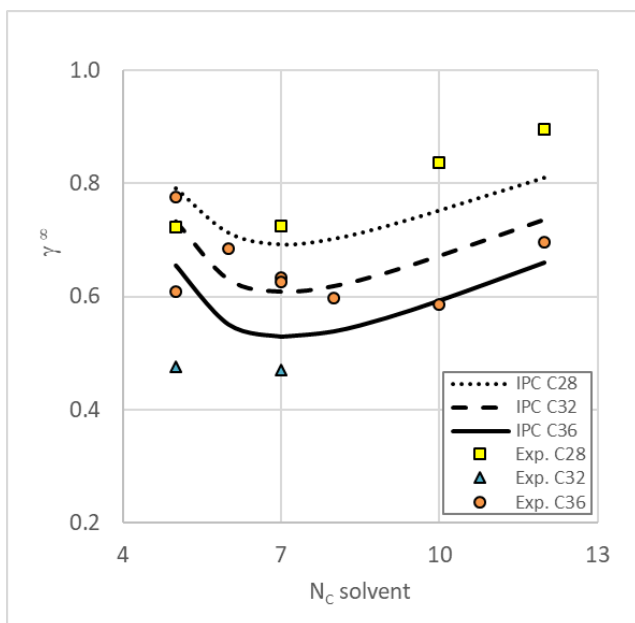
Figure O.2: Limiting activity coefficients of long n-alkanes as function of the carbon number of the n-alkane solvent. Data points are the values at 300 K of table O.2. Curves are the results of the IPC activity coefficient model (Eq. 6.38).

Madsen and Boistelle pointed out in their first article [156] that they initially used another analysis method. In here a paraffin saturated solution was weighted and after drying the amount of paraffin was measured. However, in this procedure either the paraffin co-evaporated with the solvent, or the dried sample still contained rest solvent. The last could have happened when they purified the >95% dotriacontane.

Appendix P

Topology number increments

Table P.1 gives overview of the increments needed to calculate the topology number D_j. We mention that the number of interacting spheres is given by $Z_j = 14.4 - D_j > 0$.

Table P.1: D of alkyl groups.

Main	Sub	$m_{v,j}$	$m_{v,j}$	$m_{v,j}$	$m_{v,j}$	D	example/note
CH$_4$	1	-	-	-	-	0	Methane
CH$_3$	2	-CH$_3$	-	-	-	1	Ethane
	3	-CH$_2$-	-	-	-	2	n-alkanes
	4	-CH<	-	-	-	3	2-methyl-pentane
	5	>C<	-	-	-	4	2,2-diethyl-hexane
CH$_2$	6	-CH$_3$	-CH$_3$	-	-	2	Propane
	7	-CH$_3$	-CH$_2$-	-	-	3	n-alkanes
	8	-CH$_2$-	-CH$_2$-	-	-	4	n-alkane
	9	-CH<	-CH$_2$-	-	-	5	3-methyl-pentane
	10	-CH<	-CH<	-	-	6	3,5-dimethyl-hexane
	11	>C<	-CH$_2$-	-	-	6	n-Hexane
	12	>C<	-CH<	-	-	7	2,2,4-trimethylpentane
	13	>C<	>C<	-	-	8	2,2,4,4-tetramethylpentane
CH	14	-CH$_3$	-CH$_3$	-CH$_3$	-	3	i-C4
	15	-CH$_3$	-CH$_3$	-CH$_2$-	-	4	
	16	-CH$_3$	-CH$_3$	-CH<	-	5	
	17	-CH$_3$	-CH$_3$	>C<	-	6	
	18	-CH$_3$	-CH$_2$-	-CH$_2$-	-	5	
	19	-CH$_3$	-CH$_2$-	-CH<	-	6	
	20	-CH$_3$	-CH$_2$-	>C<	-	7	

