Integrated process and solvent design for CO_2 capture using Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD)

Stavrou, Marina

DOI
10.4233/uuid:b44de3d4-25dd-453c-988c-07fca027a612

Publication date
2017

Document Version
Final published version

Citation (APA)
Stavrou, M. (2017). *Integrated process and solvent design for CO_2 capture using Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD).* [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:b44de3d4-25dd-453c-988c-07fca027a612

Integrated process and solvent design for CO_2 capture using Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD)

Integrated process and solvent design for CO_2 capture using Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD)

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben;
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
donderdag 06 juli 2017 om 10:00 uur

door

Marina-Eleni STAVROU

Dipl. Chemical Engineer,
National Technical University of Athens, Athens, Greece
Master of Science in Process Energy and Environmental Systems Engineering,
Technische Universität Berlin, Germany
geboren te Athene, Griekenland

This dissertation has been approved by the

promotor: Prof. Dr.- Ing. J. Gross
promotor: Prof. Dr.- Ing. A. Bardow

Composition of the doctoral committee:

Rector Magnificus	chairman
Prof. Dr.- Ing. J. Gross	Delft University of Technology / University of Stuttgart, promotor
Prof. Dr.- Ing. A. Bardow	RWTH Aachen, promotor
Independent members:	
Prof. dr. ir. M. T. Kreuzer	Delft University of Technology
Prof. dr. ir. P. Colonna	Delft University of Technology
Prof. dr. ir. A. I. Stankiewicz	Delft University of Technology
Prof. ir. J. Grievink	em. hgl. Delft University of Technology
Other member:	
Dr. K. Damen	Energy research Centre of the Netherlands

The work documented in this thesis has been performed within the CO_2 Catch-up R&D programme aimed at demonstrating and optimizing pre-combustion CO_2 capture technology for the energy sector. This programme is executed in a consortium of Vattenfall, the Delft University of Technology and the Energy research Centre of the Netherlands.

Published by Marina-Eleni Stavrou, Stuttgart

Printed by Baier Digitaldruck in Heidelberg

to my mother

Contents

Summary

The cost of currently available technologies for CO_2 capture should be further reduced to allow for large scale implementation of Carbon Capture and Storage. Solvents for CO_2 capture systems with physical absorption are usually selected based on heuristics, engineering expertise and experimental trials. The performance of the separation system is, however, defined by both the properties of the selected solvent and the process conditions, which should be considered simultaneously. In this thesis, the Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD) framework is extended and applied to the simultaneous optimization of process and solvent for CO_2 capture systems with physical absorption.

In Chapter 2 of the thesis, the CoMT-CAMD framework is applied for a pre-combustion CO_2 capture system. The problem of simultaneous process and solvent optimization is solved for a reasonably complex process model and for an objective function defining the overall process performance. The solvent is represented as the pure component parameters of the Perturbed Chain Statistical Associating Fluid Theory (PC-SAFT). The optimization is formulated with the pure component parameters of the solvent (PC-SAFT parameters) and with the process variables as degrees of freedom. Preceding studies have shown that process inequality constraints can be problematic for CoMT-CAMD. A constraint for the solvent for example was considered in this chapter. This work proposes a Taylor-expansion of the process constraints in the direction of the parameters of the PC-SAFT model. QSPR models for the prediction of pure component properties (ideal gas heat capacity and molar mass) as function of the PC-SAFT pure component parameters were developed. These models, together with the PC-SAFT equation, enabled predictions of full caloric properties and mass specific process streams inside the CoMT-CAMD framework. Optimal solvents that achieve a minimization of the primary energy consumption in all stages of the capture process were successfully identified.

Chapter 3 presents a method for the prediction of binary interaction parameters k_{ij} of the PC-SAFT equation of state. The correction required when PC-SAFT is applied to mixtures is linked to the level of asymmetry in the intermolecular potentials of the mixture components. Asymmetry in intermolecular potentials is quantified using relations between the PC-SAFT parameters of the pure components present in the mixture. Estimates of binary interaction parameters k_{ij} are thus possibly based only on (combinations of) pure component parameters. A QSPR method is employed and estimations of k_{ij} are obtained independent from experimental mixture data. The method is implemented to mixtures with non-associating and to mixtures with associating components.

Chapter 4 uses and assesses the model for estimating k_{ij}-values in the CoMT-CAMD framework. Phase equilibria for mixtures of the optimized fluid can be described more accurately. For the assessment of the k_{ij}-model, the case study of

CoMT-CAMD for polar solvents was examined. For the most promising solvents resulting from the CoMT-CAMD optimization, k_{ij}-values were individually identified by adjusting to experimental data and the process conditions for these solvents were optimized. That made it possible to compare the results to the CoMT-CAMD calculations with the predictive QSPR-model for k_{ij}. The comparison shows that CoMT-CAMD with our QSPR-model for k_{ij} gives good estimates for the results of individually optimized substances. An improvement is found over the model without binary correction (k_{ij} = 0).

Samenvatting

De kosten van de huidige beschikbare technologieën voor de afvang van koolstofdioxide (CO_2) moeten verder worden gereduceerd, eerde implementatie van CO_2-afvang en -opslag (Engels, Carbon Capture and Storage) op grote schaalkan worden gerealiseerd. De oplosmiddelen voor CO_2-afvang systemen door middel van fysische absorptie worden doorgaans geselecteerd op basis van een heuristische methode, technische expertise of experimentele studies. De prestaties vanhet scheidingssysteemworden echter bepaald door zowel de eigenschappen van het gekozen oplosmiddel als de procescondities. Deze dienentegelijkertijd in ogenschouw te worden genomen. In dit proefschrift, wordt de Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD) methode uitgebreid en toegepast op de gelijktijdige optimalisatie van het proces en het oplosmiddel voor CO_2-afvang systemen middels een fysisch oplosmiddel.

In hoofdstuk 2 van het proefschrift wordt de CoMT-CAMD methode aangewend voor een voorverbranding (Engels, pre-combustion) CO_2-afvang systeem. Het simultaan optimaliseren van het proces en het oplosmiddel is gedaan aan de hand van een tamelijk complex procesmodel en een doelfunctie die de globale procesprestaties definieert. Het oplosmiddel wordt gerepresenteerd door de zuivere component parameters van de geperturbeerde-keten statistische associërende vloeistoftheorie Perturbed Chain Statistical Associating Fluid Theory (PC-SAFT). De optimalisatie wordt opgesteld met de zuivere componenten parameters van het oplosmiddel (PC-SAFT parameters) en de procesvariabelen als vrijheidsgraden. Voorgaande studies hebben aangetoond dat randvoorwaarden in de vorm van ongelijkheden problematisch kunnen zijn voor CoMT-CAMD. In dit hoofdstuk werd bijvoorbeeld een randvoorwaarde voor het oplosmiddel in overweging genomen. Dit werk stelt een Taylor-expansie voor in de vorm van de procesrandvoorwaarden in de richting van de PC-SAFT parameters. QSPR modellen werden ontwikkeld voor het voorspellen van de eigenschappen van dezuivere componenten (warmtecapaciteit van ideaal gas en molaire massa) als functie van de zuivere componenten parameters van PC-SAFT. Deze modellen, samen met de PC-SAFT toestandsvergelijking, maakten het mogelijk om voorspellingen te doen van de volledige calorische eigenschappen en massa specifieke processtromen binnen de CoMT-CAMD methode. De meest optimale oplosmiddelen, die in alle fasen van het afvangproces het primaire energieverbruik minimaliseren, werden op deze manier succesvol geïdentificeerd.

Hoofdstuk 3 presenteert een methode voor het voorspellen van de binaire interactie parameters k_{ij} van de PC-SAFT toestandsvergelijking. De vereiste correctie, indien PC-SAFT wordt toegepast op mengsels, hangt samen met de mate van asymmetrie in de intermoleculaire potentialen van de mengselcomponenten. De asymmetrie in de intermoleculaire potentialen wordt gekwantificeerd middels verbanden tussen de PC-SAFT parameters van de pure componenten in het mengsel.

Schattingen van de binaire interactie parameters k_{ij} zijn dus mogelijkerwijs alleen gebaseerd op (combinaties van) pure componenten parameters. Een QSPR methode wordt toegepast om k_{ij} te schatten onafhankelijk van de experimentele data van het mengsel. Deze aanpak wordt geïmplementeerd voor zowel mengsels met niet-associërende als mengsels met associërende componenten.

In hoofdstuk 4 wordt het model toegepast en geëvalueerd voor het schatten van de k_{ij}-waarden in het CoMT-CAMD kader. Fasenevenwichten van mengsels van het geoptimaliseerde fluïdum kunnenhiermee nauwkeuriger worden beschreven. Voor de beoordeling van het k_{ij}-model, werd de case study van CoMT-CAMD voor polaire oplosmiddelen onderzocht. Voor de meest veelbelovende oplosmiddelen die uit de CoMT-CAMD optimalisatie voortvloeien, werden individuele k_{ij}-waarden geïdentificeerddoor ze aan te passen aan de experimentele data. Vervolgens werden de procescondities voor deze oplosmiddelen geoptimaliseerd. Dit maakte het mogelijk om de resultaten van de CoMT-CAMD berekeningen te vergelijken met het voorspellende QSPR-model voor k_{ij}. De vergelijking toont aan dat CoMT-CAMD met het QSPR-model goede schattingen van k_{ij} geeft voor de resultaten van de individueel geoptimaliseerde stoffen. Een verbetering is geconstateerd ten opzichte van het model dat geen gebruik maakt van de binaire correctie (k_{ij} = 0).

1

Introduction

1

Pre-combustion Carbon Capture and Storage (CCS) with physical absorption for CO_2 capture is seen as a cost-competitive low carbon emission concept for power plants. [1, 2] According to public data of the MIT Energy Initiative ('CCS Project Database' [3]): 43% of the power plant CCS projects planned to be operated worldwide consider Integrated Gasification Combined Cycle (IGCC) power plants and employ pre-combustion CO_2 capture with physical absorption. Physical absorption has long since been employed for acid gas removal from industrial gas streams [4–7]. Still, the large-scale implementation of CCS systems requires further reduction in the cost of the currently available technologies for carbon capture. [8] In that light, interrelated environmental and financial performance targets should be considered. In power plants, the performance of the carbon capture system affects directly the production price of electrical power, being crucial to the viability of the entire CCS project. [1, 9–14] Simultaneous process and solvent optimization can lead to lower-cost designs for carbon capture systems with physical absorption. The optimization of absorption systems for CO_2 capture has become a vibrant research topic. [8, 15–18]

The Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD) [19] approach enables the simultaneous process and solvent optimization with a single process-driven objective function, utilizing a physically-based thermodynamic model, the Perturbed Chain Statistical Associating Fluid Theory (PC-SAFT) [20–23]. The current thesis proposes extensions to the CoMT-CAMD framework for an overall performance optimization of a pre-combustion CO_2 capture system with physical absorption.

The process topology of the carbon capture system examined in this work resembles in its' key features the pre-combustion capture pilot plant at the IGCC power plant in Buggenum, the Willem Alexander Centrale (WAC). The current work has been part of the test and R&D program of the CO_2 Catch-up project for the demonstration of pre-combustion CO_2 capture at the site of the IGCC pilot plant in Buggenum. The test and R&D program of the CO_2 Catch-up project has been managed by Vattenfall R&D Projects and performed together with Delft University of Technology and Energy research Center of the Netherlands (ECN).

1.1. Solvents for CO_2 capture using physical absorption at IGCC power plants

Physical absorption is a mature technology and appropriate for the pre-combustion capture of CO_2 at IGCC plants. [10, 16, 24–27] A simplified process scheme for an IGCC pre-combustion capture concept is given in Figure 1.1. An example of an acid

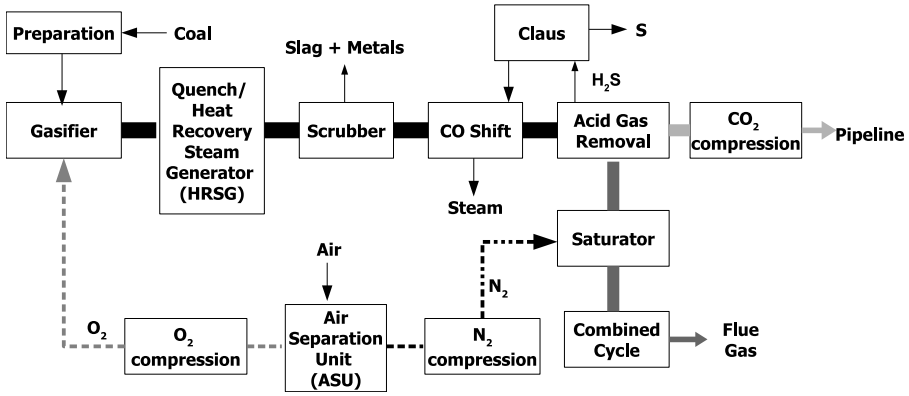


Figure 1.1: Simplified process scheme for an IGCC pre-combustion capture concept (adopted from Kunze and Spliethoff [28]).

gas removal (AGR) process for the pre-combustion capture concept is illustrated in Figure 1.2. The high partial pressure (or better: fugacity) of CO_2 in the shifted syngas stream of an IGCC process allows for high solubility of CO_2 in the solvent (absorption step). The lean solvent can be recovered through pressure reduction in flash drums (desorption step). [29] The desorption step for physical solvents is more energy efficient and cost effective, than the regeneration step of chemical solvents in systems for post-combustion CO_2 capture [15, 30, 31].

The most common industrially applied solvents for the physical absorption of CO_2 are methanol, N-methyl-2-pyrrolidone (NMP), propylene carbonate, 4-formylmorpholine and mixtures of polyethylene glycols of dimethylethers (DEPEGs). A short review about CO_2 separation processes using these solvents is given in the following.

Methanol is an inexpensive chemical and can achieve sharp separations at high pressures and low absorption temperatures. [33] Methanol is used in the Rectisol process, the earliest commercially applied process for acid gas removal. [4] Due to low operating temperatures, the Rectisol process includes solvent refrigeration steps. This leads to high primary energy consumption and to complex process designs, leading to high operating and investment costs. The Rectisol process was initially implemented in the production of syngas in the chemical industry. [4, 29, 34] For the implementation to CCS systems in IGCC power plants, the original Rectisol process scheme has to be modified. [33] Improved Rectisol-based processes with heat integration have recently been studied by Gatti et al. [34].

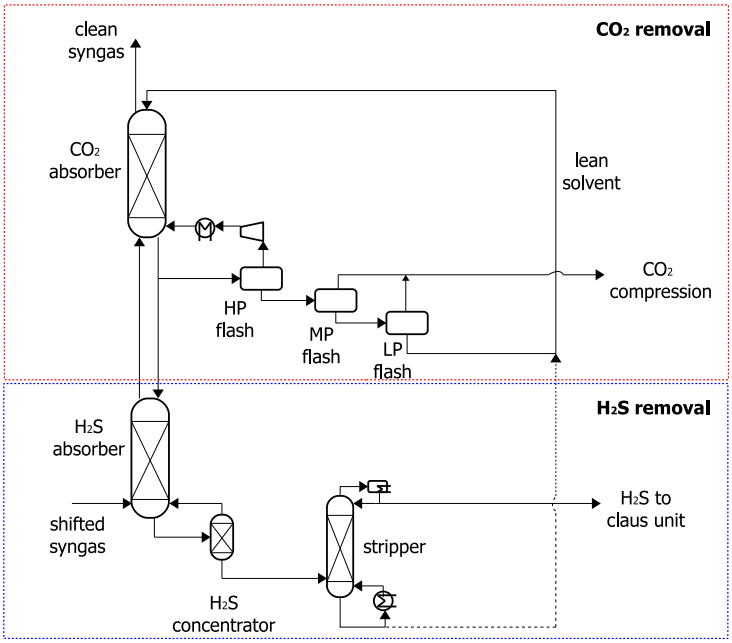


Figure 1.2: Acid gas removal process for the IGCC base case presented in Fig.1.1: Simplified flowsheet of the Selexol process for the removal of H_2S and CO_2, as described in ref. [32]. The solvent recovery from the CO_2 capturing step is carried out through pressure reduction in flash drums.

N-methyl-2-pyrrolidone (NMP) is a polar solvent, used in the Purisol process. Due to the higher selectivity of NMP towards H_2S, the Purisol process has been mainly implemented in natural gas sweetening processes. [4, 29] Additionally, NMP has been studied as a component of hybrid solvents (in mixtures with alkanolamines) for CO_2 capture. [35]

Propylene carbonate, also a polar organic solvent, is used in the Fluor Solvent process. Propylene carbonate has a higher selectivity towards CO_2 than towards light hydrocarbons or H_2. [4] The Fluor process has been mainly employed in natural gas processing. Recently, propylene carbonate has been studied as solvent for the absorption of CO_2 in a membrane gas absorption process. [36]

4-formylmorpholine, in mixture with N-acetyl-morpholine, is used in the Morphysorb process. The Morphysorb process requires low operating temperatures due to the low boiling point of the solvent. The process has been recently applied and tested for the CO_2 capture for natural gas purification. [37] 4-formylmorpholine has also been studied as solvent in a membrane gas absorption process. [36]

Finally, mixtures of polyethylene glycols of dimethylethers (DEPGs) are used in the Selexol process. The Selexol process has been already implemented for the selective removal of H_2S from flue gas streams that also contain CO_2. [4] In order to achieve high purity product streams of H_2, two Selexol units are usually employed for the sequential removal of H_2S and CO_2. Hydrocarbons of high molecular mass

are more soluble in DEPGs than CO_2 and H_2S. [4] Therefore, the Selexol process is not suitable for acid gas removal from rich hydrocarbon streams. According to a recent study of Kapetaki et al. [38], the implementation of the Selexol process in IGCC power plants may lead to CO_2 capture efficiencies[1] higher than 90%.

The field of research for superior solvents for CO_2 capture with physical absorption is constantly developing. Perfluorinated compounds (PFCs) [39–41] and siloxanes [42, 43] have been studied as promising solvents. Another prominent research field is the design of ionic liquids (IL). [44–49]

Still, for the large-scale implementation of CCS systems a new generation of designs for carbon capture is required . New designs should achieve significant cost reduction for the overall capture process. [31, 50] In this regard, the employed solvents should facilitate high selectivity for absorbing CO_2 against H_2, while they allow for a minimization of the primary energy consumption in all stages of the capture process. [31]

1.2. Solvent and process optimization with Computer Aided Molecular Design (CAMD)

The selection of physical solvents for CO_2 capture is often based on heuristics: Physical properties of the solvent, like CO_2 solubility, CO_2/H_2S selectivity, vapor pressure, thermal stability, viscosity, toxicity and corrosivity are evaluated for given process conditions. [39–41] The decision about the appropriate solvent is thus based on engineering expertise, prior knowledge of the process and experimental trials. [51] In industrial practice, the choice of solvents becomes therefore a solvent evaluation procedure rather than a true solvent optimization.

Alternatively, the solvent optimization problem can be addressed as a reverse engineering problem and be solved using Computer Aided Molecular Design (CAMD). [52, 53] The goal of the reverse engineering problem is to identify solvent molecules that meet certain property targets or, more generally, to identify the solvent molecules that achieve optimal performance regarding specified process measures. CAMD provides computational methods and systematic frameworks for the solution of this problem.

1.2.1. Property models in CAMD

Property models are essential in CAMD for the prediction of pure component parameters[2] and for the prediction of state properties[3] of pure components and mixtures.

[1] Carbon capture efficiency = $\frac{\text{Carbon in } CO_2 \text{ product}}{\text{Carbon in coal} - \text{Carbon in slag}}$ (according to ref. [38])

[2] The term 'pure component parameters' is used here to describe physical properties like molar mass, critical properties, normal boiling and melting points, the standard enthalpy and standard Gibbs energy of formation. [54]

[3] Here, the term 'state properties' refers to the thermal, caloric or transport properties of pure components and mixtures (internal energy, enthalpy, entropy, Helmholtz energy and Gibbs energy) and their derivative properties: a) temperature dependent properties (vapor pressure, liquid density, enthalpy of vaporization, ideal gas heat capacity, liquid heat capacity, surface tension and speed of sound) [54], b) transport properties (liquid and vapor viscosity, liquid and vapor thermal conductivity and diffusion coefficients) [54] and c) properties for phase equilibrium calculations (chemical potential, fugacity and

Predicted properties are used to identify the molecules that meet the property targets of the design problem. The accuracy and predictive power of the employed property model(s) define the accuracy and plausibility of the CAMD solution. [55–58] Therefore, modeling of physical properties plays a critical role in CAMD.
The property models employed in CAMD are mainly Group-Contribution (GC) methods [59–63] or Quantitative Structure Property Relationship (QSPR) models with molecular signature descriptors [64–66].

Strictly speaking, physical properties can be distinguished in pure component parameters and state properties. [54] In the CAMD literature, it is common to speak about 'primary properties' [67], as properties estimated using information only about the molecular structure. 'Primary properties' include pure component parameters and certain temperature dependent state properties at a reference temperature. The GC methods proposed by Joback and Reid [68] have been used in CAMD (e.g. ref. [69, 70]) for the estimation of 'primary properties' of pure components (i.e. the normal boiling and freezing point temperatures, critical properties, the ideal gas heat capacity and the enthalpy of vaporization at boiling point temperature). The GC methods of Joback and Reid are based on first order (UNIFAC) functional groups and were extended by Constantinou and Gani [60]. Constantinou and Gani included second-order functional groups in order to capture the differences between isomers and proximity effects. [71] Further, Marrero and Gani [62] introduced GC methods with third-order functional groups for a better description of complex heterocyclic and large poly-functional acyclic compounds (e.g. ref. [72]). Constantinou et al. [61] developed a GC method for the estimation of acentric factors. GC methods are also available for the prediction of properties that characterize the eco-toxicity of pure components. [73–75]
State properties are calculated from estimated values of 'primary properties', using basic thermodynamic relations or empirical correlations. [71, 76–78] For state properties, some works (e.g. ref. [56, 79]) use the theorem of corresponding states with GC predictions for the acentric factor and the critical properties. For phase equilibrium calculations g^E-models (usually UNIFAC) are often used or, less often, cubic equations of state (e.g. ref. [80, 81]).

Methods that use topological indices (TI) as descriptors in QSPR models are employed for the prediction of both pure component parameters and state properties. Topological indices are a particular category of molecular signature descriptors, which are derived from molecular graphs. According to Faulon et al. [82]: "The (molecular) signature is a systematic codification system over an alphabet of atom types, describing the extended valence (i.e. neighborhood) of the atoms of a molecule." Topological indices based on molecular signature take into account the molecular architecture and serve as suitable descriptors for a QSPR model for property prediction. [64] Raman and Maranas [64] implemented such QSPR models for the prediction of the boiling point temperature and the critical properties and for the prediction of the molar volume, the enthalpy of vaporization, the molar refractivity and the surface tension at a given temperature. Chemmangattuvalappil and Eden [65] used QSPR models with topological indices for the prediction of the molar

fugacity coefficients, activity and activity coefficients) [54].

volume [83] and the toxic limit concentration [84], while Patel et al. [66] proposed a QSPR model with topological indices for the prediction of flash point in CAMD. A thorough review on the QSPR methods using topological indices in CAMD can be found in the work of Harini et al. [57].

A combination of QSPR models with topological indices and GC methods, the GC$^+$ method [67], has been also implemented in CAMD. The GC$^+$ method improves property predictions for complex molecules and is implemented in a way similar to the original GC methods. For example, Samudra and Sahinidis [76] used GC$^+$ for the prediction of 'primary properties' and external property models for the calculation of heat conductivities and the volumetric heat capacity. In the work of Conte et al. [85] the GC$^+$ method was used for the prediction of surface tension and viscosity. Gonzalez et al. [86] used GC$^+$ to predict properties relevant to phase equilibrium calculations. Based on the GC$^+$ method, Hukkerikar et al. [87] proposed estimation methods for environment-related properties.

O'Connell et al. [56] classified the property models in CAMD according to the way these models are obtained and the level of empiricism they implicate (Figure 1.3). GC and QSPR models use information about the molecular structure in a sys-

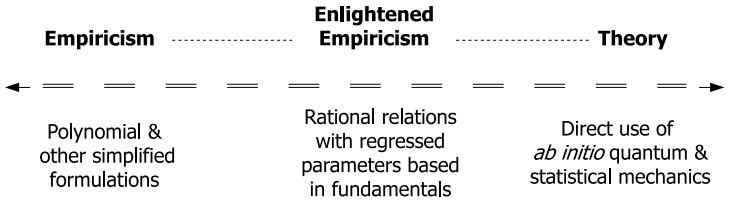


Figure 1.3: Iterative steps of property model development (adopted from O'Connell et al. [56]).

tematic way. Still, they are correlations with parameters adjusted to experimental data. As stated by Raman and Maranas [64], the QSPR models used in CAMD "do not necessarily have a casual relationship with the correlated property". In complex CAMD problems, depending on the required properties, the combination of different (GC or QSPR) models may be necessary as well. In that case, the model assumptions, the validity range of the employed models and combined uncertainties should be carefully examined. A recent review on the advantages and limitations of the various fluid property models used in CAMD is given by van Speybroeck et al. [88].

Physically-based thermodynamic models have been recently introduced in CAMD frameworks. Physically-based thermodynamic models (for example, SAFT-type models [89]) originate in statistical mechanics and have a theoretical background (right end of the scale in Figure 1.3) but also parameters. Due to the physical background of their molecular parameters, SAFT-type models give a sharp representation of the optimized molecule inside CAMD frameworks. SAFT-type models can be better extrapolated and therefore can provide safer predictions for state properties than (other) GC methods and QSPR models. With a SAFT-type equation of state (EoS) residual static state properties can be derived directly from the residual Helmholtz energy. Adjiman et al. [90] illustrate the expanding potential of the implementation

1

of SAFT-type models in CAMD. A more detailed discussion about the key-role of the PC-SAFT [20–23] model as a unified framework for property prediction in CAMD is given in Sections 1.2.4 and 1.3.1 of the current thesis.

1.2.2. The integrated solvent and process optimization problem

Solvent optimization problems are particularly demanding due to the strong correlations between the optimal properties of the solvent and the optimal process conditions. In order to achieve an optimal performance for the CO_2 capture process, the solvent design problem should thus not be decoupled and solved separately from the process design. [19, 52, 90] The solvent and the process conditions have to be optimized simultaneously. An integrated process and molecular design approach is essential.

The fully integrated process and solvent design problem is in its original formulation a mixed integer non-linear optimization program (MINLP) [53]:

$$\begin{aligned} & \min_{\overline{x},\overline{p}} f(\overline{x},\overline{p}) \\ \text{s.t.} \quad & h(\overline{x},\overline{p}) = 0 \\ & g(\overline{x},\overline{p}) \leq 0 \end{aligned} \tag{1.1}$$

where $\overline{x}$ denotes the vector of process variables defined in $\mathbb{R}^m$ and $\overline{p} = [p_1, \dots, p_D]$ denotes the vector of molecular parameters of the optimized solvent. The molecular parameters p_i are parameters required by the employed property model: the functional groups of a GC property model, the descriptors of a QSPR model or the molecular parameters of a physically-based (SAFT-type) thermodynamic model. The vector of molecular parameters $\overline{p}$ is unique for each real substance and is defined over a discrete vector space, $P^{D\times N}$, where D is the number of molecular parameters p_i and N is the number of all real substances. The non-linear objective function $f(\overline{x},\overline{p})$ is a representative measure of the overall process performance. The non-linear constraints of the problem $h(\overline{x},\overline{p})$ and $g(\overline{x},\overline{p})$ represent the equations of the process model, the equations of the property model and the constraints on $\overline{x}$ and $\overline{p}$.