Main	Sub	$m_{v,j}$	$m_{v,j}$	$m_{v,j}$	$m_{v,j}$	D	example
				Continuation of Table G.1			
	21	-CH$_3$	-CH<	-CH<	-	7	
	22	-CH$_3$	-CH<	>C<	-	8	
	23	-CH$_3$	>C<	>C<	-	9	
	24	-CH$_2$-	-CH$_2$-	-CH$_2$	-	6	
	25	-CH$_2$-	-CH$_2$-	-CH<	-	7	
	26	-CH$_2$-	-CH$_2$-	>C<	-	8	
	27	-CH$_2$-	-CH<	-CH<	-	8	
	28	-CH$_2$-	-CH<	>C<	-	9	
	29	-CH$_2$-	>C<	>C<	-	10	
	30	-CH<	-CH<	-CH<	-	9	
	31	-CH<	-CH<	>C<	-	10	
	32	-CH<	>C<	>C<	-	11	
	33	>C<	>C<	>C<	-	12	
C	34	-CH$_3$	-CH$_3$	-CH$_3$	-CH$_3$	4	neopentane
	35	-CH$_3$	-CH$_3$	-CH$_3$	-CH$_2$-	5	
	36	-CH$_3$	-CH$_3$	-CH$_3$	-CH<	6	
	37	-CH$_3$	-CH$_3$	-CH$_3$	>C<	7	
	38	-CH$_3$	-CH$_3$	-CH$_2$-	-CH$_2$-	6	
	39	-CH$_3$	-CH$_3$	-CH$_2$-	-CH<	7	
	40	-CH$_3$	-CH$_3$	-CH$_2$-	>C<	8	
	41	-CH$_3$	-CH$_3$	-CH<	-CH<	8	
	42	-CH$_3$	-CH$_3$	-CH<	>C<	9	
	43	-CH$_3$	-CH$_3$	>C<	>C<	10	
	44	-CH$_3$	-CH$_2$-	-CH$_2$-	-CH$_2$	6	
	45	-CH$_3$	-CH$_2$-	-CH$_2$-	-CH<	7	
	46	-CH$_3$	-CH$_2$-	-CH$_2$-	>C<	8	
	47	-CH$_3$	-CH$_2$-	-CH<	-CH<	9	
	48	-CH$_3$	-CH$_2$-	-CH<	>C<	10	
	49	-CH$_3$	-CH$_2$-	>C<	>C<	11	
	50	-CH$_3$	-CH<	-CH<	-CH<	10	
	51	-CH$_3$	-CH<	-CH<	>C<	11	
	52	-CH$_3$	-CH<	>C<	>C<	12	
	53	-CH$_3$	>CH<	>CH<	-CH<	13	
	54	-CH$_2$	-CH$_2$-	-CH$_2$-	-CH$_2$	8	
	55	-CH$_2$	-CH$_2$-	-CH$_2$-	-CH<	9	
	56	-CH$_2$	-CH$_2$-	-CH$_2$-	>C<	10	
	57	-CH$_2$	-CH$_2$-	-CH<	-CH<	10	

Main	Sub	$m_{v,j}$	$m_{v,j}$	$m_{v,j}$	$m_{v,j}$	D	example
					Continuation of Table G.1		
	58	-CH$_2$	-CH$_2$-	-CH<	>C<	11	
	59	-CH$_2$	-CH$_2$-	>C<	>C<	12	
	60	-CH$_2$	-CH<	-CH<	-CH<	11	
	61	-CH$_2$	-CH<	-CH<	>C<	12	
	62	-CH$_2$	-CH<	>C<	>C<	13	
	63	-CH$_2$	>CH<	>CH<	>CH<	14	
	64	-CH<	-CH<	-CH<	-CH<	12	
	65	-CH<	-CH<	-CH<	>C<	13	
	66	-CH<	-CH<	>C<	>C<	14	
	67	-CH<	>C<	>C<	>CH<	15	$D > 14.7$ Z=0!
	68	>C<	>C<	>C<	>C<	16	$D > 14.7$ Z=0!
				End of Table			

Appendix Q

Induced dipole interaction model

The relative energy difference between two different atoms in reference to two equal atoms can be defined by:

$$\Delta\epsilon_{ij} = \epsilon_{ij} - (\epsilon_{ii} + \epsilon_{jj})/2 \tag{Q.1}$$

where ϵ_{ij} is the dispersion energy between atom i and j. Consider a CH_3 end group interacting by dispersion with a CH_2 chain group. When we assume that each atom can form an induced dipole, than the two atom groups can make 12 induced dipole pairs, as indicated in Table Q.1.

Table Q.1: Example of CH3-CH2 induced dipole interaction energies.