The non-linear objective function $f(\overline{x},\overline{p})$, the non-linear property models in $h(\overline{x},\overline{p})$ and the high dimensionality of the discrete vector space $P^{D\times N}$ lead to a mixed-integer non-linear program of prohibitive size for rigorous and large process models. Therefore, a solution of problem (1.1) is attainable only through problem relaxation. CAMD methods can be distinguished according to the relaxation strategy they employ in order to handle problem (1.1).

1.2.3. CAMD methods for solvent selection and solvent design

In solvent design, most CAMD approaches treat the process and the molecular design problems separately. Some approaches start with a pre-selection of suitable candidate species (e.g. ref. [91] and [92]). In a subsequent step, the process is optimized individually for each candidate solvent, in order to identify the best

performing solvent. In other CAMD approaches [59, 60, 62, 93] the property targets and the property constraints are defined for given process conditions. The molecular structures that meet the thus defined property targets and constraints are identified through database screening. The considered property databases include either already existing components or molecular structures generated using a GC method [53]. In order to avoid the mixed-integer formulation of the molecular design problem, these approaches decouple the solution of the process design from the solution of the molecular design problem. All aforementioned CAMD approaches involve heuristic decisions, either in the step of pre-selection or by the definition of property targets. The joint search space of the optimal solution of problem (1.1) is then decomposed, leading inevitably to loss of information and sub-optimal solutions.

Eden et al. [52] introduced the concept of property integration in a CAMD approach for the simultaneous optimization of separation processes and solvents. Property integration allows for process-specific property targets and establishes an indirect link between the two sub-problems of process and molecular design. [19] A description of the approach proposed by Eden et al. [52] is given in Figure 1.4. CAMD approaches based on the concept of property integration preserve the mixed-

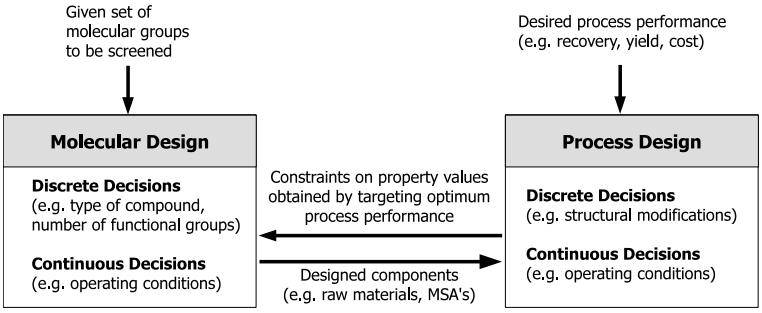


Figure 1.4: Reverse problem formulation with property integration for simultaneous separation process and product design (adopted from Eden et al. [52]).

integer formulation of the molecular design problem in (1.1). Still, the solution of the MINLP in (1.1) requires a problem relaxation. Problem (1.1) is therefore further relaxed, in order either to handle the non-linearity of the process design problem, or in order to reduce the dimension of the discrete molecular search space. Auxiliary objective functions (e.g. ref. [94]) or simplified process and property models (e.g. ref. [52]) are implemented to relax the high non-linearity in $f(\overline{x},\overline{p})$, $h(\overline{x},\overline{p})$ and $g(\overline{x},\overline{p})$. A reduction of the discrete molecular search space $P^{D\times N}$ is attained through pre-selection of certain candidate species (molecules or the functional groups of a GC model, e.g. ref. [95]). Finally, many approaches turn to advanced algorithmic techniques for the solution of the modified MI(N)LP. A detailed discussion on these approaches is given in Chapter 2.

Consider, however, the level of complexity in industrial separation processes,

like the process of capturing CO_2 for a CCS application. The implementation of simplified process models and auxiliary objective functions may neglect important interactions between the various process steps. Thereby, critical aspects of the full process performance will be missed, which may lead to fallacious property targets and consequently to sub-optimal or even erroneous solutions for the molecular design problem. Moreover, by implementing an *a-priori* reduction of the molecular search space, superior solvents may be omitted.

1.2.4. The Continuous Molecular Targeting (CoMT) - CAMD method

Continuous Molecular Targeting (CoMT), established by Bardow et al. [19], is an alternative CAMD method for the solution of the integrated solvent and molecular optimization problem (1.1). CoMT-CAMD circumvents the mixed-integer formulation of the molecular optimization problem in problem (1.1). The discrete molecular search space of the thermodynamic model representing real molecules $P^{D\times N}$ is regarded as a continuous molecular search space of real and hypothetical molecules $\mathbb{R}^N$. The molecular and process optimization variables are all defined in the continuous domain $\mathbb{R}^{m+N}$ and they can be optimized simultaneously. Problem (1.1) is thereby formulated as a non-linear program (NLP) and it can be solved using gradient based optimization algorithms. CoMT is empowered by the use of a physically-based thermodynamic model. In this work, the CoMT-CAMD approach is implemented with the Perturbed Chain - Statistical Associating Fluid Theory (PC-SAFT) model [20–23]. The PC-SAFT molecular parameters are used as variables for the molecular optimization.

Process and molecular optimization take place simultaneously using a single, well-defined objective function for the process (like annual costs or primary energy consumption). Individual property targets and auxiliary objective functions are not required. A single, process-based objective function simultaneously evaluates the impact of changes in the molecular structure of the solvent and of the process conditions to the overall process performance. Additionally, the transformation of problem (1.1) to a non-linear program (NLP) allows for rigorously considering (reasonably complex) process models.

The optimal solution obtained in the CoMT step is the vector of the PC-SAFT molecular parameters of the optimal, hypothetical, molecule and the corresponding optimal process conditions. The molecular design is completed by identifying the real components (or more generally the feasible molecular structures), which give the best achievable performance, closest to the optimal one. The efficiency of the molecular design step in CoMT-CAMD is founded on the strong physical background of the PC-SAFT molecular parameters: In PC-SAFT, similar molecular parameters correspond to similar thermodynamic properties. The performance of real components is estimated by an approximation of the objective function around the optimum of the CoMT step. The objective function is approximated using a Taylor series of 2^{nd} order as function of the molecular parameters. The best performing real molecules are expected to lie close to the local optimum of problem (1.1) in the joint search space of process variables and molecular parameters. The set of real

components that are considered during the step of molecular design are either existing components with known molecular parameters (database mapping) [96, 97] or molecular structures generated using a GC method for the PC-SAFT parameters like GPC-SAFT [98] (feasible structure mapping [99]). In the case that molecular structures are generated using a GC method, a molecular design algorithm with feasibility constraints has to be implemented. [99] The generic procedure of CoMT-CAMD is illustrated in Figure 1.5.

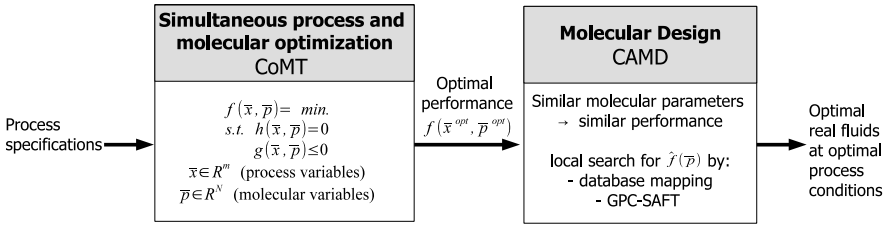


Figure 1.5: Continuous Molecular Targeting - Computer Aided Molecular Design for integrated process and fluid optimization.

1.3. The Perturbed Chain - SAFT (PC-SAFT) thermodynamic model

1.3.1. A unified framework for the calculation of physical properties in CAMD

The PC-SAFT equation of state (EoS) [20–23] formulates the residual Helmholtz energy. Any other residual (static) state property can be derived from the residual Helmholtz energy. For the calculation of full caloric properties, an ideal gas contribution is additionally required. In process simulation problems, the ideal gas heat capacity c_p^{ig} can be obtained directly from property databases, since the system components are known. During the Continuous Molecular Targeting, values of c_p^{ig} for the optimized (hypothetical) molecule are not available in databases. Further, the molecular mass of the optimized fluid is required in process calculations that include mass-specific process streams. The QSPR models developed in the course of the current thesis for the estimation of c_p^{ig} and the molecular mass of pure substances as function of the PC-SAFT pure component parameters are given in Chapter 2.

Beyond the calculation of caloric properties, the PC-SAFT model can be utilized for the prediction of several static and dynamic state properties: Surface tension for pure components and mixtures can be predicted with PC-SAFT in the framework of density functional theory (DFT). [100–102] Viscosity of pure fluids and fluid mixtures can be predicted using Rosenfeld's entropy-scaling [103, 104] and a GC method for the PC-SAFT parameters. [105] On-going work focuses on the development of models for the prediction of thermal conductivity [106, 107] and

self-diffusion coefficients.
The implementation of PC-SAFT EoS to mixtures often demands the adjustment of a binary interaction parameter, k_{ij}, to experimental data of phase equilibria. Since experimental mixture data are not accessible during molecular design, k_{ij} needs to be predicted from pure component parameters. This thesis shows that the binary interaction parameter k_{ij} can be predicted from the pure component PC-SAFT parameters using a QSPR model (Chapter 3).

The PC-SAFT EoS originates in statistical mechanics and accounts explicitly for polar and associating intermolecular interactions. [20, 21] PC-SAFT provides good predictions of pure component and mixture properties and gives better predictions in phase equilibrium calculations for complex and high pressure systems than cubic EoS or g^E-models. [108, 109] Due to the theoretical background of the model parameters, PC-SAFT links thermodynamic properties of pure fluids to molecular structure. Further, full static and dynamic state properties required in process systems' calculations can be derived from residual properties. Thereby, PC-SAFT serves as a unified framework for the calculation of thermodynamic properties in the CoMT-CAMD framework.

1.3.2. The PC-SAFT equation of state

The PC-SAFT EoS belongs to the SAFT family of equations of state. SAFT-type EoS are algebraic approximations of the residual Helmholtz energy. SAFT-type EoS are based on statistical mechanics and more specifically on Wertheim's thermodynamic perturbation theory of first-order (TPT1) [110, 111]. Perturbation theories provide models for fluid properties based on intermolecular potentials. [112, 113] A reference fluid is considered in order to describe fluid behavior resulting from repulsive intermolecular forces. Attractive intermolecular forces are treated as perturbation to the behavior of the reference fluid. Molecular distribution functions characterizing the repulsive part of the intermolecular potential are used. In the thermodynamic perturbation theory (TPT) of Wertheim the residual Helmholtz energy is expanded in a series of integrals of molecular distribution functions for the reference fluid and the association potential. [112] Chapman et al. [114] applied Wertheim's theory to mixtures and derived the Statistical Associating Fluid Theory (SAFT) EoS. In their work, Chapman et al. [114] use a reference fluid of homonuclear hard-sphere chains. The contribution to the residual Helmholtz energy due to association is derived from Wertheim's perturbation theory of first order (TPT1). In the same work [114], the effects due to dispersive forces and induction are accounted through a mean field perturbation term [112].

PC-SAFT [20–23] also utilizes the reference fluid of hard-sphere chains. Gross and Sadowski [20] applied the perturbation theory of Barker and Henderson [115] to chain molecules, to calculate the contribution to the residual Helmholtz energy due to dispersive forces. The contribution due to associating interactions is applied in a similar way as in the work of Chapman et al. [112] and Huang and Radosz [116]. The contribution of quadrupole-quadrupole, dipole-dipole and dipole-quadrupole interactions to the residual Helmholtz energy are accounted for, as described in the work of Gross [22], Gross and Vrabec [23] and Vrabec and Gross [117] respectively.

PC-SAFT is implemented to mixtures using the one-fluid (van der Waals) mixing rules.

In the PC-SAFT model, the molecules are represented as chains of spherical segments. Each pure substance is identified by a vector of molecular parameters characterizing the shape of the chain and the dispersive, associating and polar intermolecular potentials. Non-polar and non-associating molecules are characterized by the number of segments per chain molecule m (segment number), the segment diameter σ and the dispersive energy parameter ε/k. Associating molecules need two additional parameters, the association energy parameter $\varepsilon^{\mathrm{AB}}/k$ and the association volume κ^{AB}. Polar molecules require the dipole moment μ and the quadrupole moment Q as parameters.

The equations of the PC-SAFT model, that are required for calculation of thermophysical properties in the present work, are given in the following.

Helmholtz Energy

The following equations are summarized in form of the residual Helmholtz energy related to the number of chain molecules N and their kinetic energy kT as

$$\tilde{a} = \frac{A}{NkT} \tag{1.2}$$

where k is the Boltzmann constant and T the temperature. The reduced Helmholtz energy of a real fluid is therefore written as

$$\tilde{a} = \tilde{a}^{\mathrm{ig}} + \tilde{a}^{\mathrm{res}} \tag{1.3}$$

The reduced Helmholtz energy for a mixture in the ideal gas state is given as

$$\tilde{a}^{\mathrm{ig}} = \sum_i x_i \left[\ln\left(\Lambda^3 \cdot \rho_i\right) - 1\right] = \frac{1}{\rho} \cdot \sum_i \rho_i \left[\ln\left(\Lambda^3 \cdot \rho_i\right) - 1\right] \tag{1.4}$$

where Λ is the de Broglie wavelength, x_i is the mole fraction of species i in the mixture, ρ_i the molecular density of species i and ρ the density of the mixture.
The residual term of the Helmholtz energy is calculated with PC-SAFT. The reduced residual Helmholtz energy $\tilde{a}^{\mathrm{res}}$ is expressed as a sum of terms due to the hard-chain reference contribution $\tilde{a}^{\mathrm{hc}}$ and the terms expressing the contribution due to attractive interactions: the dispersive $\tilde{a}^{\mathrm{disp}}$, associating $\tilde{a}^{\mathrm{assoc}}$, quadrupolar $\tilde{a}^{\mathrm{qq}}$ and dipolar $\tilde{a}^{\mathrm{dd}}$ contribution terms, according to

$$\tilde{a}^{res} = \tilde{a}^{\mathrm{hc}} + \tilde{a}^{\mathrm{disp}} + \tilde{a}^{\mathrm{assoc}} + \tilde{a}^{\mathrm{qq}} + \tilde{a}^{\mathrm{dd}} \tag{1.5}$$

The equations for the calculation of the various contributions to the residual Helmholtz energy are adopted from ref. [20–23, 112] and they are given in Appendix A.1.

Derived thermodynamic properties from the Helmholtz energy

Compressibility factor. The compressibility factor Z is written as:

$$Z = Z^{\mathrm{ig}} + Z^{\mathrm{hc}} + Z^{\mathrm{disp}} + Z^{\mathrm{assoc}} + Z^{\mathrm{qq}} + Z^{\mathrm{dd}} = 1 + \eta \left(\frac{\partial \tilde{a}^{\mathrm{res}}}{\partial \eta}\right)_{T,\bar{x}}. \tag{1.6}$$

1

where η is the packing fraction defined in Appendix A.1. The equations for the calculation of the various contributions to the compressibility are given in ref. [20–23, 112].

Pressure. The total pressure P of the system can be calculated as:

$$P = ZkT\rho \tag{1.7}$$

using the compressibility factor Z given in Eq. (1.6).

Chemical potential, μ_i. The chemical potential μ_i of species i is the partial derivative of the Helmholtz energy A to the mole number n_i for constant temperature T and volume V:

$$\mu_i(T,V,\overline{n}) = \left(\frac{\partial A}{\partial n_i}\right)_{T,V,n_{j\neq i}}. \tag{1.8}$$

For the chemical potential μ_i of species i Eq.(1.8) leads to the following formulation:

$$\frac{\mu_i(T,\overline{\rho})}{kT} = \left(\frac{\partial\,(\tilde{a}\cdot\rho)}{\partial\rho_i}\right)_{T,\rho_{j\neq i}} \tag{1.9}$$

where $\overline{\rho}$ is the vector of component molar densities $\rho_i = x_i \cdot \rho$ of all species i of the system. For the full chemical potential of a species i holds

$$\frac{\mu_i}{kT} = \frac{\mu_i^{\mathrm{ig}}}{kT} + \frac{\mu_i^{\mathrm{res}}}{kT}. \tag{1.10}$$

From Eq.(1.4) and Eq.(1.8), for $\Lambda^3 = 1$, the chemical potential of species i in the ideal gas state is calculated by

$$\frac{\mu_i^{\mathrm{ig}}}{kT} = \ln(\Lambda^3\cdot\rho_i) = \ln(\rho_i). \tag{1.11}$$

The residual chemical potential μ_i^{res} of species i is calculated with the PC-SAFT EoS according to

$$\frac{\mu_i^{\mathrm{res}}}{kT} = \rho\cdot\left(\frac{\partial\tilde{a}^{\mathrm{res}}}{\partial\rho_i}\right)_{T,\rho_{j\neq i}} + \tilde{a}^{\mathrm{res}}. \tag{1.12}$$

Fugacity coefficient $\ln\phi_i$. The fugacity coefficient of species i is expressed as function of the temperature T and the pressure P as

$$\ln\phi_i(T,P,\overline{x}) = \frac{\mu_i^{\mathrm{res}}(T,P,\overline{x})}{kT} = \frac{\mu_i^{\mathrm{res}}(T,\overline{\rho})}{kT} - \ln Z. \tag{1.13}$$

Residual molar entropy, $\hat{s}^{\text{res}}$. In the following, molar thermodynamic properties $\hat{w}$ are related to the dimensionless properties as

$$\frac{\hat{w}}{RT} = \tilde{w} = \frac{W}{NkT}. \tag{1.14}$$

From the Maxwell relations the entropy S is given as the partial derivative of the Helmholtz energy A over the temperature T for constant volume V and constant composition $\overline{x}$ of the mixture

$$\left(\frac{\partial A}{\partial T}\right)_{V,\overline{x}} = -S. \tag{1.15}$$

Eq.(1.15) can be reformulated on the reduced Helmholtz energy $\tilde{a}$ as

$$\left(\frac{\partial \left(A/Nk\right)}{\partial T}\right)_{\overline{\rho}} = -\frac{S}{Nk} \quad \text{which gives} \quad \left(\frac{\partial \left(\tilde{a}\cdot T\right)}{\partial T}\right)_{\overline{\rho}} = -\frac{\hat{s}}{R}. \tag{1.16}$$

The residual molar entropy of a mixture of molar volume v is given by

$$\begin{aligned}\frac{\hat{s}^{\text{res}}(v,T,\overline{x})}{R} = \frac{\hat{s}^{\text{res}}(T,\overline{\rho})}{R} &= -T\cdot\left(\frac{\partial \tilde{a}^{\text{res}}}{\partial T}\right)_{\overline{\rho}} - \tilde{a}^{\text{res}} = \\ &= -T\cdot\left[\left(\frac{\partial \tilde{a}^{\text{res}}}{\partial T}\right)_{\overline{\rho}} + \frac{\tilde{a}^{\text{res}}}{T}\right]\end{aligned} \tag{1.17}$$

and

$$\frac{\hat{s}^{\text{res}}(P,T,\overline{x})}{R} = \frac{\hat{s}^{\text{res}}(T,\overline{\rho})}{R} + \ln Z. \tag{1.18}$$

Molar enthalpy, $\hat{h}$. The molar enthalpy of the mixture $\hat{h}$ is given as the summation of the ideal gas term $\hat{h}^{\text{ig}}$ and the residual term $\hat{h}^{\text{res}}$:

$$\hat{h} = \hat{h}^{\text{ig}} + \hat{h}^{\text{res}}. \tag{1.19}$$

The enthalpy of the mixture in the ideal gas state, for a reference temperature T^o, is given according to classical thermodynamics by

$$\hat{h}^{\text{ig}}(T,\overline{x}) = \sum_i x_i \hat{h}_i^{\text{ig}}(T) \tag{1.20}$$

$$\text{with} \quad \hat{h}_i^{\text{ig}}(T) - \hat{h}_i^{\text{ig}}(T^o) = \int_{T^o}^{T} \hat{c}_{p,i}^{\text{ig}}\mathrm{d}T. \tag{1.21}$$

The residual molar enthalpy $\hat{h}^{\text{res}}$ is calculated from the reduced residual Helmholtz energy and the compressibility factor. The relation between the enthalpy H and the Helmholtz energy A written for residual properties is

$$H^{\text{res}} = A^{\text{res}} - TS^{\text{res}} + (PV)^{\text{res}} = A^{\text{res}} - TS^{\text{res}} + \left[PV - (PV)^{\text{ig}}\right]. \tag{1.22}$$

Eq.(1.7) and (1.22) result to the following expressions for the residual molar enthalpy:

$$\frac{\hat{h}^{\mathrm{res}}}{RT} = \frac{H^{\mathrm{res}}}{NkT} = \tilde{a}^{\mathrm{res}} - \frac{\hat{s}^{\mathrm{res}}}{R} + (Z-1) \tag{1.23}$$

and

$$\frac{\hat{h}^{\mathrm{res}}}{RT} = -T \cdot \left(\frac{\partial \tilde{a}^{\mathrm{res}}}{\partial T}\right)_{\overline{\rho}} + (Z-1). \tag{1.24}$$

Phase equilibrium calculations

For a system with K components and M phases in equilibrium, the equilibrium conditions write:

$$T^{\mathrm{I}} = T^{\mathrm{II}} = \cdots = T^{\mathrm{M}} \tag{1.25}$$

$$P^{\mathrm{I}} = P^{\mathrm{II}} = \cdots = P^{\mathrm{M}} \tag{1.26}$$

$$\mu_i^{\mathrm{I}} = \mu_i^{\mathrm{II}} = \cdots = \mu_i^{\mathrm{M}} \quad \forall i. \tag{1.27}$$

The chemical potential μ_i may be expressed by the fugacity f_i. [54] For a reference pressure P^o

$$\mu_i(T,P,\overline{x}) = \mu_i^{\mathrm{ig,pure}}(T,P^o) + RT\ln\frac{f_i(T,P,\overline{x})}{P^o}. \tag{1.28}$$

Then, the equilibrium condition in Eq.(1.27) can be expressed by means of fugacity as

$$f_i^{\mathrm{I}} = f_i^{\mathrm{II}} = \cdots = f_i^{\mathrm{M}} \quad \forall i. \tag{1.29}$$

Using the definition of the fugacity coefficient ϕ_i of species i [118]

$$\phi_i = \frac{f_i}{x_i P} \tag{1.30}$$

the isofugacity relations in Eq.(1.29) can be written as

$$x_i^{\mathrm{I}}\phi_i(T,P,x_i^{\mathrm{I}}) = x_i^{\mathrm{II}}\phi_i(T,P,x_i^{\mathrm{II}}) = \cdots = x_i^{\mathrm{M}}\phi_i(T,P,x_i^{\mathrm{M}}) \quad \forall i. \tag{1.31}$$

In phase equilibrium calculations the equations (1.31) are solved iteratively using the PC-SAFT EoS for the calculation of the fugacity coefficients (Eq.(1.12) and (1.13)).

The application of PC-SAFT to highly asymmetric mixtures (mixtures of components with considerably different intermolecular potentials) requires the adjustment of a binary correction parameter k_{ij} to experimental mixture data.

Figure 1.6 shows results of phase equilibrium calculations with PC-SAFT for the binary systems CO_2 - acetophenone and H_2 - cyclohexane. In both examples, we observe a good description of the vapor phase. An accurate description of the liquid phase is obtained when the model is corrected with k_{ij} adjusted to experimental mixture data. However, experimental data are not available for mixtures considered during molecular design. A method for the estimation of k_{ij} as a function of the PC-SAFT pure component parameters is proposed and implemented with CoMT-CAMD in the current thesis.

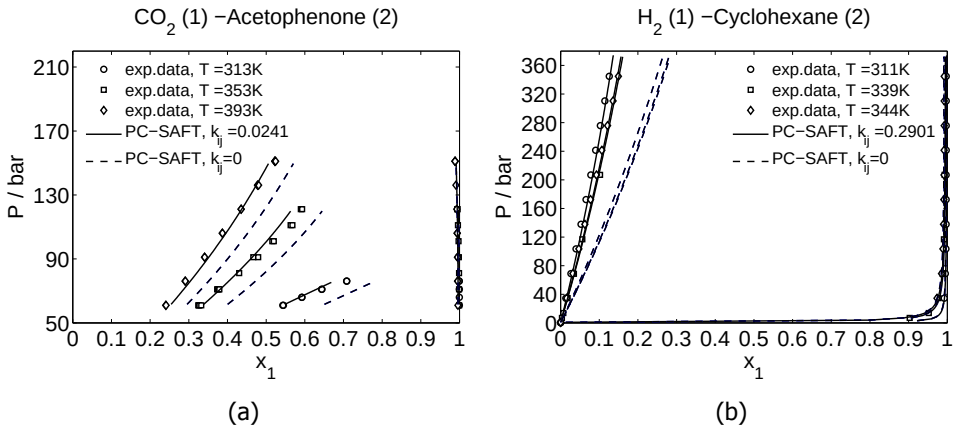


Figure 1.6: Results of vapor-liquid equilibrium calculations with PC-SAFT for asymmetric binary systems: a) CO_2 - acetophenone [119] and b) H_2 - cyclohexane [120, 121]. Phase equilibrium calculations are compared to isothermal (P-xy) experimental data. Solid lines denote phase equilibrium calculations with binary interaction parameter k_{ij} adjusted on the experimental data and dashed lines are for $k_{ij} = 0$.

1.4. Outline of the thesis

An overall performance optimization of CO_2 capture systems with physical absorption is attainable when all major process trade-offs are taken into account. For this, the solvent should be optimized simultaneously with the process conditions, using a single, process-based objective function.

In this work the solution of the integrated solvent and process optimization problem is approached with the CoMT-CAMD framework, using PC-SAFT. In CoMT-CAMD, the optimization problem is formulated as a non-linear program. The process and the solvent are described through continuous variables and are optimized simultaneously. The optimal solution of the process and solvent optimization problem is obtained without pre-selection of candidate solvents and heuristic objectives.

The work in the current thesis unfolds in two levels: a) the implementation of the CoMT-CAMD framework for the solution of the problem of optimal solvent selection for a pre-combustion CO_2 capture system and b) the development and implementation of correlation models for the prediction of auxiliary properties of the solvent, aiming to enhance the accuracy in calculations inside CoMT-CAMD.

In Chapter 2, the CoMT-CAMD method is presented in detail. The implementation and application of CoMT-CAMD is developed for the case of simultaneous solvent and process optimization of a CO_2 capture process with physical absorption for an IGCC power plant. Calculation of full caloric properties in CoMT-CAMD requires predictions of the ideal gas heat capacity c_p^{ig} of the optimized solvent. The calculation of mass specific quantities, such as specific density, requires predictions of the molar mass. QSPR models for c_p^{ig} and the molar mass as function of the PC-SAFT pure component parameters were developed and integrated in the CoMT-CAMD framework.