	C	H	H
C	ϵ_{CC}	ϵ_{CH}	ϵ_{CH}
H	ϵ_{CH}	ϵ_{HH}	ϵ_{HH}
H	ϵ_{CH}	ϵ_{HH}	ϵ_{HH}
H	ϵ_{CH}	ϵ_{HH}	ϵ_{HH}

The interaction between a CH_m and a CH_n group involves one C..C induced dipole pair, $m + n$ C..H induced dipole pairs and $m.n$ H..H induced dipole pairs. The corresponding dipole energies are ϵ_{CC}, ϵ_{CH}, and ϵ_{HH}, respectively. The relative

energy difference is given by:

$$\Delta\epsilon_{CHm,CHn} = \epsilon_{\text{CHm,CHn}} - \tfrac{1}{2}\left(\epsilon_{\text{CHm,CHm}} + \epsilon_{\text{CHn,CHn}}\right)$$
$$= \left[\epsilon_{\text{CC}} + (m+n)\,\epsilon_{\text{CH}} + mn\epsilon_{\text{HH}}\right]$$
$$- \tfrac{1}{2}\left[\left(\epsilon_{\text{CC}} + 2m\epsilon_{\text{CH}} + m^2\epsilon_{\text{HH}}\right) + \left(\epsilon_{\text{CC}} + 2n\epsilon_{\text{CH}} + n^2\epsilon_{\text{HH}}\right)\right]$$
$$= -\tfrac{1}{2}\left(m-n\right)^2\epsilon_{\text{HH}}, \tag{Q.2}$$

where ϵ_{HH}, ϵ_{CC} is the dispersion energy between two hydrogen and two carbon atoms, respectively and ϵ_{CH} is the dispersion energy between a hydrogen and a carbon atom. Eq. Q.2 shows that only the hydrogen-hydrogen induced dipole matters. Table Q.2 illustrates this.

Table Q.2: Relative interaction energies between two methyl groups.

$-2\Delta\epsilon_{ij}/\epsilon_{\text{HH}}$	j=CH$_3$	j=CH$_2$	j=CH	j=C
i=CH$_3$	0	1	4	9
i=CH$_2$	1	0	1	4
i=CH	4	1	0	1
i=C	9	4	1	0

We observe that Table Q.2 is symmetric around both matrix diagonals, and that the relative interaction energies increases quadratic with increasing difference in hydrogen number.

Appendix R

Gibbs-Duhem consistency tests

Activity models are thermodynamic consistent, when they obey the Gibbs-Duhem equation. This equation relates the chemical potential of one of the components in a mixture to that of another. In term of the activity coefficients, and for an isothermal-isobaric binary system, we have:

$$x_1 \mathrm{d} \ln \gamma_1 + x_2 \mathrm{d} \ln \gamma_2 = 0, \tag{R.1}$$

where x_j and γ_j are the mole fraction and activity coefficient of component j. To check whether a model obeys this rule, an Integral and Differential test [103] is carried out, as explained below.

R.1   Integration Test

By integration of Eq. R.1 we obtain:

$$\int_0^1 \ln \left[\frac{\gamma_1}{\gamma_2} \right] dx_1 = 0 \tag{R.2}$$

Note that in COSMO-SAC and COSMO-3D the excess volume, V^{E}, is neglected. The effect of this assumption is not significant, because the term:

$$| \frac{1}{RT} \int V^{\mathrm{E}} dP | << \int_0^1 \ln \left[\frac{\gamma_1}{\gamma_2} \right] dx_1, \tag{R.3}$$

holds for most mixtures. If we plot the function $\ln \left[\frac{\gamma_1}{\gamma_2} \right]$ as function of x_1, the horizontal axis splits the integral into two areas. The ratio of these areas has to be close to 1 (see fig. 2.1. Assuming some light deviation coming from numerical integration, we take a threshold of 0.5%.

R.2   Differential Test

Eq. R.1 differentiated towards x_1 yield:

$$x_1 \left(\frac{\partial \ln \gamma_1}{dx_1} \right)_{T,P} + x_2 \left(\frac{\partial \ln \gamma_2}{dx_1} \right)_{T,P} = 0 \tag{R.4}$$

Let A and B be the first and second term of R.4, respectively, then the quantity:

$$\frac{(A-B)^2}{AB} \tag{R.5}$$

is an indicator. A value of zero represent perfect thermodynamic consistency. A negative or large positive value indicates an incorrect model. A small positive value is usually caused by a numerical error in the differentiation. We calculated this quantity for all the systems using a finite difference to obtain the derivative of R.4. All the predictions obtained with COSMO-3D and COSMO-SAC pass the differential test.

Appendix S

Statistical Information

Table S.1: Results of the training set for the COSMOSAC [94] (SAC), the refined COSMOSAC (REF), and the COSMO-3D (3D) models.

Mixture	T (°C)	P^{dev} (%)			y^{dev} (%)			γ^{dev} (RMS)		
		SAC	REF	3D	SAC	REF	3D	SAC	REF	3D
1-hexene/ethyl acetate	60.00	4.11	2.76	2.42	1.69	1.19	1.00	0.17	0.11	0.10
2-butanone/heptane	45.00	6.72	2.62	0.49	2.78	0.68	0.58	0.45	0.18	0.05
MTBE/methyl acetate	80.00	2.17	1.08	1.16	0.47	0.18	0.19	0.04	0.02	0.02
MTBE/methyl acetate	90.00	2.02	0.94	1.03	0.49	0.21	0.21	0.04	0.01	0.02
MTBE/methyl acetate	100.00	1.95	0.85	0.95	0.32	0.13	0.12	0.04	0.01	0.02
2-methylpropanal/heptane	45.00	5.11	1.92	1.20	1.95	0.93	0.78	0.33	0.20	0.10
acetone/diisopropyl ether	70.11	7.99	6.41	2.62	2.76	2.08	0.62	0.24	0.19	0.07
acetone/diisopropyl ether	80.12	8.04	6.44	2.76	2.73	2.05	0.77	0.23	0.18	0.07
acetone/diisopropyl ether	90.14	9.25	7.67	4.15	3.21	2.53	1.26	0.28	0.23	0.12
acetone/heptane	40.00	9.02	2.99	3.35	1.92	0.47	1.16	0.53	0.22	0.17
benzene/cyclohexane	39.90	1.21	1.78	0.64	0.44	0.54	0.33	0.02	0.03	0.01
benzene/cyclohexane	40.00	1.13	1.46	0.42	0.32	0.62	0.24	0.03	0.05	0.02
benzene/heptane	25.00	3.86	0.99	1.26	1.25	0.36	0.45	0.11	0.03	0.04
benzene/heptane	40.00	3.41	0.89	0.80	0.96	0.21	0.19	0.11	0.04	0.03
benzene/hexane	30.00	3.80	1.20	1.46	1.76	0.62	0.72	0.13	0.05	0.06
benzene/hexane	40.00	3.82	1.32	1.46	2.25	1.13	1.16	0.15	0.07	0.07
benzene/hexane	50.00	3.69	1.37	1.33	1.61	0.62	0.57	0.12	0.05	0.04
benzene/hexane	60.00	3.35	1.23	1.10	1.25	0.56	0.51	0.10	0.04	0.04
cyclohexane/heptane	25.00	2.12	2.11	1.88	0.67	0.66	0.59	0.05	0.05	0.04
cyclohexane/heptane	40.00	1.90	1.89	1.66	0.59	0.58	0.50	0.04	0.04	0.04
ethyl acetate/methylcyclohexane	56.85	1.29	3.24	1.07	0.39	1.61	0.49	0.06	0.15	0.05
ethyl acetate/methylcyclohexane	76.85	0.85	2.72	1.01	0.28	1.24	0.50	0.06	0.13	0.05
ethylbenzene/benzaldehyde	75.00	2.34	0.89	1.30	1.07	0.75	0.38	0.08	0.05	0.02
ethylbenzene/benzaldehyde	95.00	2.43	1.16	1.36	0.66	0.30	0.45	0.06	0.03	0.04
heptane/3-pentanone	65.00	5.78	3.43	2.10	2.43	1.33	0.88	0.27	0.15	0.08
heptane/3-pentanone	80.00	5.88	3.60	2.23	2.53	1.51	1.03	0.28	0.17	0.11
heptane/3-pentanone	95.00	4.95	2.90	1.59	2.66	1.69	1.18	0.29	0.18	0.12
heptane/pentanal	75.00	5.41	2.58	0.97	2.48	1.49	1.21	0.26	0.15	0.11
hexane/1,4-dioxane	35.00	5.64	2.93	1.80	2.30	2.06	1.36	0.38	0.25	0.14
oxolane/benzene	60.00	0.74	2.08	0.79	0.48	0.75	0.48	0.02	0.05	0.02
oxolane/cyclohexane	40.00	2.51	1.51	1.39	0.82	0.44	0.33	0.12	0.08	0.07
oxolane/cyclohexane	60.00	2.09	1.05	0.87	0.78	0.36	0.28	0.21	0.18	0.18
oxolane/ethyl acetate	60.00	0.94	0.88	0.79	0.37	0.35	0.32	0.04	0.04	0.03
oxolane/hexane	60.00	3.76	2.88	1.89	1.37	1.01	0.57	0.17	0.13	0.08
oxolane/hexane	25.55	4.15	4.15	4.07	0.80	0.79	0.79	0.07	0.07	0.07
oxolane/hexane	30.55	3.34	3.34	3.27	0.86	0.86	0.86	0.07	0.07	0.07
oxolane/hexane	35.55	4.55	4.55	4.46	1.18	1.18	1.17	0.09	0.09	0.09
pentane/propanal	40.00	8.80	4.82	1.38	3.07	1.58	0.77	0.42	0.24	0.09
propanal/cyclohexane	45.00	5.88	1.86	0.95	2.60	0.66	0.63	0.55	0.20	0.09
Overall		**3.89**	**2.36**	**1.56**	**1.41**	**0.89**	**0.63**	**0.24**	**0.14**	**0.08**