With CoMT-CAMD, a set of optimal solvents is identified from a large database of organic molecules, without pre-selection of candidates and for a reasonably complex process model. The set of optimal solvents include both state-of-the-art and new solvents, a fact underlining the validity and the strength of the CoMT-CAMD method.

The accuracy of PC-SAFT in calculations of state properties is decisive to the plausibility of the CoMT-CAMD solution. Especially for mixtures with highly unlike components, a correction (k_{ij}) of the PC-SAFT EoS is usually required. A new method for the prediction of k_{ij}, independent of experimental mixture data and based solely on the PC-SAFT parameters of the pure components of the mixture, is presented in Chapter 3.

Chapter 4 analyzes the effect of predicted binary interaction parameter k_{ij} on the resulting list of promising candidate solvents from the CoMT-CAMD approach. The evaluation is done for the case of CO_2 capture as detailed in Chapter 2. The result of the mapping step ('mapping list') obtained using k_{ij} predictions is compared to the result of the mapping step when no correction is used for the binary mixtures of the solvent ($k_{ij} = 0$). Further, the result of the CoMT-CAMD is also compared to results from individual process optimizations with k_{ij} values adjusted to experimental data.

Chapter 5 assesses the results of the current work and discusses some ideas for further extensions to the CoMT-CAMD framework.

References

[1] M. Finkenrath, *Cost and performance of carbon dioxide capture from power generation*, Tech. Rep. (International Energy Agancy, 2011).

[2] L. Tock and F. Marechal, *Thermo-environomic optimisation strategy for fuel decarbonisation process design and analysis,* Comput. Chem. Eng. , (2015).

[3] MIT, *Energy Initiative Carbon Capture & Sequestration Technologies: Carbon Capture and Sequestration Project Database* http://sequestration.mit.edu/tools/projects/index_capture.html, (07/2015).

[4] A. L. Kohl and R. B. Nielsen, *Gas Purification* (Gulf Publishing Company, 1997).

[5] O. R. Rivas and J. M. Prausnitz, *Sweetening of sour natural gases by mixed-solvent absorption: Solubilities of ethane, carbon dioxide, and hydrogen sulfide in mixtures of physical and chemical solvents,* AIChE J. **25**, 975 (1979).

[6] R. Bucklin and R. Schendel, *Comparison of Fluor Solvent and Selexol processes,* Energy Prog.; (United States) **4:3** (1984).

[7] R. W. Rousseau, J. N. Matange, and J. K. Ferrell, *Solubilities of carbon dioxide, hydrogen sulfide and nitrogen mixtures in methanol,* AIChE J. **27**, 605 (1981).

[8] R. Aldous, C. Anderson, R. Anderson, M. Gerstenberger, B. Gurevich, B. Hooper, C. Jenkins, J. Kaldi, S. Kentish, V. Linton, S. Santos, and P. Webley, *CSLF Technology Assessment, CCS Technology Development; Gaps, Opportunities and Research Fronts*, Tech. Rep. (Cooperative Research Centre for Greenhouse Gas Technologies, 2013).

[9] P. H. M. Feron and C. A. Hendriks, CO_2 *capture process principles and costs,* Oil Gas Sci. Technol. **60**, 451 (2005).

[10] K. Damen, M. van Troost, A. Faaij, and W. Turkenburg, *A comparison of electricity and hydrogen production systems with CO_2 capture and storage. Part A: Review and selection of promising conversion and capture technologies,* Prog. Energy Combust. Sci. **32**, 215 (2006).

[11] H. Zhai and E. S. Rubin, *Comparative performance and cost assessments of coal- and natural-gas-fired power plants under a CO_2 emission performance standard regulation,* Energy Fuels **27**, 4290 (2013).

[12] E. S. Rubin, J. E. Davison, and H. J. Herzog, *The cost of CO_2 capture and storage,* Int. J. Greenh. Gas Control **40**, 378 (2015), special Issue commemorating the 10th year anniversary of the publication of the Intergovernmental Panel on Climate Change Special Report on CO_2 Capture and Storage.

[13] K. Damen, A. Faaij, and W. Turkenburg, *Pathways towards large-scale implementation of CO_2 capture and storage: A case study for the Netherlands,* Int. J. Greenh. Gas Control **3**, 217 (2009).

[14] T. Bruckner, I. A. Bashmakov, Y. Mulugetta, H. Chum, A. de la Vega Navarro, J. Edmonds, A. Faaij, B. Fungtammasan, A. Garg, E. Hertwich, D. Honnery, D. Infield, M. Kainuma, S. Khennas, S. Kim, H. B. Nimir, K. Riahi, N. Strachan, R. Wiser, and X. Zhang, *2014: Energy systems. in: Climate change 2014: Mitigation of climate change. contribution of working group III to the fifth assessment report of the intergovernmental panel on climate change,* in *Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E.Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel, and J. Minx (Cambridge University Press, 2014).

[15] H. J. Herzog, *Peer reviewed: What future for carbon capture and sequestration?* Environ. Sci. Technol. **35**, 148A (2001).

[16] D. M. D'Alessandro, B. Smit, and J. R. Long, *Carbon dioxide capture: Prospects for new materials,* Angewandte Chemie International Edition **49**, 6058 (2010).

[17] National Carbon Capture Centre, *http://www.nationalcarboncapturecenter.com/pre-combustion-co2-capture.htm*, (10/2015).

[18] National Energy Technology Laboratory U.S. Department of Energy, *http://www.netl.doe.gov/research/coal/carbon-capture/pre-combustion*, (10/2015).

[19] A. Bardow, K. Steur, and J. Gross, *Continuous-Molecular Targeting for integrated solvent and process design,* Ind. Eng. Chem. Res. **49**, 2834 (2010).

[20] 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**, 1244 (2001).

[21] J. Gross and G. Sadowski, *Application of the Perturbed-Chain SAFT equation of state to associating systems,* Ind. Eng. Chem. Res. **41**, 5510 (2002).

[22] J. Gross, *An equation-of-state contribution for polar components: Quadrupolar molecules,* AIChE J. **51**, 2556 (2005).

[23] J. Gross and J. Vrabec, *An equation-of-state contribution for polar components: Dipolar molecules,* AIChE J. **52**, 1194 (2006).

[24] M. Kanniche, R. Gros-Bonnivard, P. Jaud, J. Valle-Marcos, J.-M. Amann, and C. Bouallou, *Pre-combustion, post-combustion and oxy-combustion in thermal power plant for CO_2 capture,* Appl. Therm. Eng. **30**, 53 (2010).

[25] A. Padurean, C.-C. Cormos, and P.-S. Agachi, *Pre-combustion carbon dioxide capture by gas-liquid absorption for integrated gasification combined cycle power plants,* Int. J. Greenh. Gas Control **7**, 1 (2012).

[26] C.-C. Cormos, *Integrated assessment of IGCC power generation technology with carbon capture and storage (CCS),* Energy **42**, 434 (2012).

[27] D. Jansen, M. Gazzani, G. Manzolini, E. van Dijk, and M. Carbo, *Pre-combustion CO_2 capture,* Int. J. Greenh. Gas Control **40**, 167 (2015).

[28] C. Kunze and H. Spliethoff, *Assessment of oxy-fuel, pre- and post-combustion-based carbon capture for future IGCC plants,* Appl. Energy **94**, 109 (2012).

[29] G. Hochgesand, *Rectisol and Purisol,* Ind. Eng. Chem. **62**, 37 (1970).

[30] J. Gibbins and H. Chalmers, *Carbon capture and storage,* Energy Policy **36**, 4317 (2008).

[31] J. D. Figueroa, T. Fout, S. Plasynski, H. McIlvried, and R. D. Srivastava, *Advances in CO_2 capture technology- the U.S. Department of Energy's Carbon Sequestration Program,* Int. J. Greenh. Gas Control **2**, 9 (2008).

[32] R. P. Field and R. Brasington, *Baseline flowsheet model for IGCC with Carbon Capture,* Ind. Eng. Chem. Res. **50**, 11306 (2011).

[33] L. Sun and R. Smith, *Rectisol wash process simulation and analysis,* J. Clean Prod. **39**, 321 (2013).

[34] M. Gatti, E. Martelli, F. Marechal, and S. Consonni, *Review, modeling, heat integration, and improved schemes of Rectisol®-based processes for CO_2 capture,* Appl. Therm. Eng. **70**, 1123 (2014).

[35] F. Murrieta-Guevara, E. Rebolledo-Libreros, and A. Trejo, *Solubility of carbon dioxide in binary mixtures of N-methylpyrrolidone with alkanolamines,* J. Chem. Eng. Data **37**, 4 (1992).

[36] V. Dindore, D. Brilman, F. Geuzebroek, and G. Versteeg, *Membrane-solvent selection for CO_2 removal using membrane gas-liquid contactors,* Sep. Purif. Technol. **40**, 133 (2004).

[37] N. Palla and D. Leppin, *Technical and Operating Support for Pilot Demonstration of Morphysorb Acid Gas Removal Process*, Tech. Rep. (Gas Technology Institute, 2004).

[38] Z. Kapetaki, P. Brandani, S. Brandani, and H. Ahn, *Process simulation of a dual-stage selexol process for 95% carbon capture efficiency at an integrated gasification combined cycle power plant,* Int. J. Greenh. Gas Control **39**, 17 (2015).

[39] B. Gwinner, D. Roizard, F. Lapicque, E. Favre, R. Cadours, P. Boucot, and P.-L. Carrette, *CO_2 capture in flue gas: Semiempirical approach to select a potential physical solvent,* Ind. Eng. Chem. Res. **45**, 5044 (2006).

[40] Y. J. Heintz, L. Sehabiague, B. I. Morsi, K. L. Jones, and H. W. Pennline, *Novel physical solvents for selective CO_2 capture from fuel gas streams at elevated pressures and temperatures ,* Energy Fuels **22**, 3824 (2008).

[41] H. W. Pennline, D. R. Luebke, K. L. Jones, C. R. Myers, B. I. Morsi, Y. J. Heintz, and J. B. Ilconich, *Progress in carbon dioxide capture and separation research for gasification-based power generation point sources,* Fuel Process. Technol. **89**, 897 (2008).

[42] M. B. Miller, D.-L. Chen, H.-B. Xie, D. R. Luebke, J. K. Johnson, and R. M. Enick, *Solubility of CO_2 in CO_2-philic oligomers; COSMOtherm predictions and experimental results,* Fluid Phase Equilib. **287**, 26 (2009).

[43] M. B. Miller, D. R. Luebke, and R. M. Enick, *CO_2-philic oligomers as novel solvents for CO_2 absorption,* Energy Fuels **24**, 6214 (2010).

[44] Y. J. Heintz, L. Sehabiague, B. I. Morsi, K. L. Jones, D. R. Luebke, and H. W. Pennline, *Hydrogen sulfide and carbon dioxide removal from dry fuel gas streams using an ionic liquid as a physical solvent,* Energy Fuels **23**, 4822 (2009).

[45] O. M. Basha, M. J. Keller, D. R. Luebke, K. P. Resnik, and B. I. Morsi, *Development of a conceptual process for selective CO_2 capture from fuel gas streams using [hmim][Tf2N] ionic liquid as a physical solvent,* Energy Fuels **27**, 3905 (2013).

[46] M. Ramdin, S. P. Balaji, J. M. Vicent-Luna, J. J. Gutiarrez-Sevillano, S. Calero, T. W. de Loos, and T. J. H. Vlugt, *Solubility of the precombustion gases CO_2, CH_4, CO, H_2, N_2, and H_2S in the ionic liquid [bmim][Tf2N] from Monte Carlo simulations,* J. Phys. Chem. C **118**, 23599 (2014).

[47] M. Ramdin, A. Amplianitis, S. Bazhenov, A. Volkov, V. Volkov, T. J. Vlugt, and T. W. de Loos, *Solubility of CO_2 and CH_4 in ionic liquids: Ideal CO_2/CH_4 selectivity,* Ind. Eng. Chem. Res. **53**, 15427 (2014).

[48] F. Chong, D. Foo, F. Eljack, M. Atilhan, and N. Chemmangattuvalappil, *Ionic liquid design for enhanced carbon dioxide capture by computer-aided molecular design approach,* Clean Technol. Envir. **17**, 1301 (2015).

[49] L. F. Zubeir, G. E. Romanos, W. M. A. Weggemans, B. Iliev, T. J. S. Schubert, and M. C. Kroon, *Solubility and diffusivity of CO_2 in the ionic liquid 1-butyl-3-methylimidazolium tricyanomethanide within a large pressure range (0.01 MPa to 10 MPa),* J. Chem. Eng. Data **60**, 1544 (2015).

[50] K. A. Mumford, Y. Wu, K. H. Smith, and G. W. Stevens, *Review of solvent based carbon-dioxide capture technologies,* Front. Chem. Sci. Eng. **9**, 125 (2015).

[51] B. V. Smith and M. G. Ierapepritou, *Integrative chemical product design strategies: Reflecting industry trends and challenges,* Comput. Chem. Eng. **34**, 857 (2010).

[52] M. Eden, S. Jorgensen, R. Gani, and M. El-Halwagi, *A novel framework for simultaneous separation process and product design,* Chem. Eng. Process. Process Intensif. **43**, 595 (2004).

[53] R. Gani, *Chemical product design: challenges and opportunities,* Comput. Chem. Eng. **28**, 2441 (2004).

[54] J. Gmehling, B. Kolbe, M. Kleiber, and J. Rarey, *Chemical thermodynamics for process simulation* (Wiley-VCH Verlag & Co. KGaA, 2012).

[55] N. Churi, , and L. E. K. Achenie, *Novel mathematical programming model for computer aided molecular design,* Ind. Eng. Chem. Res. **35**, 3788 (1996).

[56] J. P. O'Connell, R. Gani, P. M. Mathias, G. Maurer, J. D. Olson, and P. A. Crafts, *Thermodynamic property modeling for chemical process and product engineering: Some perspectives,* Ind. Eng. Chem. Res. **48**, 4619 (2009).

[57] M. Harini, J. Adhikari, and K. Y. Rani, *A review on property estimation methods and computational schemes for rational solvent design: A focus on pharmaceuticals,* Ind. Eng. Chem. Res. **52**, 6869 (2013).

[58] L. Y. Ng, F. K. Chong, and N. G. Chemmangattuvalappil, *Challenges and opportunities in Computer Aided Molecular Design,* Comput. Chem. Eng. **81**, 115 (2015).

[59] R. Gani, B. Nielsen, and A. Fredenslund, *A group contribution approach to computer-aided molecular design,* AIChE J. **37**, 1318 (1991).

[60] L. Constantinou and R. Gani, *New group contribution method for estimating properties of pure compounds,* AIChE J. **40**, 1697 (1994).

[61] 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 Equilib. **103**, 11 (1995).

[62] J. Marrero and R. Gani, *Group-contribution based estimation of pure component properties,* Fluid Phase Equilib. **183-184**, 183 (2001).

[63] T. Sheldon, C. Adjiman, and J. Cordiner, *Pure component properties from group contribution: Hydrogen-bond basicity, hydrogen-bond acidity, hildebrand solubility parameter, macroscopic surface tension, dipole moment, refractive index and dielectric constant,* Fluid Phase Equilib. **231**, 27 (2005).

[64] V. Raman and C. D. Maranas, *Optimization in product design with properties correlated with topological indices,* Comput. Chem. Eng. **22**, 747 (1998).

[65] N. G. Chemmangattuvalappil and M. R. Eden, *A novel methodology for property-based molecular design using multiple topological indices,* Ind. Eng. Chem. Res. **52**, 7090 (2013).

[66] S. J. Patel, D. Ng, and M. S. Mannan, *QSPR flash point prediction of solvents using topological indices for application in computer aided molecular design,* Ind. Eng. Chem. Res. **48**, 7378 (2009).

[67] R. Gani, P. M. Harper, and M. Hostrup, *Automatic creation of missing groups through connectivity index for pure-component property prediction,* Ind. Eng. Chem. Res. **44**, 7262 (2005).

[68] K. Joback and R. Reid, *Estimation of pure-component properties from group-contributions,* Chem. Eng. Commun. **57**, 233 (1987).

[69] N. V. Sahinidis, M. Tawarmalani, and M. Yu, *Design of alternative refrigerants via global optimization,* AIChE J. **49**, 1761 (2003).

[70] W. Xu and U. M. Diwekar, *Improved genetic algorithms for deterministic optimization and optimization under uncertainty. Part II. Solvent selection under uncertainty,* Ind. Eng. Chem. Res. **44**, 7138 (2005).

[71] M. Sinha and L. E. Achenie, *Systematic design of blanket wash solvents with recovery considerations,* Adv. Environ. Res. **5**, 239 (2001).

[72] M. Mattei, G. M. Kontogeorgis, and R. Gani, *A comprehensive framework for surfactant selection and design for emulsion based chemical product design,* Fluid Phase Equilib. **362**, 288 (2014).

[73] C. Gao, R. Govind, and H. H. Tabak, *Application of the group contribution method for predicting the toxicity of organic chemicals,* Environ. Toxicol. Chem. **11**, 631 (1992).

[74] S.-T. Lin and S. I. Sandler, *Multipole corrections to account for structure and proximity effects in group contribution methods:Octanol-water partition coefficients,* J. Phys. Chem. A **104**, 7099 (2000).

[75] J. Marrero and R. Gani, *Group-contribution-based estimation of octanol/water partition coefficient and aqueous solubility,* Ind. Eng. Chem. Res. **41**, 6623 (2002).

[76] A. P. Samudra and N. V. Sahinidis, *Optimization-based framework for computer-aided molecular design,* AIChE J. **59**, 3686 (2013).

[77] L. Y. Ng, N. G. Chemmangattuvalappil, and D. K. S. Ng, *A multiobjective optimization-based approach for optimal chemical product design,* Ind. Eng. Chem. Res. **53**, 17429 (2014).

[78] O. Palma-Flores, A. Flores-Tlacuahuac, and G. Canseco-Melchor, *Optimal molecular design of working fluids for sustainable low-temperature energy recovery,* Comput. Chem. Eng. **72**, 334 (2015).

[79] F. T. Eljack, M. R. Eden, V. Kazantzi, X. Qin, and M. M. El-Halwagi, *Simultaneous process and molecular design- A property based approach,* AIChE J. **53**, 1232 (2007).

[80] A. T. Karunanithi, L. E. K. Achenie, and R. Gani, *A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures,* Ind. Eng. Chem. Res. **44**, 4785 (2005).

[81] J. M. Valdez-Patrinos, R. Vazquez-Roman, E. Moran-Corona, J. A. Inchaurregui-Mendez, and E. Quiroz-Perez, *Computer aided molecular design for undefined petroleum fractions,* Fluid Phase Equilib. **390**, 14 (2015).

[82] J.-L. Faulon, J. Donald P. Visco, and R. S. Pophale, *The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSPR studies,* J. Chem. Inf. Comput. Sci. **43**, 707 (2003).

[83] E. Estrada, *Edge adjacency relationships and a novel topological index related to molecular volume,* J. Chem. Inf. Comput. Sci. **35**, 31 (1995).

[84] R. Koch, *Molecular connectivity and acute toxicity of environmental pollutants,* Chemosphere **11**, 925 (1982).

[85] E. Conte, A. Martinho, H. A. Matos, and R. Gani, *Combined group-contribution and atom connectivity index-based methods for estimation of surface tension and viscosity,* Ind. Eng. Chem. Res. **47**, 7940 (2008).

[86] H. E. Gonzalez, J. Abildskov, R. Gani, P. Rousseaux, and B. Le Bert, *A method for prediction of UNIFAC group interaction parameters,* AIChE J. **53**, 1620 (2007).

[87] A. S. Hukkerikar, S. Kalakul, B. Sarup, D. M. Young, G. Sin, and R. Gani, *Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution+ (GC+) property models and uncertainty analysis,* Journal of Chemical Information and Modeling **52**, 2823 (2012).

[88] V. van Speybroeck, R. Gani, and R. J. Meier, *The calculation of thermodynamic properties of molecules,* Chem. Soc. Rev. **39**, 1764 (2010).

[89] L. F. Vega and G. Jackson, *20 years of the SAFT equation of state-recent advances and challenges: Symposium held in Bellaterra, Barcelona, 19-21 September 2010,* Fluid Phase Equilib. **306**, 1 (2011).

[90] C. S. Adjiman, A. Galindo, and G. Jackson, *Molecules matter: The expanding envelope of process design,* Comput. Chem. Eng. **34**, 55 (2014).

[91] E. J. Pretel, P. A. Lopez, S. B. Bottini, and E. A. Brignole, *Computer-aided molecular design of solvents for separation processes,* AIChE J. **40**, 1349 (1994).

[92] M. Cismondi and E. A. Brignole, *Molecular design of solvents: An efficient search algorithm for branched molecules,* Ind. Eng. Chem. Res. **43**, 784 (2004).

[93] P. M. Harper and R. Gani, *A multi-step and multi-level approach for computer aided molecular design,* Comput. Chem. Eng. **24**, 677 (2000).

[94] A. I. Papadopoulos and P. Linke, *Integrated solvent and process selection for separation and reactive separation systems,* Chem. Eng. Process. **48**, 1047 (2009).

[95] N. G. Chemmangattuvalappil, C. C. Solvason, S. Bommareddy, and M. R. Eden, *Combined property clustering and GC+ techniques for process and product design,* Comput. Chem. Eng. **34**, 582 (2010).

[96] M. Lampe, M. Stavrou, H. M. Bucker, J. Gross, and A. Bardow, *Simultaneous optimization of working fluid and process for organic rankine cycles using PC-SAFT,* Ind. Eng. Chem. Res. **53**, 8821 (2014).

[97] M. Stavrou, M. Lampe, A. Bardow, and J. Gross, *Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for CO_2 capture,* Ind. Eng. Chem. Res. **53**, 18029 (2014).

[98] E. Sauer, M. Stavrou, and J. Gross, *Comparison between a Homo- and a Heterosegmented Group Contribution approach based on the Perturbed-Chain Polar Statistical Associating Fluid Theory Equation of State,* Ind. Eng. Chem. Res. **53**, 14854 (2014).

[99] M. Lampe, M. Stavrou, J. Schilling, E. Sauer, J. Gross, and A. Bardow, *Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT,* Comput. Chem. Eng. **81**, 278 (2015).

[100] J. Gross, *A density functional theory for vapor-liquid interfaces using the PCP-SAFT equation of state,* J. Chem. Phys. **131**, (2009).

[101] C. Klink and J. Gross, *A density functional theory for vapor-liquid interfaces of mixtures using the Perturbed-Chain Polar Statistical Associating Fluid Theory equation of state,* Ind. Eng. Chem. Res. **53**, 6169 (2014).

[102] C. Klink, B. Plankova, and J. Gross, *Density functional theory for liquid-liquid interfaces of mixtures using the Perturbed-Chain Polar Statistical Associating Fluid Theory equation of state,* Ind. Eng. Chem. Res. **54**, 4633 (2015).

[103] Y. Rosenfeld, *Relation between the transport coefficients and the internal entropy of simple systems,* Phys. Rev. A **15**, 2545 (1977).

[104] Y. Rosenfeld, *A quasi-universal scaling law for atomic transport in simple fluids,* J. Phys. Condens. Matter **11**, 5415 (1999).

[105] O. Lötgering-Lin and J. Gross, *Group contribution method for viscosities based on entropy scaling using the perturbed-chain polar statistical associating fluid theory,* Ind. Eng. Chem. Res. **54**, 7942 (2015).

[106] M. Hopp, O. Loetgering-Lin, and J. Gross, *Calculation of thermal conductivity using PCP-SAFT based on entropy scaling,* in *28th European Symposium on Applied Thermodynamics* (2015).

[107] M. Hopp, O. Lötgering-Lin, and J. Gross, *Transportgrössen aus der Entropieskalierung mit PC-SAFT: Viskosität und Wärmeleitfähigkeit,* in *Thermodynamik Kolloquium 2015* (2015).

[108] G. M. Kontogeorgis and I. G. Economou, *Equations of state: From the ideas of van der Waals to association theories,* J. Supercrit. Fluids **55**, 421 (2010).

[109] R. Privat, F. Gaillochet, and J.-N. Jaubert, *Testing the ability of various equations of state to reproduce high-pressure isotherm crossings in the (α, P) plane,* Fluid Phase Equilib. **327**, 45 (2012).

[110] M. S. Wertheim, *Fluids with highly directional attractive forces. I. Statistical thermodynamics,* J. Stat. Phys. , 19 (1984).

[111] M. S. Wertheim, *Fluids with highly directional attractive forces: II. Thermodynamic perturbation theory and integral equations,* J. Stat. Phys. , 35 (1984).

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

[113] J.-P. Hansen and I. R. McDonald, *Chapter 5 - Perturbation Theory,* in *Theory of Simple Liquids (Fourth Edition)*, edited by J.-P. Hansen and I. R. McDonald (Academic Press, Oxford, 2013) pp. 149 – 202.

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

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

[116] S. H. Huang and M. Radosz, *Equation of state for small, large, polydisperse, and associating molecules,* Ind. Eng. Chem. Res. **29**, 2284 (1990).

[117] J. Vrabec and J. Gross, *Vapor–Liquid Equilibria simulation and an equation of state contribution for dipole–quadrupole interactions,* J. Phys. Chem. B **112**, 51 (2008).

[118] S. I. Sandler, *Chemical, Biochemical and Engineering Thermodynamics* (John Wiley & Sons, 2006).

[119] A. Bamberger and G. Maurer, *High-pressure vapor-liquid equilibria in binary mixtures of carbon dioxide and aromatic hydrocarbons: Experimental data and correlation for* co_2 *+ acetophenone,* co_2 *+ 1-chloronaphthalene,* co_2 *+ methyl benzoate and* co_2 *+ n-propylbenzene,* J. Supercrit. Fluids **7**, 115 (1994).

[120] T. E. Berty, H. H. Reamer, and B. H. Sage, *Phase behavior in the hydrogen-cyclohexane system.* J. Chem. Eng. Data **11**, 25 (1966).

[121] R. E. Thompson and W. C. Edmister, *Vapor-liquid equilibria in hydrogen-benzene and hydrogen-cyclohexane mixtures,* AIChE J. **11**, 457 (1965).

2

Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD) for Simultaneous Process and Solvent Design for CO_2 Capture

This chapter has been published as: M. Stavrou, M. Lampe, A. Bardow, J. Gross; Ind. Eng. Chem. Res. 53 (2014) 18029 – 18041

Solvent-based separation systems have a substantial potential for improvement when the solvent and the process conditions are optimized simultaneously. The fully integrated design problem, however, leads to an optimization problem of prohibitive size and complexity due to the many discrete degrees of freedom in selecting a solvent and the nonlinear nature of the process models. We here implement and extend the method of Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD) for the solvent and process optimization of a pre-combustion CO_2-capture system with physical absorption. CoMT-CAMD is a deterministic procedure that does not require a pre-selection of solvent molecules. The process topology considered in our study includes all major process operations of an existing CO_2-capture system: multi-stage absorption, desorption (two flash desorption stages with gas recycle) and CO_2 compression. We measure the process performance with a single economic objective function. The objective function captures the process trade-offs and evaluates potential process-solvent on a common basis. The solvent is represented as the pure component parameters of the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT). The optimization problem is formulated with the pure component parameters of the solvent (PC-SAFT parameters) and with the process variables as degrees of freedom. Necessary auxiliary properties of the optimized solvent like the ideal gas heat capacity and the molar mass are predicted with Quantitative Structure Property Relationship (QSPR) models, based on the pure component PC-SAFT parameters. As a result, one gets a unified thermodynamic framework for fluid properties based on the PC-SAFT model. With CoMT-CAMD we obtain a list of the best performing physical solvents for the considered CO_2-capture application. The resulting list of best performing solvents contains state-of-the-art solvents and new green solvent molecules.