Average relative deviations of pressures and gas mole fractions, and the root mean square of activity coefficients calculated according to Eq. 8.15, 8.16, and 8.17, respectively.

Table S.2: Results of the test set for the COSMOSAC [94] (SAC), the refined COSMOSAC (REF), and the COSMO-3D (3D) models.

Mixture	T (°C)	P^{dev} (%)			y^{dev} (%)			γ^{dev} (RMS)		
		SAC	REF	3D	SAC	REF	3D	SAC	REF	3D
1-hexene /ethyl acetate	40	4.68	3.49	3.19	2.09	1.64	1.41	0.25	0.19	0.17
2-methylpropanal/heptane	61.85	5.63	2.51	0.68	1.88	0.69	0.37	0.35	0.21	0.13
acetone/heptane	0	9.29	4.87	3.62	1.89	1.27	1.27	1.42	1.00	0.56
acetone/heptane	50	8.42	1.57	4.63	2.56	0.95	1.96	0.95	0.57	0.28
acetone/heptane	50	1.56	5.47	7.24	1.30	3.11	3.74	1.07	1.01	0.70
benzene/cyclohexane	10	1.32	1.82	0.17	0.58	0.72	0.19	0.05	0.06	0.01
benzene/cyclohexane	24.91	1.45	1.78	0.37	0.37	0.63	0.26	0.03	0.04	0.01
benzene/cyclohexane	25	0.35	2.49	1.24	0.17	1.08	0.63	0.01	0.09	0.05
benzene/cyclohexane	38.5	1.32	1.57	0.58	0.14	1.06	0.70	0.02	0.05	0.03
benzene/cyclohexane	50	2.67	1.01	1.11	0.31	0.69	0.50	0.05	0.04	0.03
benzene/cyclohexane	55	0.76	2.07	1.46	0.35	0.67	0.45	0.02	0.04	0.03
benzene/cyclohexane	55.05	1.13	1.65	1.04	0.26	0.48	0.33	0.02	0.04	0.02
benzene/cyclohexane	60	1.13	1.13	0.66	0.28	0.65	0.47	0.03	0.04	0.03
benzene/cyclohexane	70	1.02	1.24	0.90	0.38	0.55	0.47	0.03	0.03	0.03
benzene/cyclohexane	70	0.46	2.17	2.04	0.94	0.39	0.50	0.04	0.03	0.02
benzene/heptane	35.04	4.26	1.35	1.08	1.30	0.48	0.39	0.14	0.07	0.07
benzene/heptane	55	3.95	1.88	1.29	1.62	0.93	0.80	0.20	0.13	0.12
benzene/heptane	55	3.29	0.50	0.60	1.33	0.51	0.46	0.15	0.08	0.06
benzene/heptane	60	3.06	0.61	0.55	0.86	0.52	0.60	0.09	0.04	0.03
benzene/heptane	80	2.72	0.57	0.60	0.48	0.74	1.06	0.07	0.05	0.06
benzene/heptane	80	3.21	0.90	0.72	0.90	0.59	0.59	0.08	0.05	0.04
ethylbenzene/benzaldehyde	85	2.15	0.99	1.53	0.71	0.30	0.40	0.06	0.03	0.05
heptane/3-pentanone	26	9.34	6.80	5.69	2.13	1.28	0.64	0.41	0.26	0.20
hexane/benzene	25	3.67	1.38	1.70	1.52	0.67	0.76	0.17	0.07	0.08
hexane/benzene	25	4.35	1.59	1.95	1.57	0.69	0.77	0.11	0.04	0.05
oxolane/benzene	30	0.34	1.60	0.37	0.31	0.66	0.30	0.02	0.05	0.02
oxolane/benzene	40	0.55	1.94	0.56	0.29	0.71	0.30	0.03	0.06	0.03
oxolane/benzene	50	0.55	1.95	0.65	0.46	0.77	0.48	0.02	0.05	0.03
oxolane/cyclohexane	30	2.36	1.70	1.44	0.81	0.94	0.72	0.08	0.08	0.07
oxolane/cyclohexane	50	2.14	0.95	0.77	0.83	0.38	0.33	0.07	0.04	0.03
oxolane/ethyl acetate	40	1.39	1.36	1.25	0.38	0.37	0.34	0.06	0.06	0.06
oxolane/hexane	30	5.53	4.37	3.07	1.61	1.34	0.84	0.14	0.12	0.07
oxolane/hexane	40	4.26	3.38	2.42	1.70	1.31	0.83	0.21	0.16	0.11
oxolane/hexane	40	5.81	5.46	3.34	1.70	1.33	0.86	0.17	0.14	0.10
oxolane/hexane	50	4.75	3.61	2.30	1.47	1.07	0.61	0.13	0.10	0.07
Overall		**3.14**	**2.16**	**1.66**	**1.08**	**0.84**	**0.69**	**0.31**	**0.23**	**0.16**

[a] Average relative deviations of pressures and gas mole fractions, and the root mean square of activity coefficients calculated according to Eq. 8.15, 8.16, and 8.17, respectively.

Appendix T

Supplementary Information

Supplementary information of chapter 8 is available via the journal web-page:
https://pubs.acs.org/doi/suppl/10.1021/ie504285x
Figure T.1 depicts system S32 in chapter 8 of this thesis.

a: VLE plot   **b:** Activity coefficient plot
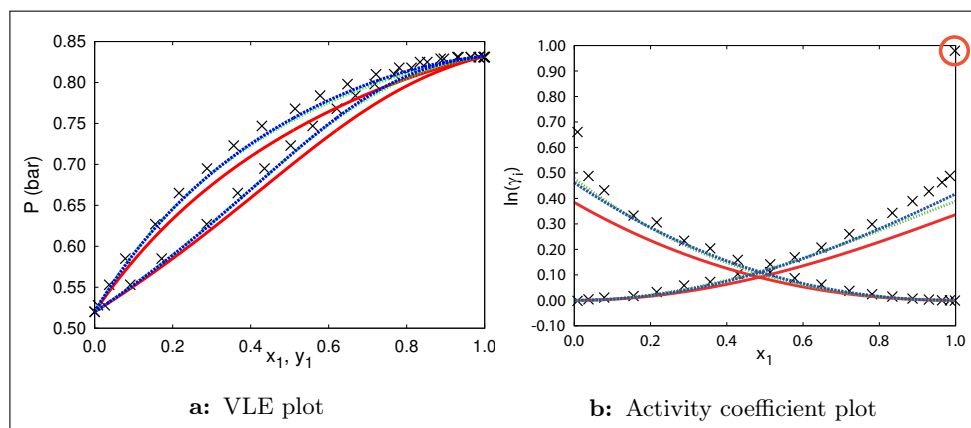
Figure T.1: Vapor-liquid equilibrium of the system oxolane-cyclohexane at $60°$ and corresponding activity coefficient plot. Data taken from [24]. Outlier is indicated by red circle. The red curve is the original COSMO-SAC 2010 model with $A_{ES} = 6525.69$ (kcal/mol)($\text{Å}^4/\text{e}^2$), green curve is the refined COSMO-SAC where $A_{ES} = 7850$ (kcal/mol)($\text{Å}^4/\text{e}^2$), and the blue curve is the COSMO-3D model.

Curriculum Vitae

Gerard Krooshof was born on the 10th of October 1961 in Utrecht. After finishing Gymnasium in 1980 at the catholic high school "De Breul" in Zeist, he studied physics, with secondary subjects in biology and chemistry, at the faculty of mathematics and physics at the State University of Utrecht. He graduated to the degree of "candidatum" in 1984. Subsequently, he continued his study in the field of experimental physics, where he carried out an extended study on the rapid thermal nitridation and silicidation of titanium for the development of new semiconductors. He discovered a new amorphous and crystalline form of titanium silicide, for which he received a university reward. He also studied the interaction of surface molecules in liquid crystals by dynamic light scattering. During his internship at Philips in Eindhoven he improved a non-destructive method for the determination of surface roughness of computer monitor screens. In 1988 he graduated and received the degree of "doctorandus" in experimental physics. Since 1989 he is employed at DSM, where he started as research project leader for the development of Stanyl fiber grade targeted for the production of technical yarn for automotive airbags and transmission belts. While studying chemical technology at Twente University he started in 1996 as group leader of the thermodynamic expertise group, where he contributed to many research projects to improve chemical process steps in the field of chemicals and polymers. After DSM sold the hydrocarbon branch and acquired Gist-Brocades and the nutritional branch of Roche, he became DSM's global champion thermodynamics in 2008. In 2015 he extended his senior expert role in thermodynamics to colloid modelling. During this role he started a PhD project at Eindhoven University of Technology in 2016, of which the results are presented in this dissertation.