2.1. Introduction

Identifying an optimal solvent for a specific process is a demanding task since the solvent selection cannot be decoupled from process design. The solvents for carbon capture by absorption, in particular, are known to critically determine the costs of possible processes. [1, 2] In pre-combustion carbon capture, for example, very different types of solvents are used in the main commercially installed processes (Selexol, Rectisol, Purisol and Fluor Solvent) [3]. On the one hand, the selected solvent determines the optimal operating point of the separation system [4–8]. On the other hand, specifications of process and operating conditions are decisive for the selection of new candidate solvents [9–12]. Wilcox et al. [13] recently revisited film theory for the mass transfer of carbon dioxide and illustrated the strong coupling between process parameters and solvent properties. In summary, the selection of an optimal solvent cannot be decoupled from the optimal selection of process conditions. The optimal performance of a solvent-based separation system is determined by both the solvent properties and the operating conditions. An integrated process and product design approach is needed. A full integration of the solvent and process optimization problems results however, in a mixed-integer non-linear programming (MINLP) problem

$$\begin{aligned} &\min_{\overline{x},\overline{p}} f(\overline{x},\overline{p}) \\ &\text{s.t.} \\ &h(\overline{x},\overline{p}) = 0 \\ &g(\overline{x},\overline{p}) \leq 0 \\ &\overline{x} \in \Re^m, \overline{p} = [p_1, \dots, p_D] \in \mathcal{P}DxN \end{aligned} \tag{2.1}$$

where $f(\overline{x},\overline{p})$ denotes the objective function of the optimization problem, $h(\overline{x},\overline{p})$ is the set of nonlinear equations representing the process model, $g(\overline{x},\overline{p})$ are the nonlinear process constraints and $\overline{x}$ is the vector of the process variables. The process model $h(\overline{x},\overline{p})$ and the process constraints $g(\overline{x},\overline{p})$ include equations of the physical property model, in which a solvent is characterized by a parameter vector $\overline{p}$. Each candidate solvent species i is represented by the parameter vector $\overline{p}_i$ in the discrete space $\mathcal{P}^{DxN}$ where D is the number of model parameters to represent a solvent and N is the number of all real species considered in the problem. The choice of a solvent species i is therefore an integer decision in the N-dimensional search space. The mathematical formulation, Eq.(2.1), shows that solving the full MINLP problem for reasonably complex practical tasks is of prohibitive dimensionality N, determined by the number of conceivable molecules. Any attempt to solve the integrated optimization problem of Eq.(2.1) has to consider a mathematical or conceptual relaxation step to allow for practical (approximate) solutions. The various methods to circumvent or simplify the full MINLP problem in Eq.(2.1) can be used to classify the vivid research in the field of process and product design. A comprehensive overview of recent developments has been given by Adjman et al. [14].

One can thereby distinguish between methods that define auxiliary objective

functions $\check{f}(\overline{p})$ to (pre)select solvents and methods that preserve the objective function $f(\overline{x},\overline{p})$ on the process level. Note that an auxiliary objective function is usually defined such that $\check{f}(\overline{p})$ does not depend on process variables $\overline{x}$, allowing the minimization of $\check{f}(\overline{p})$ as an independent subproblem. Another class of approaches introduces simplified process models $h(\overline{x},\overline{p})$ in order to make the problem solvable. One can further differentiate methods that limit the search space $\check{N}$ to certain chemical structures in advance and methods that require no pre-selection. Another level of approximation decomposes the MINLP into an optimization problem with continuous parameter vector $\overline{p}$ and a subsequent integer problem of identifying the optimal solvent.

In this regard, the structure of the MINLP problem in Eq.(2.1) was maintained in ref. [15, 16], but in order to ensure a feasible mathematical solution, either simplified process models are employed or the size of the considered integer (molecular) search space is limited. Sahinidis et al. [17] proposed an advanced global optimization strategy for the solution of the MINLP problem in Eq.(2.1). As objective function, their approach considered particular fluid properties as performance measures of the optimized fluid.

Many approaches decompose the full MINLP problem into two sub-problems, which are solved independently: First, a short-listing of the solvents is conducted based on heuristic, predefined performance measures or solvent properties $\check{f}(\overline{p})$; subsequently, the process conditions are optimized for the pre-selected solvent(s). [18–20] Pre-selection based on heuristics requires, however, practical experience and process know-how. This decomposition approach has been shown to be powerful, when expert knowledge of the process is available [21, 22]. This knowledge, however, is not always available when new designs and processes are examined.

Marcoulaki and Kokossis [23] emphasized the significance of approaches that are unbiased from pre-selection procedures and thus maintain the potential for novel solutions. They proposed a framework that allows a maximum number of chemical structures to be considered for the molecular representation and they applied a stochastic algorithm for the solution of the optimization problem. They defined the objective function either on certain physical properties of the optimized molecule or on individual process streams. [23, 24]

Recent approaches in Computer Aided Molecular Design (CAMD) reduce the initial molecular search space using elegant mathematical or algorithmic techniques: Bommareddy et al. use enhanced enumeration techniques of higher order groups in a group-contribution approach. The process design problem is first solved to identify ranges of property targets, which are further used as process property constraints in a molecular design problem. [25] Samudra and Sahinidis [26] consider a simplified property model and allow for the solution of a mixed-integer linear problem (MILP) to define the initial molecular composition. In a following step, they use molecular graph techniques to define the exact molecular structures. Papadopoulos and Linke [27] implement multi-objective optimization and an iterative molecular clustering approach to confine the initial molecular search space. For the demanding case of reactive systems, Chemmangattuvalappil and Eden [28] implement the concept of molecular signature descriptors and the principles of molecular

graph theory in property based molecular design. All the above methods are very elaborate but they still use auxiliary objective functions for the evaluation of the process performance, in order to reduce the problem complexity on the process level. Lampe et al. [29], however, showed that the dependence of process performance on molecular properties is not smooth due to process constraints. Such non-smooth objective functions of the full problem cannot be captured by auxiliary objective functions $\check{f}(\overline{p})$ that are solved independent of the process model.
Pereira et al. [30, 31] proposed a Computer Aided Molecular and Process Design (CAMPD) methodology for solvent selection among alkane-blends. This method uses a physically based thermodynamic model (SAFT-VR [32]). Thereby, variables describing the molecular structure can be directly embedded into the process model calculations. The overall performance is then evaluated with a single objective function measuring process performance.

The method of Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD), introduced by Bardow et al. [33], solves the molecular design problem simultaneously to the process optimization without pre-selection of candidate substances. CoMT-CAMD uses a physically based thermodynamic model to enable the direct coupling of the molecular structure to the process performance. In problem Eq.(2.1), the CoMT-CAMD approach relaxes the parameter vector $\overline{p}$ from discrete values to continuous space. The resulting subproblem is a non-linear optimization problem, where solvent parameters $\overline{p}$ and process variables $\overline{x}$ are optimized simultaneously. In a subsequent step, a list of best real solvents is obtained from the optimized solvent parameters $\overline{p}^{opt}$. Lampe et al. [34] successfully implemented the CoMT-CAMD method for the simultaneous process and working fluid optimization for Organic Rankine Cycles.
Bardow et al. [33] presented a first conceptual implementation of the CoMT-CAMD method using a simplified topology for the CO_2 capture. In the present work, we describe the further development of CoMT-CAMD and a comprehensive application of the method on the solvent and process optimization in separation systems using physical absorption. The further development concerns on the one hand a method for predicting the ideal gas heat capacity and the solvents' molar mass based solely on the solvent molecular parameters $\overline{p}$. This development enables the modeling of all thermal and full caloric properties of a target fluid within the CoMT-CAMD approach. On the other hand, we introduce a method to approximate non-equality process constraints during the mapping step. Further, a detailed process topology for the pre-combustion capture of carbon dioxide is now examined. We implement an economic objective function accounting for the various trade-offs in the absorption, desorption and CO_2 compression stage. We intend to show how CoMT-CAMD can be successfully performed for process models that include complex phase equilibrium calculations of multi-component streams. We demonstrate that CoMT-CAMD delivers accurate results and identifies the most promising solvents in a fully deterministic procedure.

2.2. Continuous Molecular Targeting (CoMT)-CAMD for solvent selection

The CoMT-CAMD method for the integrated process and solvent design is founded on physically based thermodynamic models. In a physically based thermodynamic model, the pure component parameters are meaningful for identifying the structure of the molecules. In CoMT-CAMD, these parameters are used as optimization variables. Thereby, the solvent optimization is integrated into the process optimization level. In this work, we use CoMT-CAMD with the PC-SAFT (Perturbed Chain Statistical Associating Fluid Theory) thermodynamic model [35–38]. The pure component parameters of PC-SAFT describe the geometry, the van der Waals attractive energy, the strength of hydrogen bonds (association) and the multipole moments of a molecule. Each real substance is uniquely represented by a set of molecular parameters. Thus N real substances define a discrete parameter vector space $\mathcal{P}^{(D\times N)}$. As a result, any attempt to iterate the discrete parameter vector $\overline{p}$ of real solvents together with the process equations leads to a mixed-integer optimization problem (cf. Eq.(2.1)). In order to circumvent the mixed-integer formulation, we relax the discrete vector space of the molecular parameters to a continuous domain. This relaxation strategy allows for the use of the parameters $\overline{p}$ of the thermodynamic model as continuous optimization variables, along with the process degrees of freedom $\overline{x}$. In CoMT-CAMD, the integrated process and product design problem is therefore formulated as a constrained, non-linear optimization problem with an objective function f to evaluate the overall process performance.

The CoMT-CAMD method is generic and its accuracy is determined by the choice of the thermodynamic model. With this study, we intend to show that the PC-SAFT equation of state allows for a comprehensive description of thermodynamic properties of real fluids with good accuracy. Traditionally, PC-SAFT describes only residual quantities; i.e. quantities in departure of the ideal gas contribution. In this work we go beyond this scope: by applying Quantitative Structure Property Relationship (QSPR) methods, we correlate various auxiliary physical properties using only the pure component parameters of the PC-SAFT model. We achieve a thermodynamically consistent description of fluids based on a single set of parameters.

CoMT-CAMD proceeds in two steps: In the first step ('CoMT step') we solve an overall performance optimization problem to define the (hypothetical) optimal molecule with parameter vector $\overline{p}^{opt}$ at optimal process conditions $\overline{x}^{opt}$. The best performing real molecules have parameter vectors $\overline{p}$ similar to $\overline{p}^{opt}$ due to the physical basis of PC-SAFT. In the second step of the method ('mapping step') we identify these real molecules with the best overall performance.

2.2.1. Continuous Molecular Targeting: Simultaneous process and molecular optimization

The simultaneous optimization of process variables and molecular parameters is performed in the first step of the CoMT-CAMD method ('CoMT step', Figure 2.1). The integrated design problem, written as a non-linear optimization problem, is solved with a gradient based non-linear programming algorithm. The mathematical

problem is here formulated in a slightly more detailed form, compared to problem (2.1), as

$$
\begin{aligned}
&\min_{\overline{x},\overline{p}} f(\overline{x},\overline{y},\overline{p}) \\
&\text{s.t} \\
&h(\overline{x},\overline{y},\overline{p}) = 0 \qquad \text{(process model)} \\
&g(\overline{x},\overline{y},\overline{p}) \leq 0 \qquad \text{(process inequality constraints)} \\
&\overline{\overline{A}} \cdot \overline{p} \leq \overline{b} \qquad \text{(convex hull on solvent parameters)} \\
&\overline{x}_{lb} \leq \overline{x} \leq \overline{x}_{ub} \\
&\overline{y}_{lb} \leq \overline{y} \leq \overline{y}_{ub} \\
&\overline{p}_{lb} \leq \overline{p} \leq \overline{p}_{ub}
\end{aligned} \tag{2.2}
$$

As opposed to Eq.(2.1), where $\overline{x}$ denoted all process variables, we here make a distinction between the process variables $\overline{x}$ that are defined as degrees of freedom of the optimization problem and the remaining process variables $\overline{y}$. We minimize a single process objective function f which depends on $\overline{x},\overline{y}$, and the molecular parameters of the solvent $\overline{p}$. Degrees of freedom for the integrated problem are the process variables $\overline{x}$ and the molecular parameters of the solvent $\overline{p}$ in the continuous domain $\mathcal{R}^n$ (with $n = m+D$ according to Eq.(2.1)). The solution of the optimization problem is a vector with the optimal molecular parameters of the solvent and the corresponding optimal process conditions. In order to avoid extreme hypothetical molecules, we constrain the search space of the molecular parameters in the CoMT step to a region which is plausible. For this purpose, we confine the parameter space to the convex hull formed by the molecular parameters of real molecules in our database (Figure 2.2). In problem Eq.(2.2) the convex hull is written as a set of linear inequality constraints [39, 40] describing the planes of the convex hull fronts. Process constraints are considered in the problem formulation of Eq.(2.2) as non-linear inequality constraints, $g(\overline{x},\overline{p},\overline{z})$.

2.2.2. Mapping step: Identification of the best performing components

The second step in the CoMT-CAMD method ('mapping step', Figure 2.1) identifies the real substances with similar behavior as the optimal molecule. Similarity in behavior is thereby measured by the objective function.

For every real substance i included in a database of molecular parameters ('mapping database'), we approximate the value of the objective function f_i by a second order Taylor series expansion around the optimal (hypothetical) molecular parameters $\overline{p}^{opt}$, as

$$
f_i \approx f^{opt} + J(\overline{p}^{opt}) \cdot (\overline{p}_i - \overline{p}^{opt})^T + \frac{1}{2} \cdot (\overline{p}_i - \overline{p}^{opt})^T \cdot H(\overline{p}^{opt}) \cdot (\overline{p}_i - \overline{p}^{opt}) \tag{2.3}
$$

Every real substance in the database is represented by a known vector of molecular parameters $\overline{p}_i$. Here, f^{opt} is the optimal value of the objective function, the

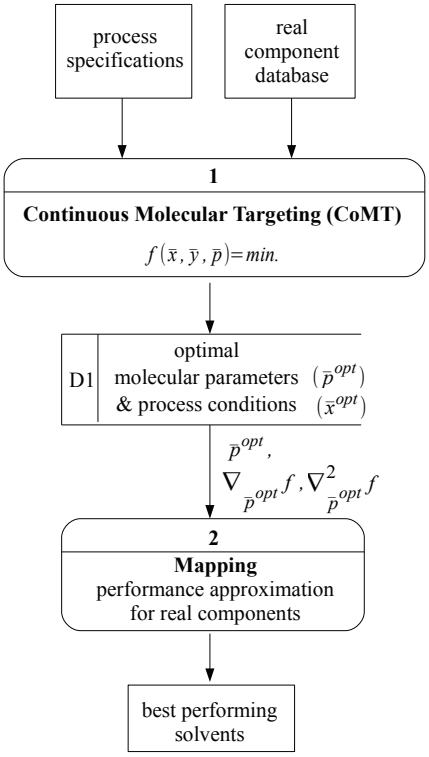


Figure 2.1: Data flow diagram of the CoMT-CAMD method.

Jacobian $J(\overline{p}^{opt})$ and the Hessian $H(\overline{p}^{opt})$ are the first and second derivatives of the objective function f^{opt} with respect to the molecular parameters at the optimum, respectively. Both, $J(\overline{p}^{opt})$ and $H(\overline{p}^{opt})$ are calculated in a subsequent step as described in section 2.3.2.

However, the derivatives of the objective function do not contain information about how changes of the molecular parameters affect the process constraints $g(\overline{x},\overline{y},\overline{p}) \leq 0$ in Eq.(2.2). Even constraints that are not active at the optimum can be violated for small changes of the optimal parameters. In order to account for possible violations of process constraints, we approximate values of the inequality constraints g_i from a Taylor series of 1st order around the value of g^{opt} at optimum, as

$$g_i \approx g^{opt} + J_g(\overline{p}^{opt}) \cdot (\overline{p}_i - \overline{p}^{opt})^T \leq 0 \tag{2.4}$$

where $J_g(\overline{p}^{opt})$ is the derivative of g with respect to $\overline{p}$ at the optimum. The components that do not violate the constraints, according to their approximation with Eq.(2.4), are further evaluated for their performance with Eq.(2.3). Thereby, the components are ranked according to the approximated value of the objective function in the so called "mapping list". The best performing components in the mapping

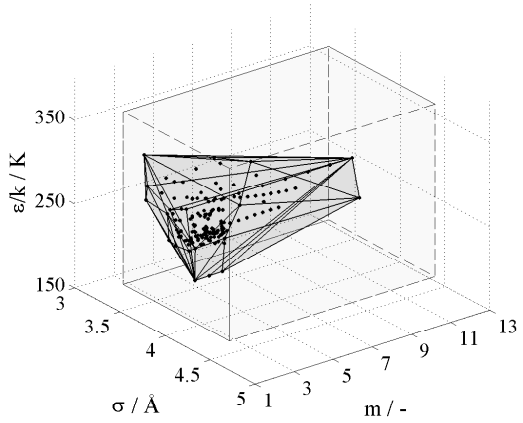


Figure 2.2: Convex hull defined by real molecules with respect to three molecular parameters ($m, \sigma, \varepsilon/k$). Real molecules (spherical symbols) define a physically attainable search space. The optimization is allowed within the convex hull (grey-shaded volume). The fronts of the convex hull (planes defined by the continuous lines) are represented in the mathematical problem of Eq.(2.2) through linear inequality constraints.

list (i.e. components with lowest values of f_i in Eq.(2.3)) should be rigorously optimized. Finally, they should be examined for their toxicity, corrosivity, flammability or any health risks if required.

2.2.3. Prediction of physical properties using PC-SAFT

The PC-SAFT thermodynamic model

The SAFT (Statistical Associating Fluid Theory) equation-of-state (EoS) proposed by Chapman et al. [41] led to a family of theoretically based engineering equations-of-state, originating in statistical mechanics and based on Wertheim's perturbation theory [42, 43]. The Perturbed Chain Statistical Associating Fluid Theory (PC-SAFT) equation-of-state (EoS) introduced by Gross and Sadowski [35] is one of the variants of the SAFT-family of models (for a review see ref. [44]). PC-SAFT is based on a coarse-grained molecular model representing each molecule as a chain of spherical segments interacting with other chain molecules.[45] PC-SAFT is formulated as a residual Helmholtz energy, as function of temperature (T) and the vector $\overline{\rho}$ of component molar densities $\rho_i = \rho \cdot x_i$ of all species i of the system, according to

$$\alpha^{\mathrm{res}}(T,\overline{\rho}) = \alpha^{\mathrm{hc}}(T,\overline{\rho}) + \alpha^{\mathrm{disp}}(T,\overline{\rho}) + \alpha^{\mathrm{assoc}}(T,\overline{\rho}) + \alpha^{\mathrm{polar}}(T,\overline{\rho}) \qquad (2.5)$$

where ρ is the molar density and x_i here denotes the molar fraction of species i. The first term α^{hc} of the residual Helmholtz energy in Eq.(2.5) describes the contribution of the hard-chain reference fluid. The additional terms account for various attractive parts of the intermolecular potential, namely contributions due to dispersive (van der Waals), associating (hydrogen bonding) and polar attractions.

Any other residual (static) thermodynamic property can be derived directly from the residual Helmholtz energy. For a full description of the PC-SAFT EoS we refer to the original literature [35–38]. We here refer to the polar version of the model, which was earlier also termed PCP-SAFT [37, 38], simply as PC-SAFT.

In PC-SAFT non-associating molecules are characterized by three pure-component parameters: the number of segments per chain molecule (m), the segment diameter parameter (σ), and the dispersion energy parameter (ε/k). For the description of polar molecules, the dipole moment (μ) and the quadrupole moment (Q) are also required. Associating interactions are determined by two additional pure-component parameters: the association energy parameter (ε^{AB}/k) and the effective association volume (κ^{AB}). In this work, we do not include association parameters of the PC-SAFT model as degrees of freedom.

Correlation of auxiliary physical properties based on PC-SAFT parameters

For process calculations with PC-SAFT we need to determine full caloric properties, such as enthalpies. For this purpose, we need a fundamental equation for the Helmholtz energy for the fluid phase region, with

$$\alpha(T,\overline{\rho}) = \alpha^{\mathrm{ig}}(T,\overline{\rho}) + \alpha^{\mathrm{res}}(T,\overline{\rho}) \tag{2.6}$$

Thus, an expression for the residual Helmholtz energy α^{res} from PC-SAFT is not sufficient; an ideal gas contribution is required. The ideal gas term is fully defined through the specific isobaric heat capacity c_p^{ig} of all involved species in pure state. While many alternative ways of defining the ideal gas contribution are possible, the route of specifying the ideal gas isobaric heat capacities c_p^{ig} of pure species is appropriate, because c_p^{ig} is available as primary measurement data for many components.

Our process calculations additionally require knowledge of the solvent's molecular mass M_i, since equipment sizing, cost functions and emission limits are often based on mass-averaged stream properties.

For known substances, the physical properties c_p^{ig} and M can easily be drawn from property databases. During molecular design in the CoMT step, this is not possible, since the iterated solvent molecule is hypothetical and only characterized by its PC-SAFT parameters. We therefore propose relations for estimating c_p^{ig} and M based only on the pure component parameters of the PC-SAFT model.

Correlations between the molar mass and the dispersive pure-component parameters of PC-SAFT ($m, \sigma, \varepsilon/k$) have been investigated by other authors in the past. Gross and Sadowski [35] examined the underlying correlations for the homologous series of n-alkanes. Tihic et al. [46] demonstrated the near-linear dependencies between molar mass and combinations of the dispersive parameters for various component families, Grenner et al. [47] individually for glycols and Nannan et al. for polyethyleneglycol dimethylethers [48]. Mac Dowell et al. have also proposed relations between the pure component molecular parameters and the molecular mass of n-alkyl-1-amines for the SAFT-VR model. [49] Based on these findings, we use the four PC-SAFT parameters $m, \sigma, \varepsilon/k$ and μ as descriptors in a quantitative

structure property relationship (QSPR) model to predict the molar mass.

$$(M)^{\text{pred}} = \beta_1 \cdot (m\varepsilon/k) + \beta_2 \cdot \left(m\sigma^3\right) + \beta_3 \cdot \left(m\sigma^3\varepsilon/k\right) + \beta_4 \cdot \left(m\sigma^3\mu\right) \tag{2.7}$$

For the prediction of c_p^{ig}, the QSPR descriptors are combinations of the three PC-SAFT parameters: m, σ and ε/k.

$$(c_p^{\text{ig}})^{\text{pred}} = \gamma_0 + \gamma_1 \cdot (m\varepsilon/k) + \gamma_2 \cdot \left(m\sigma^3\right) + \gamma_3 \cdot \left(m\sigma^3\varepsilon/k\right) \tag{2.8}$$

The heat capacity of pure substance in the ideal gas state accounts for the energy of the molecule due translation, rotation and all internal vibrations or excitations. Values of c_p^{ig} therefore strongly depend on the molecular structure. Although the PC-SAFT pure component parameters characterize intermolecular interactions, we argue that they should still be meaningful QSPR-descriptors for intramolecular energy states: The parameters m and σ should be suitable descriptors for rotational energies because they reflect a molecules' geometry in a coarse-grained manner, while ε/k measures the effective electron correlation among two sites and the parameter should thus serve as a sensible descriptor for vibrational energies.

2.3. Application of CoMT-CAMD to solvent selection for CO_2 capture with physical absorption

2.3.1. Process description and process specifications

The CoMT-CAMD method is here implemented for the integrated solvent and process optimization of a carbon capture system with physical absorption. We examine the pre-combustion carbon dioxide (CO_2) capture from the syngas stream of a coal-fired Integrated Gasification Combined Cycle (IGCC) power plant [50–52]. The process topology resembles in its key features the 'CO_2 Catch-up' pilot plant in Buggenum, the Willem Alexander Centrale (WAC), operated by NUON-Vattenfall. [53] The syngas feed includes three components: hydrogen (H_2), carbon dioxide (CO_2) and water (H_2O). We particularly account for the presence of water in the system, since water has a significant impact on the phase equilibrium calculations. The mole fractions of hydrogen $x_{H_2}^{\text{sg}}$ and water $x_{H_2O}^{\text{sg}}$, as well as the temperature T^{sg} and the pressure P^{sg} of the syngas feed (Table 2.2) considered here are the same as for the syngas stream that enters the carbon capture pilot plant in Buggenum. The full process topology includes the major unit operations of the absorption, the desorption and the carbon dioxide compression stage (Figure 2.3). The solvent make-up stream consists of pure solvent at ambient pressure and temperature. All four substances regarded in the flowsheet (CO_2, H_2, H_2O and the solvent) are present in the gas stream exiting the absorber ('hydrogen outlet') and the gas stream after compression ('pipeline gas'). The absorber is defined with a fixed number of 7 equilibrium stages N_{abs}, with a constant pressure of $2\,\text{MPa}$ on the same pressure-level as the syngas stream, and with a constant CO_2 capture rate of 90% in the absorber. The value of 7 equilibrium stages is the result of a sensitivity analysis, where we observed little response of both, the optimal solvent

parameters $\overline{p}^{opt}$, optimal process settings $\overline{x}^{opt}$, together with only a moderate decline of the objective function for $N_{abs} > 7$ (see Appendix B.1.1). The desorption includes a first flash stage which feeds a hydrogen-rich gas back to the absorber, followed by an additional desorption flash. Both desorption stages are adiabatic (pressure-reduction) flash units.

For the rich gas and the pipeline compressors we consider a polytropic, adiabatic compression model as described in Smith [54]. The stream, process and equipment specifications are summarized in Table 2.2.

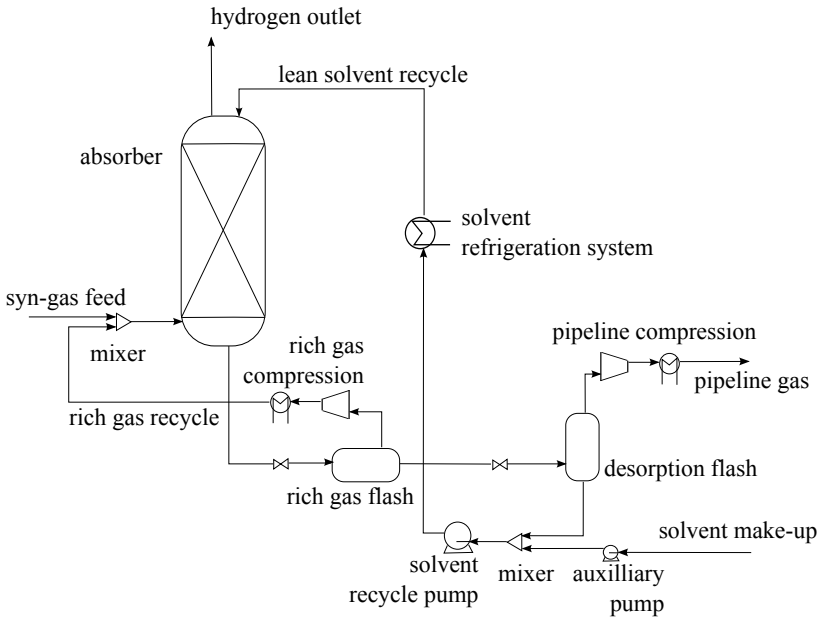


Figure 2.3: Flowsheet for CO_2 capture with physical absorption, including absorption, desorption in two flash vessels and gas compression steps.

A more detailed description of the process model is given in Appendix B.1 of this thesis.

2.3.2. Description of the optimization problem

For our optimization problem (Eq.(2.2)) we define an economic objective function as the sum of several contributions of the operating costs,

$$f_{\text{tot}} = \sum_i f_i = f_{\text{electr.demand}} + f_{\text{H2 loss}} + f_{\text{solvent loss}} + f_{\text{utilities}} \tag{2.9}$$

The term $f_{\text{electr.demand}}$ summarizes the costs of electricity for the solvent recycle pump and for the compressors (CO_2-compression to pipeline-pressure and rich gas compression). We consider an electricity production price equal to $55\,€/\text{MWh}$ [56].

Table 2.1: Stream, process and equipment specifications adopted from the Buggenum pilot plant [53]. The pipeline pressure is taken from ref. [55]; efficiencies from ref. [50].

Stream specifications		
Syngas feed composition	$x^{sg}_{H_2}$ / mol/mol	0.5480
	$x^{sg}_{CO_2}$ / mol/mol	0.4485
	$x^{sg}_{H_2O}$ / mol/mol	0.0035
Syngas feed temperature	T^{sg} / K	313.00
Rich gas recycle temperature	T_{rg} / K	313.00
Pipeline gas temperature	T^{ppl} / K	313.00
Pipeline gas pressure	P^{ppl} / MPa	11.00
Solvent make-up temperature	T^{MU} / K	298.15
Solvent make-up pressure	P^{MU} / MPa	0.1013
Process specifications		
CO_2 capture rate in absorber	α_{CO_2} / –	0.90
Absorption pressure	P_{abs} / MPa	2.00
Equipment specifications		
Absorber - number of stages	N_{abs} / –	7
Rich gas compressor - number of stages	$N_{comp-rg}$ / –	1
Pipeline compressor - number of stages	$N_{comp-des}$ / –	1
Compressors - efficiency	η_{comp} / –	0.82
Solvent recirculating pump - efficiency	η_{pump} / –	1.00
Auxiliary pumps - efficiency	$\eta_{aux.pump}$ / –	0.75

$f_{\text{utilities}}$ takes the operating cost of the cooling systems into account, given that this is measured through the electric energy demand for their operation [54]: The two heat exchangers (after the rich-gas compression and after the pipeline compression, (Figure 2.3)) are operated with cooling water. Thus, the cost contribution is due to the electrical duty of the cooling-water pumps. For cooling of the lean solvent recycle stream we consider a vapor compression refrigeration system with an exergetic efficiency η_{ex} of 25% [54, 57]. The term $f_{\text{H2 loss}}$ is the lost profit due to hydrogen loss and evaluates the lost opportunity to produce electricity from the hydrogen that is captured in the CO_2-rich stream. Finally, the term $f_{\text{solvent loss}}$ expresses the costs for the fresh solvent. This term accounts for the solvent loss both in the absorption and desorption stage. We consider an average price of high molecular mass chemicals of $4{,}000$€/T. All cost-contributions are expressed per ton of captured CO_2 in the absorber. A more detailed description of the objective function is given in Appendix B.2.

The process model is given as a set of non-linear equality constraints $h(\overline{x}, \overline{y}, \overline{p}) = 0$. The absorber and the flash units are described by the so called MESH equations [58]. The process optimization variables $\overline{x} = \{P^{\text{rg}}, P^{\text{des}}, T^{\text{lean}}\}$ include the desorption pressure of both desorption stages $P^{\text{rg}}, P^{\text{des}}$ and the temperature of the lean solvent T^{lean}. The solvent optimization variables are the PC-SAFT molecular parameters of the solvent $\overline{p} = \{m, \sigma, \varepsilon/k, \mu\}$. Thus, we limit the presentation to non-associating solvents. As non-linear inequality process constraint $g(\overline{x}, \overline{y}, \overline{p})$, we limit the solvent emission m_s in the hydrogen outlet of the absorber to a maximum of $20\,\text{mg/Nm}^3$ [59].The convex hull of the molecular parameter search space is given by the set of linear inequality constraints (Eq.(2.2)). [40]

For the solution of the optimization problem, we use a commercial NLP solver [60] with an interior Conjugate Gradient (CG) algorithm [61]. The solver delivers only the approximate Hessian of the Lagrangian at the optimum. For the mapping step, however, we need the Hessian and the Jacobian of the objective function and of the constraints at optimum (Eq.(2.3) and (2.4)). We currently calculate these derivatives with finite differences around the optimum with respect to the optimal solvent parameters $\overline{p}^{opt}$. The computation of each Hessian element $\left(\frac{\partial^2 f}{\partial p_i \partial p_j}\right)_{p_{k\neq i,j}}$ with forward finite differences demands four discretization points and one function evaluation for each discretization point. We evaluate the objective function f^*at each discretization point with $p_i^+ = p_i^{opt} + \Delta p_i$, $p_j^+ = p_j^{opt} + \Delta p_j$ and $p_{k\neq i,j}^+ = p_{k\neq i,j}^{opt}$ solving the following optimization problem:

$$
\begin{aligned}
& f^+(\overline{x}, \overline{y}, \overline{p} = \overline{p}^+) = \min_{\overline{x}} f^+(\overline{x}, \overline{y}, \overline{p} = \overline{p}^+) \\
& \text{s.t.} \\
& h(\overline{x}, \overline{y}, \overline{p} = \overline{p}^+) = 0 \qquad \text{(process model)} \\
& g(\overline{x}, \overline{y}, \overline{p} = \overline{p}^+) = 0 \qquad \text{(process ineq. constraint)} \\
& \overline{x}_{lb} \leq \overline{x} \leq \overline{x}_{ub}, \quad \overline{y}_{lb} \leq \overline{y} \leq \overline{y}_{ub}
\end{aligned} \tag{2.10}
$$

The resulting Jacobian and Hessian (to be used in Eq.(2.3)) are thus defined for a surface of optimal process settings $\overline{x} = \{P^{\text{rg}}, P^{\text{des}}, T^{\text{lean}}\}$. We verified that for

the derivative calculations the use of central finite differences is not decisive for the accuracy of the Taylor approximation in the mapping step. The size of the differentiation step however should be selected carefully. The implementation of automatic analytic differentiation techniques can therefore be beneficial. Lampe et al. [29] successfully used an automatic differentiation tool in the CoMT-CAMD application for the Organic Rankine Cycle (ORC) process model.

2.3.3. Predicting phase equilibria with PC-SAFT

The PC-SAFT pure component parameters of carbon dioxide, hydrogen and water are taken from the literature. Carbon dioxide is considered as a quadrupolar molecule. [37] The vapor-liquid equilibrium of pure hydrogen covers low temperatures, where quantum corrections are needed. We therefore considered hydrogen as a simple Lennard-Jones fluid and adopted parameters from Hirschfelder et al. [62], based on data of second Virial coefficients. The pure component parameters for water are taken from ref. [36] [1]. The heat capacity of ideal gas states, $c_p^{\mathrm{ig}}(T)$, for CO_2, H_2 and H_2O is taken from the literature ref. [65].

For the vapor-liquid equilibrium calculations of the binary mixtures of CO_2/H_2O, CO_2/H_2 and H_2/H_2O we use non-zero binary interaction parameters k_{ij}. According to our correlation results, the maximum Absolute Average Deviation in both the liquid and the vapor phase occurs for the binary system CO_2/H_2 with $\mathrm{AAD}(x^{\mathrm{L}}_{CO_2/H_2}) = 1.6521\,\%$ and $\mathrm{AAD}(x^{\mathrm{V}}_{CO_2/H_2}) = 1.3677\,\%$ [2]. We note that for the three binary mixtures, the binary interaction parameter k_{ij} is treated as temperature independent. This is a simplification for the binary systems containing H_2O, which however avoids additional non-linearities in the NLP optimization problem. For the binary mixtures of the solvent with CO_2, H_2 and H_2O respectively, we have set k_{ij} values equal to zero. Methods to predict k_{ij} values as a function of the PC-SAFT parameters are an objective of our current and future work. In order to provide some confidence in the phase equilibrium calculations of a relevant process solvent, Figure 2.4 shows the phase equilibrium of the ternary system $CO_2 - H_2 - \mathrm{methanol}$ at $2\,\mathrm{MPa}$ and $288\,\mathrm{K}$[3]. The calculation results of PC-SAFT, with $k_{ij} = 0$ for all pairs of substances involving methanol, can be considered in good agreement to the experimental data.

For the phase equilibrium calculations, we solve the adiabatic-isobaric flash problem. The sub-problem of the isothermal-isobaric two-phase flash is solved with the successive substitution Rachford-Rice algorithm (Michelsen and Mollerup) [66].

[1] A summary of the parameters is published in ref. [63], available online as Supporting Information to the work of Stavrou et al. [64]

[2] The calculated values of the binary interaction parameters k_{ij} and the correlation results are published in ref. [63], available online as Supporting Information to the work of Stavrou et al. [64]

[3] Results for the same ternary mixture at other temperatures are published in ref. [63], available online as Supporting Information to the work of Stavrou et al. [64]

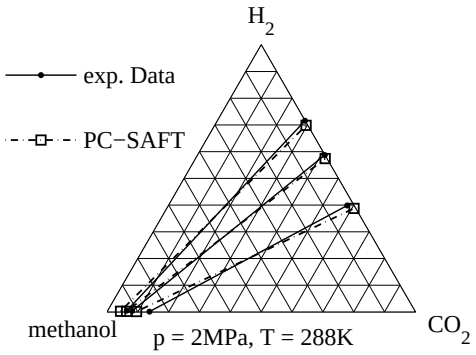


Figure 2.4: Ternary phase diagram for $CO_2 - H_2 - MeOH$ at $P = 2\,\mathrm{MPa}$ and $T = 288\,\mathrm{K}$. Comparison of predictions ($k_{ij} = 0$ for the methanol pairs methanol-CO_2, methanol-H_2) obtained from PC-SAFT (squares) to experimental data (circular points). Experimental data are taken from ref. [67].

2.3.4. Mapping database of the PC-SAFT pure component parameters

Our study is based on a PC-SAFT pure component parameter database covering 657 real, non-associating components. The pure component parameters were adjusted to (quasi)experimental data taken from the DIPPR database [68]. The value of the dipole moment μ_i for the polar components was taken directly from the DIPPR database. The n-polyethylene glycol dimethylethers included in our database (denoted hereafter as 'DEPG-n') are considered to be non-polar components and their molecular parameters have been adopted from the work of Nannan et al. [48].

For the design of the mapping database, we limit the database to components fulfilling the physical bounds of given process conditions: Substances are considered too volatile to be solvents if they have a normal boiling point temperature lower than $300\,\mathrm{K}$. Further, we require a melting point temperature lower than $298.15\,\mathrm{K}$ and we exclude amines, nitriles and nitrates known to react with CO_2.

As a result, the mapping database includes 612 polar and non-polar, non-associating components. We further divide the mapping database in two subordinate databases: The first includes 168 non-polar, non-associating components (denoted hereafter as the 'database of non-polar components'). The second includes 444 polar, non-associating components (denoted hereafter as the 'database of polar components'). It is sensible to introduce these subordinate databases since the components are described by different (sub)sets of PC-SAFT parameters (cf. Section 2.2.3). The definition of the mapping database was made prior to the CoMT step, so that the same database is used both, for the calculation of the convex hull during the optimization, and for the mapping afterward. A list of the 612 components comprising the mapping database is given in Appendix C.

It is straightforward to extend the approach formally to associating mixtures by including the corresponding pure component parameters as variables. Still, asso-

ciating components are currently not included in either of our databases because associating compounds in mixture with CO_2, H_2 and H_2O are expected to require non-zero binary interaction parameters. Vapor-liquid equilibrium calculations for these mixtures with zero binary interaction parameters k_{ij} would therefore not allow the desired level of accuracy of CoMT-CAMD. A sound implementation of associating solvents for the present mixture would thus require predictive methods for the binary interaction parameters k_{ij} which is an active topic of current research (see e.g., Haslam et al. [69])

2.4. Results

2.4.1. Prediction of pure component ideal gas heat capacity and molar mass using PC-SAFT parameters

For the prediction of the ideal gas heat capacity c_p^{ig} and of the molar mass M with the QSPR method as described in section 2.2.3, we use Eq.(2.8) and Eq.(2.7) respectively. For the prediction of c_p^{ig}, we have used constant temperature data for $300\,\mathrm{K}$. For both c_p^{ig} and M, the QSPR coefficients in Eq.(2.8) and Eq.(2.7) have been estimated using ordinary multivariate normal maximum likelihood estimation, assuming normally distributed errors. We have used the multivariate linear regression algorithm *mvregress* provided in the Matlab Statistics Toolbox [70]. The squared correlation coefficient R^2 used to evaluate the quality of the fitting is given by

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(w_i^{exp} - w_i\right)}{\sum_{i=1}^{n}\left(w_i^{exp} - \overline{w}^{exp}\right)} \tag{2.11}$$

We adjusted the QSPR coefficients separately for the two major component classes defined in section 1.3.4. For the ideal gas heat capacity c_p^{ig} we find values of $R^2 = 0.991$ for non-polar and $R^2 = 0.970$ for polar components. The QSPR multivariable regression for the molar mass M results to $R^2 = 0.966$ for non-polar and $R^2 = 0.932$ for polar components.[4] From all training sets, we excluded the inorganic components. Halogenated components are not taken into account for the QSPR model of the molar mass M, since they have very different QSPR-characteristics compared to other organic components. For the training set of c_p^{ig} values, we consider components for which data at $300\,\mathrm{K}$ were available in the DIPPR database. Figure 2.5 illustrates the QSPR results for c_p^{ig} and molar mass M for non-polar components.[5] We assess the accuracy of the QSPR correlation as sufficient. In particular, we do not expect the prediction of c_p^{ig} and of the molar mass M to cause significant error propagation in the process calculations of the CoMT step.

[4] The size of the training sets, the squared correlation coefficients R^2 and the resulting coefficients of the QSPR models are published in ref. [63], available online as Supporting Information to the work of Stavrou et al. [64]

[5] The corresponding results for polar components are published in ref. [63], available online as Supporting Information to the work of Stavrou et al. [64]

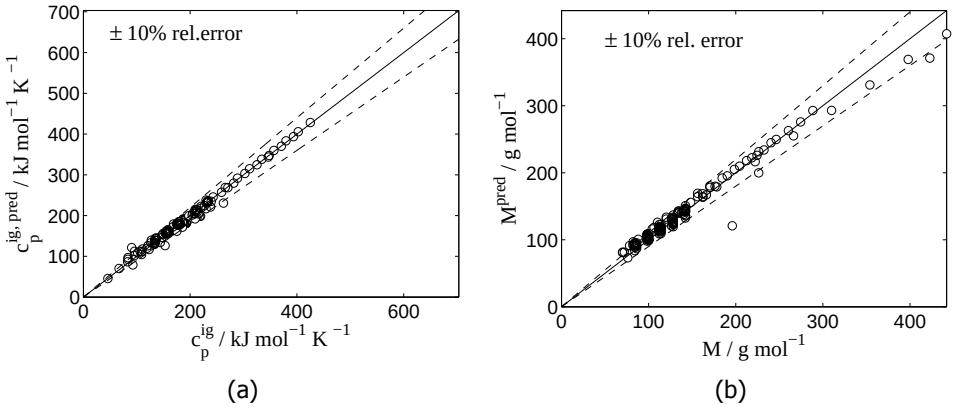


Figure 2.5: Literature values vs. QSPR results for (a) the ideal gas specific heat capacity at $T = 300\,\mathrm{K}$ and (b) the molar mass of non-polar components. The dashed lines indicate $\pm 10\,\%$ relative error.

2.4.2. Simultaneous solvent and process optimization: CoMT-CAMD for polar and non-polar solvents

Here, we present the results of CoMT-CAMD for the simultaneous process and solvent optimization for a CO_2 pre-combustion capture process. Non-polar components are fully described by three PC-SAFT parameters $(m, \sigma, \varepsilon/k)$. Polar components also require the dipole moment μ. Due to this difference in parameter sets, we choose to solve the CoMT-CAMD problem for polar and non-polar components separately.

The pure component parameters of PC-SAFT are known to be correlated. We have made an attempt of defining less correlated coordinates by combining pure component parameters: the molecular 'volume' $(m \cdot \sigma^3)$, the molecular 'energy' $(m \cdot \varepsilon/k)$ and $(m \cdot \sigma^3 \cdot \varepsilon/k)$ as an expression that approximately scales the dimensionless Helmholtz energy [71]. The molecular optimization variables $\overline{p}$ for polar components are the dispersive PC-SAFT parameters $(m, \sigma, \varepsilon/k)$ and the reduced (dimensionless) dipole moment $\mu^* = \mu/_{m \cdot \sigma^3 \cdot \varepsilon/k}{}^{-1/2}$.

In both cases, the process optimization variables $\overline{x} = \{P^{\mathrm{rg}}, P^{\mathrm{des}}, T^{\mathrm{lean}}\}$ are the pressure of the rich gas flash, the pressure of the desorption flash and the temperature of the lean solvent, respectively. The process inequality constraint limits the solvent emission in absorber m_S to a maximum value of $20\,\mathrm{mg/Nm^3}$.

The CoMT optimization for the non-polar solvents is initialized with the molecular parameters of Pentaethylene Glycol Dimethylether (DEPG-5). DEPG-5 is a common component in mixtures of physical solvents for CO_2 absorption and therefore represents a suitable starting point. The result of the CoMT step and of the CoMT-CAMD method for non-polar solvents is summarized in Table 2.2. In Table 2.2, we compare the molecular parameters, the optimal operating conditions, and the value of the objective function for the initial substance DEPG-5, for the optimal (hypothetical) fluid and for the best real molecule (in the non-polar database). The CoMT optimization for polar solvents uses the parameters of N-methyl-2-pyrrolidone as

Table 2.2: Comparison of starting point, the result of the CoMT-step and the result of the CoMT-CAMD method for non-polar solvents. The CoMT-step defines the optimal (hypothetical) non-polar molecule. The 'mapping step' (which completes CoMT-CAMD) identifies the real molecule with best overall performance

	starting molecule	optimal nonpolar molecule	best nonpolar real molecule
	DEPG-5	hypothetical	1,5,9-cyclododecatriene
$M/\mathrm{g/mol}$	266.33	174.22	162.27
$m/-$	7.8895	3.1544	2.7575
$\sigma/$Å	3.5853	4.5177	4.5709
$\varepsilon/k/\mathrm{K}$	253.93	360.37	377.14
		optimal process conditions	
$P^{\mathrm{des}}/\mathrm{MPa}$	0.197	0.239	0.234
$P^{\mathrm{rg}}/\mathrm{MPa}$	0.711	0.689	0.705
$T^{\mathrm{lean}}/\mathrm{K}$	265.3	273.4	274.2
$m_S/\mathrm{mg/Nm^3}$	0.0008	4.208	13.61
$f_{\text{electr.demand}}/€/\mathrm{T_{CO_2\,capt.}}$	7.894	7.228	7.197
$f_{\text{H2 loss}}/€/\mathrm{T_{CO_2\,capt.}}$	0.158	0.188	0.219
$f_{\text{solvent loss}}/€/\mathrm{T_{CO_2\,capt.}}$	0.000016	0.089	0.316
$f_{\text{utilities}}/€/\mathrm{T_{CO_2\,capt.}}$	2.478	1.199	1.563
$f/€/\mathrm{T_{CO_2\,capt.}}$	10.53	8.704	9.295

starting values. Table 2.3 compares results of the starting condition (with optimized process conditions $\bar{x}$) to the results of the the CoMT-step and to results of the CoMT-CAMD method. Table 2.3 shows the pure component parameters, the optimal process conditions, and the value of the objective function for N-methyl-2-pyrrolidone (starting molecule), for the optimal (hypothetical) molecule and for the best real polar molecule (propylene carbonate). We note that the optimization converges to the same optimum from many starting points (real or hypothetical) in both cases. The computations have been performed on a standard desktop PC. For non-polar solvents the problem has 77 model variables in total, including the 3 molecular optimization variables and the 3 process optimization variables. In our implementation the NLP solver converges to the optimal solution in 44 iterations. For polar solvents the problem has 78 model variables in total, including the additional molecular optimization variable of dipole moment. For this problem the NLP solver converges to the optimal solution in 51 iterations.

In both cases, polar and non-polar components, the simultaneous process and solvent optimization leads to a successful minimization of the objective function. The optimal molecular parameters (hypothetical fluid) and the optimal operating process conditions obtained in the CoMT step define the best overall performance. The result is a lower bound to the minimization problem, as the best solution of the

Table 2.3: Comparison of starting point, the result of the CoMT-step and the result of the CoMT-CAMD method for polar solvents. The CoMT-step defines the optimal (hypothetical) polar molecule. The "mapping"-step (which completes CoMT-CAMD) identifies the real molecule with best overall performance.

	starting molecule	optimal polar molecule	best polar real molecule
	N-methyl-2-pyrrolidone	hypothetical	propylene carbonate
$M/\mathrm{g/mol}$	99.131	100.47	102.09
$m/-$	3.2008	3.2860	3.3130
$\sigma/$Å	3.5095	3.3814	3.3598
$\varepsilon/k/\mathrm{K}$	314.59	308.57	313.92
μ/D	4.0771	4.9655	4.9765
	optimal process conditions		
$P^{\mathrm{des}}/\mathrm{MPa}$	0.228	0.235	0.236
$P^{\mathrm{rg}}/\mathrm{MPa}$	0.650	0.631	0.633
$T^{\mathrm{lean}}/\mathrm{K}$	276.0	275.7	276.4
$m_S/\mathrm{mg/Nm^3}$	19.28	4.143	2.697
$f_{\text{electr.demand}}/€/\mathrm{T_{CO_2\,capt.}}$	7.143	7.067	7.107
$f_{\text{H2 loss}}/€/\mathrm{T_{CO_2\,capt.}}$	0.016	0.005	0.004
$f_{\text{solvent loss}}/€/\mathrm{T_{CO_2\,capt.}}$	0.375	0.071	0.044
$f_{\text{utilities}}/€/\mathrm{T_{CO_2\,capt.}}$	0.569	0.518	0.518
$f/€/\mathrm{T_{CO_2\,capt.}}$	8.104	7.662	7.674

minimization problem that is mathematically achievable with the selected thermodynamic model (PC-SAFT). The cost break-down on the various specific cost terms illustrates the benefit of using a single objective function: although some terms of the objective function are clearly dominant, the process trade-offs are balanced to enable an overall performance optimization. The best real component determined in the mapping-step is similar to the optimal hypothetical fluid, both, in terms of the pure component parameters and in terms of the corresponding process conditions. The performance of the best real component lies in the immediate vicinity of the optimum.

Tables 2.2 and 2.3 additionally show that state-of-the-art separation processes can be substantially improved, when solvent parameters and process conditions are optimized simultaneously. We compare the optimal performance of DEPG-5 (a typical component of the Selexol mixture) to the optimal performance of the best real molecules (1,5,9-cyclododecatriene and propylene carbonate) as measured by our objective function. There is a significant reduction of the specific cost: by $12\,\%$ for the best non-polar component and by $27\,\%$ for the best polar component. This improvement is even more significant if one takes into account that some cost contributions are fixed and cannot be improved by the solvent. Namely, the

compression of the CO_2 captured in the absorber (which is the same for all cases) from a pressure of $2\,\mathrm{MPa}$ (syngas feed) to the pipeline pressure of $11\,\mathrm{MPa}$ leads to a cost of $2\,€/\mathrm{T}_{\mathrm{CO_2capt}}$.

The CoMT-CAMD method approximates the solution of the full MINLP problem in Eq.(2.1). More specifically it is the mapping step, i.e. the Taylor expansion Eq.(2.3) used to identify best performing real substances and the corresponding best process settings $\bar{x}$, where the procedure is approximate. In order to assess the quality of the mapping step, we have conducted individual process optimizations for all components included in both databases. The mapping list with the 10 best non-polar components is given in Table 2.4 and for the polar components in Table 2.5. The ranking predicted in the mapping-step ('mapping ranking') is compared to the ranking after the individual process optimizations ('real ranking'). The process optimization variables $\bar{x}^{opt} = \{P^{\mathrm{des}}, P^{\mathrm{rg}}, T^{\mathrm{lean}}\}^{opt}$ of the CoMT step were used as degrees of freedom for the individual process optimizations. These optimal process conditions along with the value of the process inequality constraint at optimum $\mathrm{m_S}^{opt}$ for each substance are also listed in Tables 2.4 and 2.5. The melting point temperature T_{mp} is the lowest temperature we allowed during each individual process optimization to avoid solvent solidification.

The results presented by Tables 2.4 and 2.5 are the core results of this study. Since the approximation made in the mapping step (Eq.(2.3) and (2.4)) is the only approximation towards the real solution of the full MINLP problem (Eq.(2.1)), it is important to find the rank list from the CoMT-CAMD procedure ('mapping ranking') in good agreement to the rank list of the individually optimized components. Tables 2.4 and 2.5 confirm that the 'mapping ranking' is very similar to the 'real ranking' and that all best components are feasible with respect to the process constraint ($\mathrm{m_S}^{opt} \leq 20\,\mathrm{mg/Nm^3}$).

A comparison between the optimal process conditions for the optimal molecule (Tables 2.2 and 2.3) and the optimal process conditions for the best real components (Tables 2.4 and 2.5, respectively) show that the identified molecules have strong similarities to the optimal molecule. The PC-SAFT parameters of the hypothetical molecule take on reasonable values, in a range, where real molecules occur. This is a significant observation, since no a priori constraints on parameter combinations are applied, beyond the convex hull.

We note that the specific cost values reported in this study serve as meaningful performance indicators. They should not, however, be interpreted as full cost estimates for a Carbon Capture and Sequestration (CCS) system, as usually presented in a techno-economic analysis of CCS systems (see e.g., ref. [72] and [73]). In particular, investment cost was neglected in our work. However, the cost objective function implemented here carries essential attributes of the full cost estimation and should allow discriminating between different solvents and process conditions.

2.4.3. Assessing the proposed solvents

Since the problem of solvent selection for CO_2 capture with physical absorption is well established both in the literature and in the industrial practice, it is possible to validate (or at least evaluate) the proposed solvents. An important sign for

Table 2.4: Best non-polar candidate solvents identified by CoMT-CAMD. The ranking predicted in the mapping step ('mapping ranking') is compared to the ranking after individual process optimization ('real ranking'). The optimal process conditions $\{P^{\mathrm{des}}, P^{\mathrm{rg}}, T^{\mathrm{lean}}\}^{opt}$ are given. The value of $\mathrm{m_S}$ indicates a components' feasibility against the inequality process constraint ${\mathrm{m_S}}^{opt} < 20\,\mathrm{mg/Nm^3}$. The melting point temperature T_{mp} is the lower temperature bound to avoid solidification of the solvent.