List of publications related to this thesis

[1] J.J. Gutiérrez-Sevillano, K. Leonhard, J.P.J.M. Van Der Eerden, T.J.H. Vlugt, and G.J.P. Krooshof. COSMO-3D: Incorporating three-dimensional contact information into the cosmo-sac model. *Industrial and Engineering Chemistry Research*, 54(7):2214–2226, 2015. cited By 5.

[2] G.J.P. Krooshof. Lattice-free UNIQUAC and COSMOSPACE. 29th European symposium on applied thermodynamics (ESAT) 18-21 May 2017, Bucarest, Romenia, 2017.

[3] G.J.P. Krooshof. A topology-based dispersion activity coefficient model derived from PC-SAFT. CERE discussion meeting, June 2019, Helsingør, Danmark, 2019.

[4] G.J.P. Krooshof, R. Tuinier, and G. de With. Generalization of Guggenheim's combinatorial activity coefficient equation. *Journal of Molecular Liquids*, 266:467–471, 2018. cited By 0.

[5] G.J.P. Krooshof, R. Tuinier, and G. de With. On the calculation of nearest neighbors in activity coefficient models. *Fluid Phase Equilibria*, 465:10–23, 2018. cited By 1.

[6] G.J.P. Krooshof, R. Tuinier, and G. de With. Dispersion activity coefficient models. Part 1: Cubic equations of state. *Fluid Phase Equil.*, 2019.

[7] G.J.P. Krooshof, R. Tuinier, and G. de With. Dispersion activity coefficient models. Part 2: Perturbed hard-sphere chain equations of state. *Fluid Phase Equil.*, 2019.

[8] G.J.P. Krooshof, R. Tuinier, and G. de With. Dispersion activity coefficient models. Part 3: A structural preserving group contribution method. *to be published*, 2019.

Other publications by the author

[1] A. Böttger, D. Frenkel, J.G.H. Joosten, and G. Krooshof. Smectic ordering in nematic and smectic liquid-crystalline films probed by means of surface light scattering. *Physical Review A*, 38(12):6316–6323, 1988. cited By 14.

[2] A. Klamt, G.J.P. Krooshof, and R. Taylor. Cosmospace: Alternative to conventional activity-coefficient models. *AIChE Journal*, 48(10):2332–2349, 2002. cited By 80.

[3] A. Klamt, G.J.P. Krooshof, and R. Taylor. Comment on "towards the development of theoretically correct liquid activity coefficient models". *Journal of Chemical Thermodynamics*, 41(11):1312–1313, 2009. cited By 1.

[4] I.A. Kouskoumvekaki, G.J.P. Krooshof, M.L. Michelsen, and G.M. Kontogeorgis. Application of the simplified pc-saft equation of state to the vapor-liquid equilibria of binary and ternary mixtures of polyamide 6 with several solvents. *Industrial and Engineering Chemistry Research*, 43(3):826–834, 2004. cited By 30.

[5] G.J.P. Krooshof. 10 ways to improve UNIFAC. UNIFAC consortium meeting, Sept. 2011, Oldenburg, Germany, 2011.

[6] G.J.P. Krooshof. Can molecular modeling meet the industrial need for robust and quick predictions? 27th European symposium on applied thermodynamics (ESAT) 6-9 July 2014, Eindhoven, Netherlands, 2014.

[7] G.J.P. Krooshof, F.H.P.M. Habraken, W.F. Van Der Weg, L. Van Den Hove, K. Maex, and R.F. De Keersmaecker. Study of the rapid thermal nitridation and silicidation of ti using elastic recoil detection. i. ti on si. *Journal of Applied Physics*, 63(10):5104–5109, 1988. cited By 15.

[8] G.J.P. Krooshof, F.H.P.M. Habraken, W.F. Van Der Weg, L. Van Den Hove, K. Maex, and R.F. De Keersmaecker. Study of the rapid thermal nitridation and

silicidation of ti using elastic recoil detection. ii. ti on sio2. *Journal of Applied Physics*, 63(10):5110–5114, 1988. cited By 25.

[9] G.J.P. Krooshof and R. Meier. Are there ways to improve the accuracy of predictive methods in the field of thermodynamic properties? Congres: Industrial use of molecular thermodynamics (InMoTher), 19-20 March 2012, Lyon, France, 2012.

[10] K. Maex, R.F. de Keersmaecker, M. van Rossum, W.F. van der Weg, and G. Krooshof. Sequence of phase formation during ion beam mixing of ti thin films on si. *Vide, les Couches Minces*, 42(236):141–143, 1987. cited By 10.

[11] K. Maex, R.F. De Keersmaecker, M. Van Rossum, W.F. Van Der Weg, and G. Krooshof. The structure and composition of layers obtained by ion beam mixing in the ti/si and co/si systems. *Nuclear Inst. and Methods in Physics Research, B*, 19-20(PART 2):731–736, 1987. cited By 8.

[12] A.I.C. Morão, E. Zondervan, G. Krooshof, R. Geertman, and A.B. de Haan. Synthesis tool for separation processes in the pharmaceutical industry. *Computer Aided Chemical Engineering*, 29:276–280, 2011. cited By 1.

[13] J. Tinge, J. Krijgsman, T. Smeets, P. Vergossen, G. Krooshof, and E. Hoving. Dewatering caprolactam solutions by direct osmosis using osmotic pressure of concentrated aqueous ammonium sulphate solutions. *Desalination*, 199(1-3):312–314, 2006. cited By 0.

[14] J.T. Tinge, G.J.P. Krooshof, T.M. Smeets, F.H.P. Vergossen, J. Krijgsman, E. Hoving, and R.M. Altink. Direct osmosis membrane process to de-water aqueous caprolactam with concentrated aqueous ammonium sulphate. *Chemical Engineering and Processing: Process Intensification*, 46(6):505–512, 2007. cited By 6.

[15] T. Tork, G. Sadowski, W. Arlt, A. De Haan, and G. Krooshof. Modelling of high-pressure phase equilibria using the sako-wu-prausnitz equation of state: I. pure-components and heavy n-alkane solutions. *Fluid Phase Equilibria*, 163(1):61–77, 1999. cited By 6.

[16] T. Tork, G. Sadowski, W. Arlt, A. De Haan, and G. Krooshof. Modelling of high-pressure phase equilibria using the sako-wu-prausnitz equation of state: Ii. vapour-liquid equilibria and liquid-liquid equilibria in polyolefin systems. *Fluid Phase Equilibria*, 163(1):79–98, 1999. cited By 38.

[17] R. Tuinier and G.J.P. Krooshof. Interfacial tension between benzene and water in the presence of caprolactam. *Journal of Colloid and Interface Science*, 382(1):105–109, 2012. cited By 4.

[18] G.H. Van Bochove, G.J.P. Krooshof, and T.W. De Loos. Modelling of liquid-liquid equilibria of mixed solvent electrolyte systems using the extended electrolyte nrtl. *Fluid Phase Equilibria*, 171(1-2):45–58, 2000. cited By 42.

[19] G.H. Van Bochove, G.J.P. Krooshof, and T.W. De Loos. Two- and three-liquid phase equilibria in the system water + 2-heptanone + caprolactam + ammonium sulfate: Experiments and modeling. *Fluid Phase Equilibria*, 194-197:1029–1044, 2002. cited By 16.

[20] R.E. Van Vliet, T.P. Tiemersma, G.J. Krooshof, and P.D. Iedema. The use of liquid-liquid extraction in the epdm solution polymerization process. *Industrial and Engineering Chemistry Research*, 40(21):4586–4595, 2001. cited By 10.

[21] C. Versluis, G. Krooshof, and F. Van Gulick. Thermal-impact damage of polyamide and polyester yarns. *Textile Research Journal*, 67(3):211–216, 1997. cited By 0.

Patents

[1] G.W.A. Hangx, G.J.P. Krooshof, and A. de Rijke. EU Patent PCTEP2006012307: Process for the separation of formic acid from methyltetrahydrofuran., 2016.

[2] S. Jahromi, G.J.P. Krooshof, A.L. Quiceno-Gomez, H. Suzuki, and G. Hoffmann. EU Patent PCT/EP2008/005514: A laminate and composite layer comprising a substrate and coating, and a proces and apparatus for preparation thereof., 2009.

[3] A.V. Peters, G.J.P. Krooshof, N.M.H. Beckers, and J. Krijgsman. US Patent 12160349: Process for the isolation of an organic amine., 2010.