IUPAC Name	ranking		f	${\mathrm{m_S}}^{opt}$	P_{des}^{opt}	P_{rg}^{opt}	T_{lean}^{opt}	$\mathrm{T_{mp}}$
	mapping	real	€/$\mathrm{T_{CO_2\,capt.}}$	mg/Nm³	MPa	MPa	K	K
1,5,9-cyclododecatriene	1	1	9.295	13.61	0.234	0.705	274.2	256.4
bis(alpha-methyl-benzyl)-ether	2	4	9.399	0.750	0.233	0.705	278.0	243.2
bicyclohexyl	3	3	9.351	17.51	0.253	0.676	281.8	276.8
2-ethylnapthalene	4	5	9.419	2.120	0.219	0.698	274.2	265.8
n-heptylbenzene	5	2	9.342	1.650	0.226	0.699	267.6	225.2
1,2,4-triethylbenzene	6	7	9.489	7.450	0.222	0.702	267.9	195.2
n-hexylbenzene	7	6	9.457	4.710	0.223	0.701	267.9	212.0
n-pentylbenzene	8	14	9.661	15.66	0.217	0.707	268.3	198.2
1-n-propylnapthalene	9	8	9.510	0.710	0.218	0.699	272.5	264.6
n-undecane	10	16	9.750	12.26	0.219	0.711	263.1	247.6

Table 2.5: Best polar candidate solvents identified by CoMT-CAMD. The ranking predicted in the mapping step ('mapping ranking') is compared to the ranking after individual process optimization ('real ranking'). The optimal process conditions $\{P^{\mathrm{des}}, P^{\mathrm{rg}}, T^{\mathrm{lean}}\}^{opt}$ are given. The value of $\mathrm{m_S}$ indicates a components' feasibility against the inequality process constraint ${\mathrm{m_S}}^{opt} < 20\,\mathrm{mg/Nm^3}$. The melting point temperature T_{mp} is the lower temperature bound to avoid solidification of the solvent.

IUPAC Name	ranking		f	${\mathrm{m_S}}^{opt}$	P_{des}^{opt}	P_{rg}^{opt}	T_{lean}^{opt}	$\mathrm{T_{mp}}$
	mapping	real	€/$\mathrm{T_{CO_2\,capt.}}$	mg/Nm³	MPa	MPa	K	K
propylene carbonate	1	1	7.674	2.697	0.236	0.633	276.4	224.9
gamma-valerolactone	2	2	7.795	20.00	0.195	0.674	274.5	242.2
methyl maleic anhydride	3	5	8.384	19.94	0.245	0.642	281.2	281.2
n-methyl-2-pyrrolidone	4	3	8.104	19.28	0.228	0.650	276.0	249.2
diethyl sulfate	5	6	8.407	9.190	0.229	0.668	271.1	248.2
butyric anhydride	6	7	8.524	9.940	0.230	0.677	266.1	199.9
isobutyric anhydride	7	4	8.247	20.00	0.185	0.729	262.7	219.7
diphenyldichlorosilane	8	12	8.869	0.050	0.225	0.684	269.8	251.2
4-formylmorpholine	9	27	10.18	10.47	0.213	0.684	293.7	293.7
2-ethylhexyl acrylate	10	14	9.022	4.882	0.229	0.695	265.2	183.2

the validity of the proposed method is the fact that the mapping list of the polar candidates (Table 2.5) includes three state-of-the-art solvents: Propylene Carbonate (Fluor solvent process) [3], n-Methyl-2-Pyrrolidone (Purisol process) [3] and 4-Formylmorpholine (Morphysorb process) [74]. We note that the solvents identified in this study are different from the solvents identified with CoMT-CAMD for the pre-combustion CO_2 capture in the previous work of Bardow et al. [33]. This is the result of implementing a different process topology and an economic objective function. In addition, an extended mapping database is used in this work compared with our earlier study. Generally, the identified solvents are a result of how the objective function is defined together with the process specifications we have laid out in Table 2.1. A strength of the CoMT-CAMD approach is that any change in the process specifications or the objective function very swiftly leads to a new list of promising solvents. Comparing the mapping lists of polar and non-polar components with each other (Tables 2.4 and 2.5) we observe, that the additional degree of freedom of polar substances is used in the optimization, leading to a lower value of the objective function. We also see that the best performing molecules possess certain functional groups. Almost all of the best performing polar molecules contain carbonylic groups, while phenyl rings are present in almost all identified polar and non-polar components. The phenyl rings and the carbonylic oxygen provide favorable interactions with CO_2. These findings are reasonable and confirm earlier studies: Karazian et al. [75] demonstrated the specific interactions between CO_2 and the carbonylic oxygen using IR spectroscopy. Nelson et al. [76] showed, using ab initio calculations, that CO_2 will primarily interact with the carbonylic oxygen and that interactions with the phenyl ring can also be present but are weaker. Similar observations about the presence of certain functional groups in solvents for the physical absorption of CO_2 are also discussed in the work of Miller et al. [11].

Enick et al. [10] mentioned siloxanes as potential solvents for the physical absorption of CO_2 due to good CO_2 solubilities. In our work, siloxanes are considered during the mapping of polar components. They are feasible components regarding the process constraint but they didn't appear in the 10 best polar components.

DEPGs are state-of-the-art solvents and they are feasible components according to our model but they are not included in the top-list of non-polar solvents (Table 2.4). The ether-oxygen in the DEPG molecule provides also a favorable interaction site for CO_2, leading to relatively high CO_2 solubility. In CoMT-CAMD though components are evaluated for their overall performance. DEPGs, mainly due to their high molecular volume and the low absorption temperature required, lead to higher electricity demand and utilities' cost. The result is that according to our objective function they demonstrate a lower overall performance than small molecules with phenyl rings. If solvent loss gets a higher economic weight than in our objective function (e.g. by introducing an additional economic penalty for air pollution), the findings can be very different. DEPG components have a low vapor pressure leading to very low solvent loss (cf. Table 2.2).

The objective function evaluates the solvent performance beyond individual solvent properties and leads to molecules that: have sufficiently small molecular volume (to keep the specific cost of electricity demand in the pump low), are selec-

tive towards CO_2 and not towards H_2 (to achieve the same CO_2 absorption rate for smaller solvent volumetric flows and lower hydrogen loss), are hydrophobic (to avoid water accumulation in the system), are less volatile and achieve the necessary cooling on a low cost.

Out of 612 non-associating components with diverse structural properties, CoMT-CAMD identifies those substances that optimize the overall process performance, without any prior judgment on favored functional groups and without using heuristics. The CoMT-CAMD results allow for conclusions about the proper molecular structure of the solvent and the optimal process conditions for a defined objective function through a continuous, deterministic mathematical procedure. In the mapping step, the database search could also be replaced by a molecular design algorithm to generate new molecular structures. Instead of mapping the optimal molecular parameters onto an existing database of pure component PC-SAFT parameters, one can map the optimal molecular parameters onto PC-SAFT parameters derived from a group-contribution approach. This would allow identifying substances beyond an existing database. This combinatorial approach requires a group contribution approach [77] and has already been implemented for the design of ORC fluids [78].

CoMT-CAMD also identifies new components which were previously not regarded for CO_2 capture. An interesting example is the component gamma-Valerolactone, which is the second best polar solvent in the mapping list (Table 2.5). Gamma-Valerolactone has already been described by other authors as a promising green solvent for bio-refineries. [79, 80]

Further refinements of the flowsheet are possible, for example by considering additional flashes for the desorption or a heat integration scheme [81]. Any modification of the process specifications or of the objective function will result in a new rank list of solvents together with corresponding optimal process settings.

2.5. Conclusions

We analyze and further develop CoMT-CAMD for the simultaneous process and solvent optimization by considering a physical absorption process, namely a pre-combustion capture process of CO_2, where the absorption solvent is a degree of freedom. We developed QSPR models for the ideal gas heat capacity and the molar mass of the optimized solvent based on pure component parameters of the PC-SAFT equation of state and found very satisfying accuracy of these approaches for organic solvents. With these extensions, PC-SAFT provides a unified thermodynamic framework for predicting solvent properties throughout all steps of a systematic solvent and process optimization using CoMT-CAMD.

We observed that the rank list from the CoMT-CAMD method is a good approximation to the real rank list of the full MINLP problem of solvent and process design. We came to this conclusion by individually optimizing the process conditions of 612 substances (representing our database) in order to assess the results obtained by CoMT-CAMD. The agreement between the predicted rank list of best solvents is in very satisfying agreement to the actual rank list of individually optimized solvents. Moreover the predicted optimal process setting to each member of the most promis-

ing solvents is in very good agreement to the individually optimized process setting. Further, the predicted rank list of best solvents includes only feasible components without a priori pre-selecting fluids.

The process topology examined here represents the CO_2-capture process of an existing IGCC power plant [53], including all three major steps of absorption, desorption and CO_2 compression and an emission limit for the solvent as a process constraint. CoMT-CAMD allows to measure the performance of the system with a single objective function that can be formulated freely according to the problem at hand. Our cost function summarizes several contributions to simultaneously evaluate all process trade-offs on a common basis. Since pre-combustion capture of CO_2 is a well-studied process, it was reassuring to find that the rank-list of CoMT-CAMD comprises several state-of-the-art solvents. But, also new promising solvent candidates were identified.

CoMT-CAMD is a deterministic approach that does not involve any pre-selection or heuristic decisions. It is a versatile and generic method. CoMT-CAMD can be implemented with any physically based thermodynamic model. It is a suitable method for the integrated process and molecular design for a wide spectrum of existing applications as well as for new process designs, when limited process know-how is available.

References

[1] J. Urech, L. Tock, T. Harkin, A. Hoadley, and F. Marechal, *An assessment of different solvent-based capture technologies within an IGCC-CCS power plant,* Energy **64**, 268 (2014).

[2] H. Zhai and E. S. Rubin, *Comparative performance and cost assessments of coal- and natural-gas-fired power plants under a* CO_2 *emission performance standard regulation,* Energy Fuels **27**, 4290 (2013).

[3] A. L. Kohl and R. B. Nielsen, *Gas Purification* (Gulf Publishing Company, 1997).

[4] R. Bucklin and R. Schendel, *Comparison of Fluor Solvent and Selexol processes,* Energy Prog.; (United States) **4:3** (1984).

[5] K. Thambimuthu, M. Soltanieh, J. C. Abanades, R. Allam, O. Bolland, J. Davison, P. Feron, F. Goede, A. Herrera, M. Iijima, D. Jansen, I. Leites, P. Mathieu, E. Rubin, D. Simbeck, K. Warmuzinski, M. Wilkinson, R. Williams, M. Jaschik, A. Lyngfelt, R. Span, and M. Tanczyk, *IPCC special report on carbon dioxide capture and storage,* in *IPCC*, edited by Z. Abu-Ghararah, T. Yashima, B. Metz, O. Davidson, H. de Coninck, M. Loos, and L. Meyer (Cambridge University Press, 2005).

[6] A. Padurean, C.-C. Cormos, and P.-S. Agachi, *Pre-combustion carbon dioxide capture by gas-liquid absorption for integrated gasification combined cycle power plants,* Int. J. Greenh. Gas Control **7**, 1 (2012).

[7] B. P. Spigarelli and S. K. Kawatra, *Opportunities and challenges in carbon dioxide capture,* J. CO_2 Util. **1**, 69 (2013).

[8] M. E. Boot-Handford, J. C. Abanades, E. J. Anthony, M. J. Blunt, S. Brandani, N. Mac Dowell, J. R. Fernandez, M.-C. Ferrari, R. Gross, J. P. Hallett, R. S. Haszeldine, P. Heptonstall, A. Lyngfelt, Z. Makuch, E. Mangano, R. T. J. Porter, M. Pourkashanian, G. T. Rochelle, N. Shah, J. G. Yao, and P. S. Fennell, *Carbon capture and storage update,* Energy Environ. Sci. **7**, 130 (2014).

[9] Y. J. Heintz, L. Sehabiague, B. I. Morsi, K. L. Jones, and H. W. Pennline, *Novel physical solvents for selective* CO_2 *capture from fuel gas streams at elevated pressures and temperatures* , Energy Fuels **22**, 3824 (2008).

[10] R. M. Enick, P. Koronaios, C. Stevenson, S. Warman, B. Morsi, H. Nulwala, and D. Luebke, *Hydrophobic polymeric solvents for the selective absorption of CO_2 from warm gas streams that also contain H_2 and H_2O,* Energy Fuels **27**, 6913 (2013).

[11] M. B. Miller, D. R. Luebke, and R. M. Enick, *CO_2-philic oligomers as novel solvents for CO_2 absorption,* Energy Fuels **24**, 6214 (2010).

[12] B. Gwinner, D. Roizard, F. Lapicque, E. Favre, R. Cadours, P. Boucot, and P.-L. Carrette, *CO_2 capture in flue gas: Semiempirical approach to select a potential physical solvent,* Ind. Eng. Chem. Res. **45**, 5044 (2006).

[13] J. Wilcox, P. Rochana, A. Kirchofer, G. Glatz, and J. He, *Revisiting film theory to consider approaches for enhanced solvent-process design for carbon capture,* Energy Environ. Sci. **7**, 1769 (2014).

[14] C. S. Adjiman, A. Galindo, and G. Jackson, *Molecules matter: The expanding envelope of process design,* Comput. Chem. Eng. **34**, 55 (2014).

[15] O. Odele and S. Macchietto, *Computer aided molecular design: a novel method for optimal solvent selection,* Fluid Phase Equilib. **82**, 47 (1993).

[16] A. P. Duvedi and L. E. Achenie, *Designing environmentally safe refrigerants using mathematical programming,* Chem. Eng. Sci. **51**, 3727 (1996).

[17] N. V. Sahinidis, M. Tawarmalani, and M. Yu, *Design of alternative refrigerants via global optimization,* AIChE J. **49**, 1761 (2003).

[18] R. Gani, *Chapter 14 Case studies in chemical product design use of CAMD techniques,* in *Chemical Product Design: Toward a Perspective Through Case Studies*, Comput. Aided Chem. Eng., Vol. 23, edited by R. G. Ka M. Ng and K. Dam-Johansen (Elsevier, 2007) pp. 435 – 458.

[19] A. Papadopoulos, S. Bard, A. Chremos, E. Forte, T. Zarogiannis, P. Seferlis, S. Papadokonstantakis, C. Adjiman, A. Galindo, and G. Jackson, *Efficient screening and selection of post-combustion CO_2 capture solvents,* Chem. Eng. Trans. **39**, 211 (2014).

[20] J. Salazar, U. Diwekar, K. Joback, A. H. Berger, and A. S. Bhown, *Solvent selection for post-combustion CO_2 capture,* Energy Procedia **37**, 257 (2013).

[21] M. Mattei, G. M. Kontogeorgis, and R. Gani, *A comprehensive framework for surfactant selection and design for emulsion based chemical product design,* Fluid Phase Equilib. **362**, 288 (2014).

[22] K. V. Gernaey and R. Gani, *A model-based systems approach to pharmaceutical product-process design and analysis,* Chem. Eng. Sci. **65**, 5757 (2010).

[23] E. Marcoulaki and A. Kokossis, *On the development of novel chemicals using a systematic synthesis approach. Part I. Optimisation framework,* Chem. Eng. Sci. **55**, 2529 (2000).

[24] E. Marcoulaki and A. Kokossis, *On the development of novel chemicals using a systematic optimisation approach. Part II. Solvent design,* Chem. Eng. Sci. **55**, 2547 (2000).

[25] S. Bommareddy, N. G. Chemmangattuvalappil, C. C. Solvason, and M. R. Eden, *Simultaneous solution of process and molecular design problems using an algebraic approach,* Comput. Chem. Eng. **34**, 1481 (2010).

[26] A. P. Samudra and N. V. Sahinidis, *Optimization-based framework for computer-aided molecular design,* AIChE J. **59**, 3686 (2013).

[27] A. I. Papadopoulos and P. Linke, *Integrated solvent and process selection for separation and reactive separation systems,* Chem. Eng. Process. **48**, 1047 (2009).

[28] N. G. Chemmangattuvalappil and M. R. Eden, *A novel methodology for property-based molecular design using multiple topological indices,* Ind. Eng. Chem. Res. **52**, 7090 (2013).

[29] M. Lampe, M. Stavrou, H. M. Bucker, J. Gross, and A. Bardow, *Simultaneous optimization of working fluid and process for organic rankine cycles using PC-SAFT,* Ind. Eng. Chem. Res. **53**, 8821 (2014).

[30] F. Pereira, E. Keskes, A. Galindo, G. Jackson, and C. Adjiman, *Integrated solvent and process design using a SAFT-VR thermodynamic description: High-pressure separation of carbon dioxide and methane,* Comput. Chem. Eng. **35**, 474 (2011).

[31] F. E. Pereira, E. Keskes, A. Galindo, G. Jackson, and C. S. Adjiman, *Integrated design of* CO_2 *capture processes from natural gas,* in *Process Systems Engineering* (Wiley-VCH Verlag GmbH & Co. KGaA, 2008) pp. 231–248.

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

[33] A. Bardow, K. Steur, and J. Gross, *Continuous-Molecular Targeting for integrated solvent and process design,* Ind. Eng. Chem. Res. **49**, 2834 (2010).

[34] M. Lampe, J. Gross, and A. Bardow, *Simultaneous process and working fluid optimisation for organic rankine cycles (orc) using PC-SAFT,* in *22nd European Symposium on Computer Aided Process Engineering*, Computer Aided Chemical Engineering, Vol. 30, edited by I. D. L. Bogle and M. Fairweather (Elsevier, 2015) pp. 572 – 576.

[35] 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**, 1244 (2001).

[36] J. Gross and G. Sadowski, *Application of the Perturbed-Chain SAFT equation of state to associating systems,* Ind. Eng. Chem. Res. **41**, 5510 (2002).

[37] J. Gross, *An equation-of-state contribution for polar components: Quadrupolar molecules,* AIChE J. **51**, 2556 (2005).

[38] J. Gross and J. Vrabec, *An equation-of-state contribution for polar components: Dipolar molecules,* AIChE J. **52**, 1194 (2006).

[39] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, *The Quickhull algorithm for Convex Hulls,* ACM Trans. Math. Softw. **22**, 469 (1996).

[40] M. Kleder, *Matlab(r) Central The Mathworks Inc.* .

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

[42] M. S. Wertheim, *Fluids with highly directional attractive forces. I. Statistical thermodynamics,* J. Stat. Phys. , 19 (1984).

[43] M. S. Wertheim, *Fluids with highly directional attractive forces: II. Thermodynamic perturbation theory and integral equations,* J. Stat. Phys. , 35 (1984).

[44] L. F. Vega and G. Jackson, *20 years of the SAFT equation of state-recent advances and challenges: Symposium held in Bellaterra, Barcelona, 19-21 September 2010,* Fluid Phase Equilib. **306**, 1 (2011).

[45] X. Tang and J. Gross, *Modeling the phase equilibria of hydrogen sulfide and carbon dioxide in mixture with hydrocarbons and water using the PCP-SAFT equation of state,* Fluid Phase Equilib. **293**, 11 (2010).

[46] 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 Equilib. **248**, 29 (2006).

[47] A. Grenner, G. M. Kontogeorgis, N. von Solms, and M. L. Michelsen, *Application of PC-SAFT to glycol containing systems - PC-SAFT towards a predictive approach,* Fluid Phase Equilib. **261**, 248 (2007).

[48] N. R. Nannan, C. M. De Servi, T. van der Stelt, P. Colonna, and A. Bardow, *An equation of state based on PC-SAFT for physical solvents composed of polyethylene glycol dimethylethers,* Ind. Eng. Chem. Res. **52**, 18401 (2013).

[49] N. Mac Dowell, F. E. Pereira, F. Llovell, F. J. Blas, C. S. Adjiman, G. Jackson, and A. Galindo, *Transferable saft-vr models for the calculation of the fluid phase equilibria in reactive mixtures of carbon dioxide, water, and n-alkylamines in the context of carbon capture,* J. Phys. Chem. B **115**, 8155 (2011).

[50] P. Chiesa, S. Consonni, T. Kreutz, and R. Williams, *Co-production of hydrogen, electricity and* CO_2 *from coal with commercially ready technology. Part A: Performance and emissions,* Int. J. Hydrogen Energy **30**, 747 (2005).

[51] E. Martelli, T. Kreutz, M. Carbo, S. Consonni, and D. Jansen, *Shell coal IGCCS with carbon capture: Conventional gas quench vs. innovative configurations,* Applied Energy **88**, 3978 (2011).

[52] E. S. Rubin, C. Chen, and A. B. Rao, *Cost and performance of fossil fuel power plants with* CO_2 *and* CH_4 *capture and storage,* Energy Policy **35**, 4444 (2007).

[53] K. Damen, R. Gnutek, J. Kaptein, N. R. Nannan, B. Oyarzun, C. Trapp, P. Colonna, E. van Dijk, J. Gross, and A. Bardow, *Developments in the pre-combustion* CO_2 *capture pilot plant at the Buggenum IGCC,* Energy Procedia **4**, 1214 (2011).

[54] R. Smith, *Chemical process design and integration* (John Wiley, 2005).

[55] J. Davison, *Performance and costs of power plants with capture and storage of* CO_2*,* Energy **32**, 1163 (2007).

[56] *Commission of the european communities, energy sources, production costs and performance of technologies for power generation, heating and transport,* (2008).

[57] A. Arora and S. Kaushik, *Theoretical analysis of a vapour compression refrigeration system with R502, R404A and R507A,* Int. J. Refrig. **31**, 998 (2008).

[58] L. T. Biegler, *Nonlinear programming - concepts, algorithms, and applications to chemical processes* (MOS-SIAM, 2010).

[59] E. Union, *COUNCIL DIRECTIVE 1999/13/EC*, Tech. Rep. (1999) on the limitation of emissions of volatile organic compounds due to the use of organic solvents in certain activities and installations.

[60] I. Ziena Optimization: Evanston, *Knitro,* (2011).

[61] J. Nocedal and S. J. Wright, *Numerical optimization* (Springer, 2006).

[62] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular theory of gases* (John Wiley & Sons, 1987).

[63] M. Stavrou, M. Lampe, A. Bardow, and J. Gross, *Supporting information: Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for* CO_2 *capture,* Ind. Eng. Chem. Res. **53**, 18029 (2014).

[64] M. Stavrou, M. Lampe, A. Bardow, and J. Gross, *Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for* CO_2 *capture,* Ind. Eng. Chem. Res. **53**, 18029 (2014).

[65] R. C. Reid, J. M. Prausnitz, and J. M. Poling, *The properties of gases and liquids* (McGraw-Hill, 1954).

[66] M. L. Michelsen and J. M. Möllerup, *Thermodynamic models: Fundamentals & Computational aspects* (Tie-Line Publications, 2007).

[67] K. Bezanehtak, F. Dehghani, and N. R. Foster, *Vapor-liquid equilibrium for the carbon dioxide + hydrogen + methanol ternary system,* J. Chem. Eng. Data **49**, 430 (2004).

[68] AIChE-DECHEMA, *Pure Component Database DIPPR-801*, Tech. Rep. (AIChE-DECHEMA, 2011).

[69] A. J. Haslam, A. Galindo, and G. Jackson, *Prediction of binary intermolecular potential parameters for use in modelling fluid mixtures,* Fluid Phase Equilib. **266**, 105 (2008).

[70] Mathworks, *Matlab(r), mathworks - documentation center,* (2015).

[71] T. van Westen, T. J. H. Vlugt, and J. Gross, *Determining force field parameters using a physically based equation of state,* J. Phys. Chem. B **115**, 7872 (2011).

[72] K. Damen, M. van Troost, A. Faaij, and W. Turkenburg, *A comparison of electricity and hydrogen production systems with CO_2 capture and storage. Part A: Review and selection of promising conversion and capture technologies,* Prog. Energy Combust. Sci. **32**, 215 (2006).

[73] E. S. Rubin, C. Short, G. Booras, J. Davison, C. Ekstrom, M. Matuszewski, and S. McCoy, *A proposed methodology for CO_2 capture and storage cost estimates,* Int. J. Greenh. Gas Control **17**, 488 (2013).

[74] Thyssenkrupp-Industrial-Solutions, *uhde-brochures-pdf-en-14.pdf,* (2014).

[75] S. G. Kazarian, M. F. Vincent, F. V. Bright, C. L. Liotta, and C. A. Eckert, *Specific intermolecular interaction of carbon dioxide with polymers,* J. Am. Chem. Soc. **118**, 1729 (1996).

[76] M. R. Nelson and R. F. Borkman, *Ab initio calculations on CO_2 binding to carbonyl groups,* J. Phys. Chem. A **102**, 7860 (1998).

[77] E. Sauer, M. Stavrou, and J. Gross, *Comparison between a Homo- and a Heterosegmented Group Contribution approach based on the Perturbed-Chain Polar Statistical Associating Fluid Theory Equation of State,* Ind. Eng. Chem. Res. **53**, 14854 (2014).

[78] M. Lampe, C. Kirmse, E. Sauer, M. Stavrou, J. Gross, and A. Bardow, *Computer-aided Molecular Design of ORC working fluids using PC-SAFT,* in *Proceedings of the 8th International Conference on Foundations of Computer-Aided Process Design - FOCAPD* (2014).

[79] J. H. Clark, F. E. I. Deswarte, and T. J. Farmer, *The integration of green chemistry into future biorefineries,* Biofuels, Bioprod., Biorefin. **3**, 72 (2009).

[80] D. M. Alonso, S. G. Wettstein, and J. A. Dumesic, *Gamma-valerolactone, a sustainable platform molecule derived from lignocellulosic biomass,* Green Chem. **15**, 584 (2013).

[81] M. Gatti, E. Martelli, F. Marechal, and S. Consonni, *Review, modeling, heat integration, and improved schemes of Rectisol®-based processes for CO_2 capture,* Appl. Therm. Eng. **70**, 1123 (2014).

3

Estimation of the Binary Interaction Parameter k_{ij} of the PC-SAFT Equation of State Based on Pure Component Parameters Using a QSPR Method

This chapter has been published as: M. Stavrou, A. Bardow, J. Gross; Fluid Phase Equilib. 416 (2016) 138 – 149

Statistical Associating Fluid Theory (SAFT) equations of state (EoS) for mixtures require cross-interaction parameters. For real systems, combining rules, such as the Lorenz-Berthelot combining rules, have to be corrected using at least one binary interaction parameter, k_{ij}. Values of k_{ij} are usually adjusted to experimental data of phase equilibria. Here, we correlate k_{ij} to the pure component parameters of the Perturbed Chain - Statistical Associating Fluid Theory (PC-SAFT) EoS, using a Quantitative Structure Property Relationship (QSPR) model. The coefficients of the proposed QSPR model are regressed separately for mixtures with non-associating components and for mixtures with associating components. The QSPR model is validated using the statistical measures of the QSPR method. We compare the values of k_{ij} that are estimated from the QSPR model to values of k_{ij} estimated from London's dispersive theory. Phase equilibrium calculations carried out with these two approaches of estimating k_{ij} values are compared to experimental data. The estimation of k_{ij} values as function of the pure component PC-SAFT parameters can be applied to problems of process design and in Computer Aided Molecular Design (CAMD), to allow for calculations that are reasonably accurate and independent from the availability of experimental mixture data.

3.1. Introduction

Statistical Associating Fluid Theory (SAFT) equations of state (EoS) are expressions of the residual Helmholtz energy that represent approximate algebraic solutions of equations from statistical mechanics [1–4]. For mixtures, a SAFT EoS requires the intermolecular potential ϕ_{ij} between the unlike molecules i and j. Usually, a conformal solution theory is applied, requiring cross-interaction parameters, such as a size-parameter σ_{ij} and an energy parameter ε_{ij}. Strictly speaking the $i-j$ cross-interaction potential, just as the potential between like $i-i$ species, has to be determined by quantum mechanical calculations. For van der Waals (dispersive) interactions, however, combining rules are often successful for estimating of cross-interaction parameters. Most variants of the SAFT-type EoS use the Lorenz-Berthelot combining rules: $\sigma_{ij} = 1/2(\sigma_i + \sigma_j)$ and $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$. For an accurate correlation of real systems, the combining rules have to be corrected using at least one interaction parameter, k_{ij} for every binary pair. As a modification of the Lorenz-Berthelot combining rules, the binary interaction parameter k_{ij} is introduced as a correction to the dispersive energy parameter for the binary pair: $\varepsilon_{ij} = (1-k_{ij})\sqrt{\varepsilon_i \varepsilon_j}$. Binary interaction parameters are often adjusted to experimental phase equilibrium data of the binary mixtures.

SAFT-type EoS were successfully applied to a wide range of complex systems, that include polar and associating components, polymers, ionic liquids, pharmaceuticals and bio-molecules (see for example ref. [1–3, 5–8]). SAFT-type EoS are often more accurate and more predictive than cubic EoS and extrapolate more reliably than g^E-models [1]. For mixtures of substances where the intermolecular potentials are dominated by van der Waals interactions, SAFT-type EoS often give accurate results, even without binary correction ($k_{ij} \simeq 0$). Many mixtures of industrial interest however require adjusting a binary interaction parameter k_{ij}.

Experimental data of phase equilibria are not always available for the estimation of k_{ij} values. Consider the design of novel processes or a molecular design problem. In Computer Aided Molecular Design (CAMD) for example, thermodynamic models are required to evaluate some objective function quantifying the performance of the designed molecule [9, 10]. CAMD problems are formulated as reverse property prediction problems [11]. In CAMD problems it is not possible to adjust binary interaction parameters on experimental mixture data. Early stages of process design and CAMD rely on predictions of mixture behavior, which in our context means predictions of k_{ij} values.

O'Connell et al. [10] studied several thermodynamic property models and suggested SAFT-type EoS to have a potential for accurate property predictions in CAMD. In a recent study, Ng et al. [12] gave a thorough review of the various approaches and advances in CAMD methods. Among the most recent advances are CAMD frameworks that use a SAFT-type model [13–19]. Adjiman et al. [20] presented the rapid progress made in the field of SAFT-based CAMD and discussed the role of SAFT-type EoS in tackling previously challenging problems in molecular and process design. In a SAFT-based CAMD framework, variables that characterize the structure of the optimized molecule are utilized as additional degrees of freedom inside the process optimization. The molecular optimization can thereby be integrated with

the process optimization problem, based on an objective function for the entire process. The problem of integrated process and fluid design has in the past been circumvented, by defining individual or collected property targets for the optimized fluid [21–23].

Pereira et al. [13, 15] introduced the framework of Computer Aided Molecular and Process Design (CAMPD) that uses the SAFT-VR EoS [24]. In their work, the authors applied CAMPD for the solvent selection among alkane blends. The molecular search space was defined over the homologous series of n-alkanes and the number of carbon atoms of the n-alkane was treated as a continuous molecular optimization variable. The necessary SAFT-VR parameters were expressed as functions of the molecular mass and therefore as functions of the optimized number of carbon atoms. The values of k_{ij} for the SAFT-VR EoS were adjusted to experimental phase equilibrium data of binary mixtures relevant to the examined system and they were held constant throughout CAMPD. Recently, Burger et al. [25] proposed a CAMPD method that uses the SAFT-γ Mie GC EoS [26] and a hierarchical optimization approach. The authors demonstrated the proposed method for the problem of solvent selection adopted from Pereira et al. In their work, Burger et al. [25] applied a group contribution (GC) approach for SAFT and considered additional chemical families extending the search space to linear alkyl ethers.

The framework of Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD), established by Bardow et al. [14], uses the PC-SAFT EoS [27–30]. CoMT-CAMD considers the parameters of the thermodynamic model (e.g. representing a solvent) as additional degrees of freedom in an integrated process and fluid optimization problem. In CoMT-CAMD the discrete search space of parameters representing a real substance is relaxed to a continuous parameter domain. Specifically, the molecular optimization variables (PC-SAFT pure component parameters) are treated as continuous variables. This relaxation allows formulating a non-linear optimization problem with a single objective function, whereby a detailed process model can be maintained without the need for pre-selecting candidate molecules. CoMT-CAMD has been implemented, so far, for the selection and design of working fluids for organic rankine cycles (ORCs) [16, 19] and for the selection and design of physical solvents for CO_2 capture [18, 19]. In phase equilibrium calculations no binary interaction correction was introduced to the PC-SAFT EoS for binary systems involving the optimized fluid ($k_{ij} = 0$).

The result of a CAMD approach is determined by the employed thermodynamic model. The accuracy of the SAFT model therefore plays a crucial role in the predictive capability of the CAMD framework.

In order to improve predictions of mixture properties (with $k_{ij} = 0$), many studies have revisited the Lorentz-Berthelot combining rules [31–38]. More recently, Haslam et al. [39] derived a new combining rule using a generalization of the Hudson-McCoubrey combining rules, including additional terms for the dipole-dipole and the dipole-induced dipole interactions. The authors presented results of their theory with the SAFT-VR EoS [24] for mixtures with large non-polar and polar components and for mixtures with a single associating component. Singh et al. [40] and Leonhard et al. [41] proposed the prediction of pure component param-

eters based on quantum mechanical calculations and a new combining rule for the unlike-dispersion interaction, derived from London's dispersive theory. Leonhard et al. applied the proposed combining rule with the PC-SAFT EoS. Both of the above approaches have a theoretical background and gave good results for mixtures.

Group contribution approaches are promising for estimating binary interaction corrections k_{ij} to the Lorenz-Berthelot combining rules. Peters et al. [42] proposed a group contribution method for determining binary interaction parameters of the PC-SAFT EoS for polymer-solvent mixtures. Inspired by London's dispersive theory, Huynh et al. [43, 44] proposed the correlation of k_{ij} to pseudo-ionization energies of the constituents of the mixture. The method was developed for the GC-SAFT EoS [45] and was shown to lead to very good results. London's dispersive theory applied to the Mie potential function has been used by Coutinho et al. [46] to derive a new combining rule for the cross-energy parameter in cubic EoS and in the Cubic Plus Association EoS [47].

Due to the analogy between the combining rules of SAFT-type EoS and the combining rules in cubic EoS, we mention the work of Shacham et al. [48] and the work of Abudour et al. [49]. Both studies proposed the prediction of k_{ij} values for cubic EoS using a Quantitative Structure Property Relationship (QSPR) method with automatically generated molecular descriptors.

The CoMT-CAMD approach requires the prediction of mixture properties (i.e. estimating a suitable k_{ij}) based on pure component parameters of PC-SAFT for the optimized fluid. On that grounds, two of the approaches for estimating k_{ij} mentioned above particularly lend themselves for the CoMT-CAMD framework: the estimation of k_{ij} based on predicted ionization potentials and the QSPR estimation method.

In this work, we relate the binary interaction parameter k_{ij} to pure component parameters of both substances, in order to allow for predicting mixture properties in CAMD applications, especially in the CoMT-CAMD framework. Two approaches are examined. Values of k_{ij} are estimated based on London's dispersive theory using experimental ionization potentials. Further, we propose a multilinear regression model for estimating k_{ij} based only on the PC-SAFT pure component parameters. For the development of the proposed model, a QSPR method is applied.

3.2. Estimation of k_{ij} based on London's dispersive theory

For a mixture of conformal fluids, i.e. fluids with the same functional form of van der Waals intermolecular pair potentials, theoretical expressions of the k_{ij} value can be derived from London's dispersive theory. Most common is the expression based on the work of Hudson and McCoubrey [50]. Hudson and McCoubrey wrote the equation of London's attractive potential ϕ_{ij}^{disp} as function of the ionization potentials I_i and I_j and the (scalar valued) static polarizabilities α_i and α_j of the two constituents of the mixture

$$\phi_{ij}^{\mathrm{disp}} = -\frac{3}{2}\frac{I_i I_j}{(I_i + I_j)}\frac{\alpha_i \alpha_j}{r_{ij}^6} \tag{3.1}$$

Eq.(3.1) combined with the attractive part of the Lennard-Jones potential

$$\phi_{ij}^{\mathrm{LJ}} = 4\varepsilon_{ij}\left(\left[\frac{\sigma_{ij}}{r_{ij}}\right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}}\right]^{6}\right) \tag{3.2}$$

and using the arithmetic mean for the segment diameter parameter σ_{ij}

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right) \tag{3.3}$$

led to an expression for the depth of the attractive potential well due to dispersive interactions ε_{ij}, i.e. the Hudson-McCoubrey rules

$$\varepsilon_{ij} = \left[2 \cdot \frac{\left(I_i I_j\right)^{1/2}}{\left(I_i + I_j\right)}\right] \cdot \left[2^6 \cdot \frac{\sigma_i^3 \sigma_j^3}{\left(\sigma_i + \sigma_j\right)^6}\right] \cdot \sqrt{\varepsilon_i \varepsilon_j} \quad . \tag{3.4}$$

For a SAFT-type EoS Eq.(3.4) leads to an approximation of k_{ij} as function of the segment diameter parameters and the ionization potentials of the two Lennard-Jones fluids in the mixture

$$k_{ij} = 1 - \left[2 \cdot \frac{\left(I_i I_j\right)^{1/2}}{\left(I_i + I_j\right)}\right] \cdot \left[2^6 \cdot \frac{\sigma_i^3 \sigma_j^3}{\left(\sigma_i + \sigma_j\right)^6}\right] \quad . \tag{3.5}$$

For details on the derivation of Eq.(3.4) and Eq.(3.5) we refer to the original work of Hudson and McCoubrey [50] and to the study of Haslam et al. [39].

The expression in Eq.(3.5), derived from London's theory, accounts only for the asymmetry in dispersive interactions. For mixtures of polar or associating components, however, k_{ij} does not only serve as correction to the dispersive intermolecular potential, as already pointed out by Hiza and Duncan [32] and Kontogeorgis [51]. Rather, k_{ij} corrects a SAFT-type model regarding any other model deficiency, including those arising from attractive interactions, not explicitly accounted for. However, London's dispersive theory is appealing for (CoMT-) CAMD applications, and more generally for predicting mixture properties, because only pure component properties of the mixtures' constituents are required. In this work, we evaluate phase equilibria predicted using k_{ij} values from Eq.(3.5) applied with the PC-SAFT model, using experimental values for the required ionization potentials.

3.3. Multivariate regression model for k_{ij} prediction

3.3.1. Contribution of asymmetric intermolecular potentials to the value of k_{ij}

QSPR studies focus mainly on the prediction of properties for pure substances. In most QSPR studies, QSPR software like CODESSA [52] or DRAGON [53] is employed to generate constitutional, topological, geometrical, electrostatic, quantum chemical or thermodynamic molecular descriptors (e.g. ref. [49, 54, 55]). The

significant descriptors are then selected using various stochastic or deterministic methods. Subsequently, the selected descriptors are combined to generate QSPR models. The resulting QSPR models are evaluated and compared based on multivariate statistics. In this study, we aim to develop a QSPR model for the estimation of k_{ij}, whereby k_{ij} is a mixture attribute. While a large number of descriptors are available for pure substances, the number of descriptors characterizing molecular pairs is very limited [56]. Here, we define the QSPR descriptors in an ad hoc manner, based on two principles. The descriptors for the estimation of k_{ij} should: First, be a function of the PC-SAFT pure component parameters only, and second, effectively relate dissimilarities in the molecular structure of two components to the k_{ij} value of their binary mixture.

In PC-SAFT, each pure component is identified by a unique set of molecular parameters: the segment number m, the segment size parameter σ, the dispersive energy parameter ε/k, the dipole moment μ, the quadrupole moment Q, the association energy parameter ε^{AB}/k and the effective association volume κ^{AB}. The equation of the QSPR model for the estimation of the value of k_{ij} (k_{ij}^{QSPR}) as function of the molecular parameters of the mixture components is therefore

$$k_{ij}^{\mathrm{QSPR}} = \sum_{L=1}^{N_d} c_L \cdot D_L(\overline{p}_i, \overline{p}_j) \tag{3.6}$$

where N_d is the number of descriptors, $D_L(\overline{p}_i, \overline{p}_j)$ are the descriptors as function of the PC-SAFT molecular parameters $\overline{p}_i = \{m_i, \sigma_i, \varepsilon_i/k, \mu_i, Q_i, \varepsilon_i^{AB}/k\}$ and $\overline{p}_j = \{m_j, \sigma_j, \varepsilon_j/k, \mu_j, Q_j, \varepsilon_j^{AB}/k\}$ of the two components, i and j, of the mixture, and c_L are the corresponding regression coefficients.

Let $p_{k,i}$ be the k^{th} element of the PC-SAFT pure component parameter vector $\overline{p}_i$, e.g. $p_{1,i} = m_i$. Relations such as ratios of the pure component parameters, $a_{ij} = p_{k,i}/p_{k,j}$ or absolute differences, $\delta_{ij} = |p_{k,i} - p_{k,j}|$ can be used as measures of the asymmetry of intermolecular potentials in the binary mixture. Parameter ratios a_{ij} and absolute differences δ_{ij} can also be defined over combinations of the PC-SAFT parameters: $a_{ij} = h(\overline{p}_i)/h(\overline{p}_j)$ and $\delta_{ij} = |h(\overline{p}_i) - h(\overline{p}_j)|$. The departure of a_{ij} from unity and of δ_{ij} from zero quantify the difference between the $i-i$ and the $j-j$ property for the two components of the mixture, respectively. In order to make the parameter ratio a_{ij} invariant for interchanging the component indices i and j for the pair of substances, we further define the ratio operator

$$\langle a_{ij} \rangle = \begin{cases} a_{ij}^{\frac{(a_{ij}-1)}{|a_{ij}-1|}} & a_{ii} \neq a_{jj} \\ 1 & a_{ii} = a_{jj} \end{cases} \tag{3.7}$$

Measures of this type were combined to form the candidate descriptors for the QSPR model.

Unlike intermolecular potentials are caused by asymmetric dispersive, polar and associating forces. In order to quantify the contribution of the aforementioned

asymmetries to the k_{ij} value, we defined eight descriptors. The initial form of the descriptors was mainly motivated by parameter combinations as they appear in the mathematical formulation of the PC-SAFT EoS. The exponents used in the final mathematical formulation of the descriptors are defined empirically.

Dispersive interactions

We define the descriptor D^{LJ} to express the asymmetry in the dispersive intermolecular potential, as

$$D^{\text{LJ}} = 1 - \left\langle \frac{\sigma_i^3(\varepsilon_i/k)^2}{\sigma_j^3(\varepsilon_j/k)^2} \right\rangle \tag{3.8}$$

by taking into account the segment diameter σ_i, σ_j and the Lennard-Jones energy potentials $\varepsilon_i/k, \varepsilon_j/k$. Our starting point for this descriptor has been the ratio $\langle \frac{m_i^2\sigma_i^3(\varepsilon_i/k)^2}{m_j^2\sigma_j^3(\varepsilon_j/k)^2} \rangle$ as motivated from the first order term of the perturbation theory for dispersive attraction. The exponents were varied empirically, usually in the integer-range of ± 2. D^{LJ} will approach zero for very similar components and if no other types of interactions exist (e.g. polar or associating interactions).

Polar interactions

For contributions to the value of k_{ij} due to polar interactions, we define four descriptors. The contribution due to asymmetry in dipole-dipole interactions is expressed through the descriptors $D^{dd,a}$ and $D^{dd,b}$. Descriptor $D^{dd,a}$ measures the absolute difference in the reduced dipole moments.

$$D^{dd,a} = \left| \frac{\mu_i}{\sqrt{m_i\sigma_i^3(\varepsilon_i/k)}} - \frac{\mu_j}{\sqrt{m_j\sigma_j^3(\varepsilon_j/k)}} \right| \tag{3.9}$$

The difference in the reduced dipole moments contributes in a different way to the asymmetry in mixtures with only one dipolar component than in the case when both mixture components are dipolar. In order to distinguish these two cases, we introduce the descriptor $D^{dd,b}$.

$$D^{dd,b} = \sqrt{\mu_i\mu_j} \cdot \left(\mu_i - \mu_j\right)^2 \tag{3.10}$$

A further contribution due to asymmetry in quadrupole-quadrupole interactions is given by the descriptor D^{qq}. Descriptor D^{qq} considers the difference of the scaled quadrupole moments over the potential energy of the chain molecule. The difference in the quadrupole moments is scaled by the ratio of segment diameters according to

$$D^{qq} = \left[\frac{Q_i}{m_i(\varepsilon_i/k)} - \frac{Q_j}{m_j(\varepsilon_j/k)} \right]^2 \cdot \left\langle \frac{\sigma_i^5}{\sigma_j^5} \right\rangle \tag{3.11}$$

Finally, we account for the case that dipole-dipole and quadrupole-quadrupole interactions occur simultaneously with the descriptor D^{dq}.

$$D^{dq} = \left| \frac{\mu_i}{\sqrt{m_i \sigma_i^3 (\varepsilon_i/k)}} - \frac{\mu_j}{\sqrt{m_j \sigma_j^3 (\varepsilon_j/k)}} \right| \cdot \left[\frac{Q_i}{m_i(\varepsilon_i/k)} - \frac{Q_j}{m_j(\varepsilon_j/k)} \right]^2 \tag{3.12}$$

Association

Naive combining rules are not suited to describe cross-association of two substances, because cross-association can not be detached from electrostatic concepts of partial charge distributions in individual molecules. In that light, it is surprising that simple combining rules have shown promising results for a collection of chemical species [28, 51, 57]. (A group contribution approach on the other hand is in our view a rather promising concept). In studies where combining rules for cross-association are applied, a binary interaction correction is often applied to the dispersive interactions, rather than to the cross-association.

Consequently, mixtures with (cross-)associating substances are demanding for our approach, where the binary mixture shall be described based on pure component parameters of both species. With some reservation, we include associating mixtures in this study. It is important to note that the k_{ij} parameters adjusted to such mixtures usually show rather high (positive or negative) values, because of the uncertainty in cross-associating interactions. We account for the contribution of self- and cross-association effects with three descriptors. Descriptor $D^{assoc,s}$ measures the strength of self-association for component i against the strength of self-association for component j.

$$D^{assoc,s} = \sqrt{(\varepsilon_i/k)(\varepsilon_j/k)} \cdot \left[\sigma_i^3 \left(\varepsilon_i^{AB}/k \right) - \sigma_j^3 \left(\varepsilon_j^{AB}/k \right) \right]^2 \tag{3.13}$$

Kontogeorgis and Folas [58] distinguish between five different types of cross-association. We turn to the effect of cross-association between two dipolar components, when at least one of them is self-associating. Kleiner and Sadowski [57] proposed an approach to account for the cross-association that is likely to occur for such mixtures; they refer to these interactions as 'induced association'. We also observed that the combining rule for induced association [28, 59]

$$\varepsilon^{A_i B_j} = \frac{1}{2} \left(\varepsilon_i^{AB} + \varepsilon_j^{AB} \right) \tag{3.14}$$

$$\kappa^{A_i B_j} = \sqrt{\kappa^{A_i B_i} \kappa^{A_j B_j}} \cdot \left(\frac{\sigma_i \sigma_j}{1/2 \left(\sigma_i + \sigma_j \right)} \right)^3 \tag{3.15}$$

noticeably improves the accuracy of phase equilibrium calculations for this type of mixtures. The descriptor for the contribution due to induced association is defined as

$$D^{assoc,c} = \sqrt{\mu_i \mu_j} \cdot \frac{\left[(\varepsilon_i^{AB}/k) + (\varepsilon_j^{AB}/k) \right]^2}{\sqrt{\left(m_i \sigma_i^3 \cdot m_j \sigma_j^3 \right)}} \tag{3.16}$$

The indices s and c in Eq.(3.13) and (3.16) are for "self" and "cross" association contributions, respectively. The contribution to the value of k_{ij} due to induced association accounts for the mutual strength of the dipole moments [60]. The contribution of the association energy parameters and the dipole moments in the descriptor $D^{assoc,c}$ is scaled by the molecular volume of the two components. This provides a better description for mixtures with small but strong associating molecules (e.g. mixtures of acetic acid) [61]. In mixtures with one associating and one non-polar, non-associating component, the effect of induced-association should not be accounted for. In this case, the descriptor $D^{assoc,c}$ is therefore equal to zero and no discrete decisions are necessary to distinguish between cases. The contribution of induced association in the asymmetry of a mixture is different when both components are self-associating species than when the mixture contains only one self-associating component and one dipolar, non-associating component. We introduced the third descriptor $D^{assoc,sc}$, in order to decouple the two cases when describing these types of mixtures with the same model simultaneously. The descriptor $D^{assoc,sc}$ is active only for mixtures with two self-associating components

$$D^{assoc,sc} = (\mu_i \mu_j) \cdot \sqrt{(\varepsilon_i^{AB}/k)(\varepsilon_j^{AB}/k)} \cdot \left| \left[\sigma_i^3 (\varepsilon_i^{AB}/k)\right]^2 - \left[\sigma_j^3 (\varepsilon_j^{AB}/k)\right]^2 \right| \quad . \quad (3.17)$$

3.3.2. Pure component parameters

The necessary PC-SAFT pure component parameters were either adopted from ref. [27–30, 62] or, if not available, they were identified in the present work. In those cases, the pure component parameters $(m, \sigma, \varepsilon/k, \varepsilon^{AB}/k)$ were adjusted to experimental data of vapor pressure and liquid density. We have implemented the $2B$ association scheme [63] for all associating components. The association energy parameter ε^{AB}/k and the effective association volume κ^{AB} are known to be strongly correlated, which is why a constant effective association volume $\kappa^{AB} = 0.03$ was used, as previously suggested by Ruether and Sadowski [64]. Dipole moments were taken from the DIPPR database [65] and they originate either from *ab initio* calculations or from measurements. The quadrupole moments were taken from ref. [29].

3.3.3. Database of k_{ij} values adjusted to experimental data of phase equilibria

For isolating the effect of the different descriptors, we define four classes of substances: non-associating, non-polar (nAnP), non-associating, dipolar (nAdP), non-associating, quadrupolar (nAqP) and associating, dipolar (AdP). From the four classes of pure components, we define 10 groups of binary mixtures. The 10 groups of binary systems considered in our study and the number of mixtures per group are listed in Table 3.1.

For every binary mixture in our database, we adjusted the binary interaction parameter k_{ij} using experimental vapor-liquid equilibrium (VLE) data. The selected experimental data comprised sets of isothermal (P,x) and (P,x,y) data obtained from the Dortmund Database (DDB) [66], which passed the standard thermody-

namic consistency tests, namely the point-to-point test and the area-test. [67] For mixtures with several experimental isotherms, we adjusted a single temperature-independent value of k_{ij}. The objective function for adjusting the k_{ij} values is

Table 3.1: The database of binary mixtures is divided in 10 groups (here denoted as G1 to G10). The binary mixtures are categorized according to the type of components they consist of: (a) non-associating, non-polar (nAnP), (b) non-associating, dipolar (nAdP), (c) non-associating, quadrupolar (nAqP) and (d) associating, dipolar (AdP). The number of mixtures for each subgroup in the database is listed.

		comp. j			
		nAnP	nAdP	nAqP	AdP
comp. i	nAnP	$^{(G1)}48$			
	nAdP	$^{(G2)}47$	$^{(G6)}23$		
	nAqP	$^{(G3)}44$	$^{(G4)}32$	$^{(G5)}5$	
	AdP	$^{(G7)}67$	$^{(G8)}90$	$^{(G9)}43$	$^{(G10)}53$

formulated on the combined residuals of the mole fraction of the liquid phase and the residuals of the pressure for every experimental point [68]:

$$F = \left[\sum_{l=1}^{n_{\text{exp}}} \frac{(\Delta x_l)^2 \cdot \left(\Delta P_l^{rel}\right)^2}{(\Delta x_l)^2 + \left(\Delta P_l^{rel}\right)^2} \right]^{1/2} \tag{3.18}$$

with n_{exp} as the number of experimental data points, $\Delta x_l = \left| x_l^{calc} - x_l^{exp} \right|$ the absolute residuals in the mole fraction of the liquid phase resulting from isobaric-isothermal flash calculations and $\Delta P_l^{rel} = \left| \frac{P_l^{calc}}{P_l^{exp}} - 1 \right|$ the relative residuals in pressure resulting from bubble point calculations.

3.3.4. Quantitative Structure Property Relationship (QSPR) for predicting k_{ij}

Multivariate regression

The QSPR model for predicting k_{ij} was built in steps: Initially, the QSPR model was developed for the simplest case; for predicting k_{ij} in mixtures of two non-associating, non-polar components (group 1), using only the descriptor for the asymmetry in the dispersive intermolecular potential D^{LJ}. Additional groups of binary mixtures were considered progressively and the model was supplemented with the necessary descriptors. In this way, the model was extended to all groups of mixtures with non-associating components (groups 1 to 6). Subsequently, the mixtures of associating components (groups 7 to 10) were considered. The QSPR

model for all groups of binary mixtures reads

$$k_{ij}^{\text{QSPR}} = \sum_{L=1}^{N_d} c_L \cdot D_L \tag{3.19}$$

with model coefficients

$$c_L \in \{c^{\text{LJ}}, c^{\text{dd,a}}, c^{\text{dd,b}}, c^{\text{qq}}, c^{\text{dq}}, c^{\text{assoc,s}}, c^{\text{assoc,c}}, c^{\text{assoc,sc}}\}$$

and with corresponding descriptors D_L

$$D_L \in \{D^{\text{LJ}}, D^{\text{dd,a}}, D^{\text{dd,b}}, D^{\text{qq}}, D^{\text{dq}}, D^{\text{assoc,s}}, D^{\text{assoc,c}}, D^{\text{assoc,sc}}\}$$

as introduced in Eq.(3.8) to (3.17). A non-weighted non-linear least squares problem is solved for the regression of the model coefficients c_L. The objective function for the model regression is defined as the total sum of the squared residuals between the values of k_{ij} individually adjusted to experimental data k_{ij}^{fit} and the values of k_{ij} calculated from the QSPR model k_{ij}^{QSPR} in Eq.(3.19) as

$$\min_{\underline{c}} Q = \sum_{m=1}^{N_{\text{tr}}} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}}\right)^2 \tag{3.20}$$

with N_{tr} the number of mixtures used for the regression. The least squares minimization was conducted using the solver *nlinfit* provided in Matlab [69]. For the 95% confidence interval of the model coefficients $\Delta c_{L,95}$ we use the *t*-distribution with $N_{\text{tr}} - N_d - 1$ degrees of freedom [70]

$$\Delta c_{L,95} = t_{0.95,N_{\text{tr}}-N_d-1} \cdot \sqrt{S^2} \cdot \sqrt{V_{LL}} \tag{3.21}$$

with S^2 as the estimated model variance

$$S^2 = \frac{\sum_{m=1}^{N_{\text{tr}}} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}}\right)^2}{N_{\text{tr}} - N_d - 1} \tag{3.22}$$

and with V_{LL} as the corresponding diagonal element of the coefficient covariance matrix.

Training and test set

The quality and the morphology of the data sets used to derive and validate the QSPR model coefficients (QSPR database) is decisive. The set used to derive the QSPR model coefficients (training set) should span the whole region of the descriptor space, it should be diverse and it should include data points close to the data points used for the external validation of the model (test set) [71, 72]. In order to ensure these conditions, we define the training and test set respectively in a three-step procedure: In step 1, we *a priori* detect and remove from the QSPR database

the binary mixtures with unreliable k_{ij} values [70, 73]. In step 2, we divide the QSPR database in training and test set, using a uniform design method for subset selection [74]. In step 3, we iteratively remove the outliers from the training set. The steps are discussed in more detail in the following.

In step 1 we use as measure for the quality of the adjusted k_{ij} values the value of the objective function F in Eq. (3.18), divided by the total number of the experimental data points, according to

$$f = F/n_{\text{exp}} \quad . \tag{3.23}$$

The value of k_{ij} for a mixture m is considered as unreliable if $f_m > 0.95 \cdot \max(\overline{f})$ and the absolute average residuals AAD-x and AAD-P^{rel} are higher than 15%. Those cases either correspond to binary mixtures with experimental data of poor quality or to mixtures, where PC-SAFT leads to inaccurate results with the chosen pure component parameters.

After unreliable data were removed, the curated database (N_m binary mixtures) is divided into the training and the test set in step 2. In our case, we define the split fraction between the training and the test set $N_{\text{tr}}/N_{\text{ts}}$ equal to 9 (with $N_{\text{tr}}+N_{\text{ts}} = N_m$). For the selection of the training set, we have used an in-house implementation of the Kennard and Stone algorithm. The Kennard and Stone (KS) algorithm [75] is a well-established method for uniform design in the field of chemometrics [74]. The selection principle of the algorithm [75] ensures that the points which are excluded (test set) are close to the points of the design set (training set). Here, we implemented the KS algorithm individually for each one of the 10 binary mixture groups ('clusters') defined in Section 3.3. For all groups we used the same fraction of training to test set, as for the complete database ($N_{tr,group}/N_{ts,group} = 9$). Implementing the KS algorithm individually for each group is a way to ensure that all groups are represented in the training set, even if the number of available data per group is limited [76] (e.g. binary mixtures with two non-associating, quadrupolar components). Since the KS algorithm uses the Euclidean norm of the descriptor vector, the results are sensitive to descriptor scaling [77]. For better scaling of the highly irregular multidimensional descriptor space, we used the standardized values of the descriptors [78]

$$\hat{D}_{L,m} = \frac{D_{L,m} - \overline{D}_L}{S_{DL}} \tag{3.24}$$

with $\hat{D}_{L,m}$ the standardized value of descriptor D_L for mixture m, $\overline{D}_L$ denotes the average value of descriptor D_L for all mixtures, and S_{DL} is the standard deviation of descriptor D_L for all mixtures of the database. In our implementation, the KS algorithm is initialized on the boundaries of the multidimensional descriptor space. Thus, the designed training set is expected to be sensitive to outliers [76].

We exclude outliers from the training set iteratively in step 3. We fit the model coefficients N_m times, excluding one mixture β at a time ("Leave-One-Out" proce-

dure), and we calculate the corresponding sum of the residuals $r_{N_m-1,\beta}$

$$r_{N_m-1,\beta} = \sum_{\substack{m=1 \\ m\neq\beta}}^{N_m-1} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}}\right)^2 \tag{3.25}$$

with $\beta \in \{1, 2, \ldots, N_m\}$. We determine the mean value

$$\bar{r}_{N_m-1} = \frac{1}{N_m} \cdot \sum_{\beta}^{N_m} r_{N_m-1,\beta} \tag{3.26}$$

and the standardized value of the residual for each mixture β around the mean value $\bar{r}_{N_m-1}$ as

$$\hat{r}_{N_m-1,\beta} = \frac{r_{N_m-1,\beta} - \bar{r}_{N_m-1}}{S_r} \tag{3.27}$$

where S_r is the standard deviation of the residuals. Mixtures that led to $\hat{r}_{N_m-1,\beta}$ values greater than three standard deviation units $(3S_r)$ were considered as outliers and were excluded. [70, 71, 79]

Model validation

Model validation is an integral part of any QSPR method. On the one hand, the goodness-of-fit, the robustness of the model and its internal predictive power are measured by internal validation techniques. On the other hand, the actual predictive power of the model is evaluated by external validation. External validation is defined through data that was not used for the model regression [72].

The goodness-of-fit is measured by the coefficient of multiple determination R^2. The coefficient of multiple determination estimates the portion of the variation in the independent variable (here k_{ij}^{fit}) that is explained by the regression [72, 73, 79]. R^2 is calculated over all mixtures of the training set (N_{tr}) and is defined as

$$R^2 = 1 - \frac{\sum_{m=1}^{N_{\text{tr}}} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}}\right)^2}{\sum_{m=1}^{N_{\text{tr}}} \left(k_{ij,m}^{\text{fit}} - \bar{k}_{ij}^{\text{fit}}\right)^2} \tag{3.28}$$

where $\bar{k}_{ij}^{\text{fit}}$ is the average value of k_{ij}^{fit} for all mixtures of the training set. A QSPR model with R^2 higher than 0.6 can be considered predictive [73]. However, Tropsha [73] also points out, that the predictive power of a QSPR model needs to be further evaluated for compounds that were not included in the training set.

The Leave-One-Out (LOO) cross-validated correlation coefficient Q_{LOO}^2 and the Leave-Many-Out (LMO) cross-validated correlation coefficient Q_{LMO}^2 measure the robustness of the model and its internal predictive power. The cross-validated correlation coefficients should be calculated over a large number of trials. In both the

LOO and LMO procedures, a certain subset of the training set is omitted at each trial.

For the calculation of the LOO cross-validated correlation coefficient Q^2_{LOO} the number of trials is equal to the total number of mixtures in the QSPR database N_m. At each trial a mixture is individually omitted. The QSPR model coefficients are regressed using the remaining $N_m - 1$ mixtures as training set. Using this regression, one can now predict the k_{ij} value ($k_{ij}^{QSPR/-m}$) for mixture m (which was temporarily excluded from the training set). The definition of Q^2_{LOO} according to the guidlines of the Organization of Economic Co-operation and Development (OECD) [72] is

$$Q^2_{\text{LOO}} = 1 - \frac{\sum_{m=1}^{N_m}\left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{QSPR/-m}\right)^2}{\sum_{m=1}^{N_m}\left(k_{ij,m}^{\text{fit}} - \overline{k}_{ij}^{\text{fit}}\right)^2} \quad . \tag{3.29}$$

At each trial of the LMO procedure, more mixtures than one are simultaneously excluded. Here, we have calculated the value of Q^2_{LMO} over $2N_m$ number of trials. At each trial, 50% of the complete database was excluded [71] ($N_{\text{LMO}} = 0.5N_m$). The value of the LMO cross-validated correlation coefficient for each trial $Q^2_{LMO/50,trial}$ is calculated as

$$Q^2_{LMO/50,trial} = 1 - \frac{\sum_{m=1}^{N_{\text{LMO}}}\left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{QSPR/-m}\right)^2}{\sum_{m=1}^{N_m}\left(k_{ij,m}^{\text{fit}} - \overline{k}_{ij}^{\text{fit}}\right)^2} \tag{3.30}$$

with $k_{ij,m}^{QSPR/-m}$ representing the predicted k_{ij} value for each mixture m that is part of the excluded set. The standard deviation and the average value of $Q^2_{LMO/50,trial}$ over all trials are used as indicators of the robustness in the model prediction. A robust model should remain invariant to changes of the training set and it is thus expected to exhibit small difference between the average value of Q^2_{LMO} and R^2. [80] We should note that the LMO procedure demands to generate as many different training sets as the number of trials. The Kennard and Stone algorithm described in Section 3.4.2 is initialized on a fixed point of the descriptor space and therefore provides a single design for the training set. Thus, for the calculation of Q^2_{LMO} , we used a straightforward random selection of the training set.

An important indicator of the actual predictive capability of the model is the external explained variance Q^2_{ext}. The external explained variance is a measure for the quality of prediction for data that were not included in the set used for the model development. There are several approaches in the literature for the calculation of Q^2_{ext} [71, 72, 81, 82]. In our implementation, the training set is much larger than the test set. In this case, the definition of Consonni et al. [81] is appropriate, with

$$Q^2_{ext} = 1 - \frac{\left[\sum_{m=1}^{N_{\text{ts}}}\left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}}\right)^2\right]/N_{\text{ts}}}{\left[\sum_{m=1}^{N_{\text{tr}}}\left(k_{ij,m}^{\text{fit}} - \overline{k}_{ij}^{\text{fit}}\right)^2\right]/N_{\text{tr}}} \quad . \tag{3.31}$$

The numerator in Eq.(3.31) sums over data of the test set, whereas the denominator sums over the training set.

3.4. Results

3.4.1. Estimation of k_{ij} based on London's dispersive theory for mixtures

Calculating k_{ij} values from London's dispersive theory, according to Eq.(3.5), requires segment diameters σ_i and σ_j and the ionization potentials I_i and I_j. Experimental values of the ionization potentials were taken from the literature [83]. We limit consideration to mixtures of two non-polar, non-associating components (group 1). Figure 3.1 compares the values of k_{ij} calculated with Eq.(3.5) to the values of k_{ij} individually adjusted on experimental data. A detailed list of the examined binary mixtures with the corresponding values of $k_{ij}(\sigma, I)$ and k_{ij}^{fit} is given in Table E.1 in Appendix E.

k_{ij} calculated with Eq.(3.5) cannot take on negative values. For mixtures of non-polar, non-associating components this limitation is not strongly restricting the results, because the negative k_{ij}^{fit} values are all fairly close to zero. Figure 3.2 shows that for the mixtures of group 1 with negative values of k_{ij}^{fit} (mixtures # 1,9,15,19,26,30,34,44,45) the results of phase equilibrium calculations remain good for calculations with the slightly positive k_{ij} values from Eq.(3.5). Mixtures with polar and associating components, though, often demand more negative k_{ij} values. In this case, the predictions of k_{ij} with Eq.(3.5) are expected to lead to more significant errors.

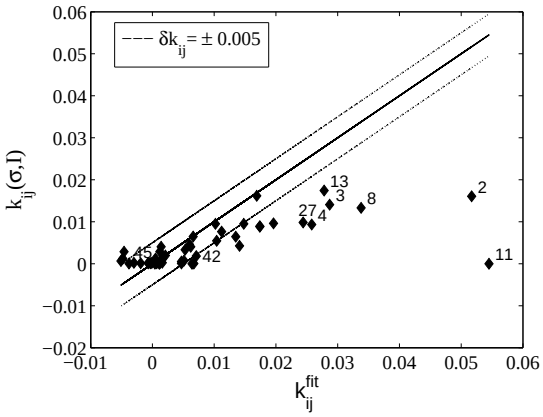


Figure 3.1: Comparison of $k_{ij}(\sigma, I)$ values estimated from the London's dispersive theory (Eq. (3.5)) to k_{ij}^{fit} values individually adjusted to experimental data for mixtures of two non-polar, non-associating components (group 1). The dashed lines illustrate the absolute deviations of ± 0.005 for the value of k_{ij}^{fit}.

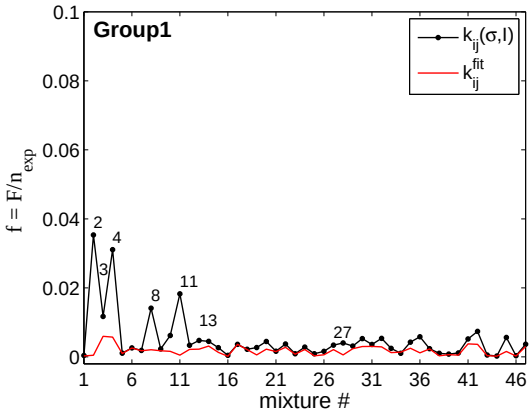


Figure 3.2: Deviations of PC-SAFT EoS from experimental binary phase equilibrium data of two non-associating, non-polar components (group 1). The deviations are defined as $f = F/n_{\text{exp}}$ according to Eq.(3.23). The red line serves as a reference and is obtained for individually optimized k_{ij}^{fit} values. The symbols (connected by black line) are obtained using estimated values $k_{ij}(\sigma, I)$. For hexadecane no ionization potential σ was available, which is why $k_{ij}(\sigma, I)$ was set to zero for mixtures #12 and #43.

Higher deviations in the prediction of k_{ij} are observed for binary mixtures including carbon monoxide (mixtures 2, 3 and 4 in Figures 3.1 and 3.2) and for binary mixtures including methane (mixtures 8, 13 and 27). These mixtures demand higher values of k_{ij} for an adequate description. We conjecture that in those cases, k_{ij} does not only correct the dispersive intermolecular potential of the mixture but compensates other model deficiencies. Below we show that a QSPR model based on semi-empirical descriptors takes the necessary contributions into account to model the value of k_{ij} for these mixtures.

3.4.2. QSPR estimation of k_{ij} - Model regression and assessment of the results

The descriptors of the QSPR model for the estimation of k_{ij} values were constructed using the PC-SAFT molecular parameters. The initial structure of the descriptors was defined in an ad hoc manner and their exact structure was adjusted empirically, by varying exponents, as described in Section 3.1. In the course of this process, different sets of descriptors were generated. For each set of descriptors, the coefficients of the QSPR model were regressed as described in Section 3.4.1. The performance of each QSPR model was evaluated using the internal and external validation techniques described in Section 3.4.3. The training and the test set remained unchanged. The QSPR model that achieved the best overall description for all groups of mixtures based on the value of R^2 has been obtained with the descriptors defined in Eq.(3.8) to (3.17). Here, we limit our discussion to the results of this model.

Mixtures with associating components are the most difficult to describe. For the

regression of the model coefficients, we therefore distinguish two cases: In the first case, we estimate the model coefficients only for mixtures of non-associating components ("1st regression case": groups 1 to 6). In the second case, we adjust the model only to mixtures containing associating components ("2nd regression case": groups 7 to 10).

In the first case, the database of mixtures with only non-associating components includes 199 binary mixtures. Unreliable data and outliers have been removed according to the rationale described in Section 3.4.2. Then, the training set consists of 151 and the test set of 22 binary mixtures. The adjusted model coefficients c_L along with the values of their 95% confidence intervals $\Delta c_{L,95}$ are given in Table 3.2. For all regressed model coefficients we observed $|\Delta c_{L,95}| << |c_L|$, which qualifies a stable regression model [84]. For mixtures of groups 1 to 6 the values of k_{ij} predicted with the QSPR model are given in Appendix E (Tables E.3 to E.8). In

Table 3.2: Results of the regressed model coefficients for mixtures of non-associating components (groups 1 to 6). The values of the coefficients c_L and their 95% confidence interval $\Delta c_{L,95}$ for the QSPR model defined in Eq.(3.19) and (3.21), with $c^{\text{assoc,s}} = 0$, $c^{\text{assoc,c}} = 0$ and $c^{\text{assoc,sc}} = 0$.

	mixtures with non-associating components	
	$c_L \cdot 10^4$	$\Delta c_{L,95} \cdot 10^4$
c^{LJ}	-103.285	± 0.065
$c^{\text{dd,a}}$	57.533	± 0.235
$c^{\text{dd,b}}$	-25.064	± 0.095
c^{qq}	0.114	± 0.001
c^{dq}	-0.645	± 0.002

the second case, for associating mixtures the database includes 253 mixtures with at least one associating component. After unreliable data and outliers have been removed, the training set consists of 167 and the test set of 23 binary mixtures. The adjusted model coefficients c_L for mixtures with at least one associating component and their 95% confidence intervals $\Delta c_{L,95}$ are given in Table 3.3. For this "2nd regression case" we also find $|\Delta c_{L,95}| << |c_L|$. The regression model for mixtures of the groups 7 to 10 can therefore be characterized stable as well. For mixtures of groups 7 to 10 the predicted values of k_{ij} are given in Appendix E (Tables E.9 to E.12). The multiple correlation coefficient R^2 for the QSPR model adjusted to mixtures of non-associating components is 88.4%. For the QSPR model adjusted to mixtures of associating components R^2 is 70.1%. In both cases, the model can, in the jargon of QSPR methods, be considered to be predictive. Figure 3.3 shows the values of the multiple correlation coefficient (R^2_{group}) calculated individually for each group of binary mixtures. Values of R^2_{group} have been calculated using Eq.(3.20), with the summation running over all mixtures of the training set for each particular group. The results of the internal and external validation of the QSPR model for both regression cases are summarized in Table 3.4.

Table 3.3: Results of the regressed model coefficients for mixtures with at least one associating component (groups 7 to 10). The values of the coefficients c_L and their 95% confidence interval $\Delta c_{L,95}$ for the QSPR model defined in Eq.(3.19) and (3.21).

	mixtures with associating components	
	$c_L \cdot 10^4$	$\Delta c_{L,95} \cdot 10^4$
c^{LJ}	-149.236	± 0.591
$c^{\mathrm{dd,a}}$	20.178	± 0.841
$c^{\mathrm{dd,b}}$	-23.180	± 0.210
c^{qq}	-0.428	± 0.006
c^{dq}	0.089	± 0.006
$c^{\mathrm{assoc,s}}$	-12.432	± 0.062
$c^{\mathrm{assoc,c}}$	-4.797	± 0.018
$c^{\mathrm{assoc,sc}}$	0.592	± 0.006

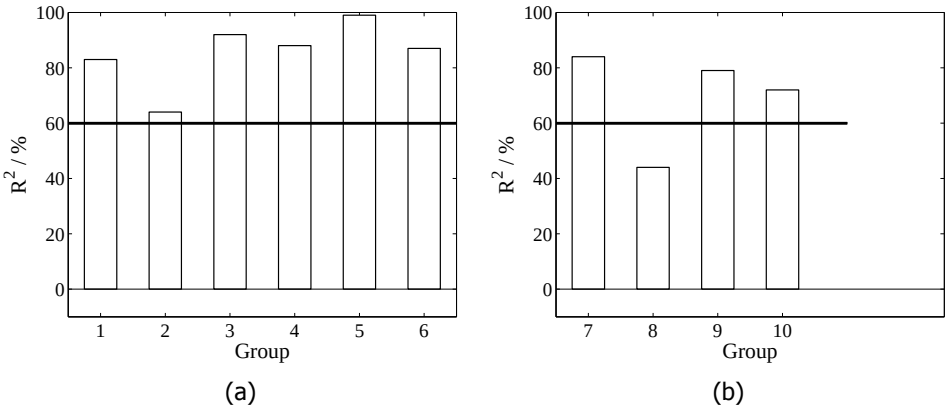


Figure 3.3: Coefficient of multiple determination R^2 for: (a) mixtures of two non-associating components (groups 1 to 6) and (b) mixtures including at least one associating component (groups 7 to 10).

The QSPR model adjusted to mixtures of non-associating components describes all types of non-associating mixtures reasonably well. The results are satisfactory, especially considering the diversity of mixtures in our database. The value of the LOO cross-validated coefficient Q^2_{LOO} and the average value of the LMO cross-validated coefficient $\hat{Q}^2_{\mathrm{LMO}/50}$ (Table 3.4) are close to the value of R^2. The proposed model for mixtures with non-associating, polar or non-polar components can be thus characterized as sufficiently robust [80]. Further, the low standard-deviation of $Q^2_{\mathrm{LMO}/50}$ (5.4%) over 346 trials indicates that the model is robust against strong variations of the training set.

For the model adjusted to mixtures of associating components ("2nd regression

Table 3.4: Results of internal and external validation of the QSPR model (Eq.(3.19)) developed for mixtures of non-associating and mixtures of associating components.

mixtures with non-associating components			
$R^2(\%)$	$Q^2_{\text{LOO}}(\%)$	$\hat{Q}^2_{\text{LMO}/50}(\%)$	$Q^2_{ext}(\%)$
88.4	86.5	84.5	92.4
mixtures with associating components			
$R^2(\%)$	$Q^2_{\text{LOO}}(\%)$	$\hat{Q}^2_{\text{LMO}/50}(\%)$	$Q^2_{ext}(\%)$
70.1	64.3	57.9	82.3

case") the coefficients of multiple determination R^2_{group} (explained variance) for the individual groups are also high, particularly for the groups 7, 9 and 10. Group 8 contains mixtures of species with widely differing pure component parameters which we call asymmetry. To properly and transferably account for this asymmetry in the resulting k_{ij} values was only partially successful in our QSPR approach. We note that mixtures of group 8 represent 34% of the training set used for the adjustment of the QSPR model coefficients in the "2nd regression case". When mixtures of group 8 are omitted during the Leave-One-Out (LOO) or Leave-Many-Out (LMO) validation tests, the QSPR model demonstrates a much better overall performance in terms of the Q^2_{LOO} and $\hat{Q}^2_{\text{LMO}/50}$ QSPR validation measures. The model adjusted to mixtures of associating components is more uncertain than the QSPR model for mixtures of non-associating components. The choice of the training set acts on the result. This is reflected directly in a) the high standard deviation of the $Q^2_{\text{LMO}/50}$ coefficient over 380 trials (22.6%) and b) the higher difference between the value of R^2 and the values of the cross validated coefficients $\hat{Q}^2_{\text{LMO}/50}$ and Q^2_{LOO} (Table 3.4).

The external explained variance Q^2_{ext} by the QSPR model adjusted to mixtures of non-associating components is 92%. For the QSPR model adjusted to mixtures of associating components Q^2_{ext} is 82%. The high values of Q^2_{ext} indicate the high predictive power of the QSPR model in both cases. Nevertheless, the results of Q^2_{ext} should be interpreted cautiously. One should bear in mind that the size of the test set is much smaller and less diverse than the size of the training set ($N_{\text{tr}}/N_{\text{ts}} = 9$).

Generally, the proposed QSPR model is stable and robust and has (in the sense of a QSPR method) a good predictive capacity. The statistical measures (Table 3.4) indicate that the QSPR model predicts k_{ij} values in binary mixtures of non-associating components with sufficient accuracy. For mixtures of associating components, the model is also reliable (in the jargon of QSPR). However, the estimations of k_{ij} for mixtures containing one dipolar, non-associating component and one associating component are more uncertain and should be used with reservation.

3.4.3. Evaluating phase equilibria with predicted k_{ij} values

The k_{ij} values used for the regression of the QSPR model (k_{ij}^{fit}) are not experimental data, but parameters fitted to experimental data. We therefore find it important to evaluate the QSPR model beyond just the statistical measures of the QSPR method. We have examined the results of phase equilibrium calculations with the predicted k_{ij} values for each group of mixtures. The parity plots and the diagrams that show the results in phase equilibrium calculations for all groups of mixtures are given in Appendix E (Figures E.1 to E.4).

Here, we discuss in more detail the results for mixtures of group 4 (mixtures of one non-associating, quadrupolar and one non-associating, dipolar component). We find the performance for mixtures of group 4 representative of the overall performance of the QSPR predictions of k_{ij} for non-associating mixtures ($R^2_{\text{group 4}} = 88.1\% \simeq R^2$). Moreover, the prediction for mixtures of group 4 is of practical interest, for example by the application of CAMD for polar solvents, for the CO_2 capture with physical absorption [18].

Figure 3.4 compares the estimated values, k_{ij}^{QSPR}, to the reference values, k_{ij}^{fit}, for group 4. Further, Figure 3.5 illustrates the accuracy in phase equilibrium calculations with PC-SAFT for the mixtures of group 4 with k_{ij} values estimated with the QSPR model adjusted to non-associating mixtures (Table 3.2). The results in phase equilibrium calculations are compared to experimental VLE data. We consider the same experimental data that were used for the individual fitting of k_{ij} values. We assess the results using the average value f of the combined residuals in the liquid mole fraction and the pressure, over the number of experimental data points for each mixture as defined in Eq.(3.23). In Figure 3.5, we also present the results in phase equilibrium calculations, when no correction is used for the PC-SAFT EoS ($k_{ij} = 0$), as well as the results achieved when k_{ij} is individually adjusted to the experimental data, namely $f(k_{ij}^{\text{fit}})$. We observe that for the majority of the mixtures in group 4 (in both the training and the test set) phase equilibrium calculations with the QSPR estimations of k_{ij} give lower residuals than if no correction is used ($k_{ij} = 0$).

It is however difficult to quantify the relationship between the error in k_{ij} prediction $\Delta k_{ij} = |k_{ij}^{\text{fit}} - k_{ij}^{\text{QSPR}}|$ and the subsequent error in phase equilibrium calculations $\Delta f = |f(k_{ij}^{\text{fit}}) - f(k_{ij}^{\text{QSPR}})|$. For different mixtures, the sensitivity of the objective function F (and thereby f) to the value of k_{ij} is different, depending on different factors, like azeotropy or supercritical regions. Let us, for example, examine the performance of the k_{ij} prediction and phase equilibrium calculations for mixture #4 (carbon dioxide - 1-bromobenzene) compared to mixture #20 (1-hexene - ethylene). For mixture #4, the error in the estimation of k_{ij} is much higher than the error in the estimation of k_{ij} for mixture #20 (Figure 3.4, Table S5). Still, the error in phase equilibrium calculations is almost identical for both mixtures (Figure 3.5).

Finally, the k_{ij} values estimated from London's dispersive theory $k_{ij}(\sigma, I)$ (Eq.(3.5)) were used in phase equilibrium calculations for the mixtures of group 4. A list of the estimated values $k_{ij}(\sigma, I)$ is given in Table E.2 in Appendix E. In Figure 3.6, phase equilibrium calculations carried out with the estimated values k_{ij}^{QSPR} are com-

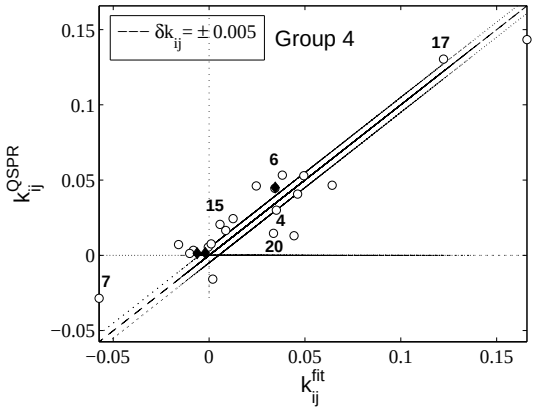


Figure 3.4: Comparison of the estimated k_{ij}^{QSPR} values with k_{ij}^{fit} values individually adjusted to experimental data for mixtures of one non-associating, dipolar and one non-associating, quadrupolar component (group 4). Open symbols indicate mixtures of the training set, whereas solid symbols correspond to the mixtures of the test set. The dashed lines illustrate the absolute deviations of ± 0.005 for the value of (k_{ij}^{fit}).

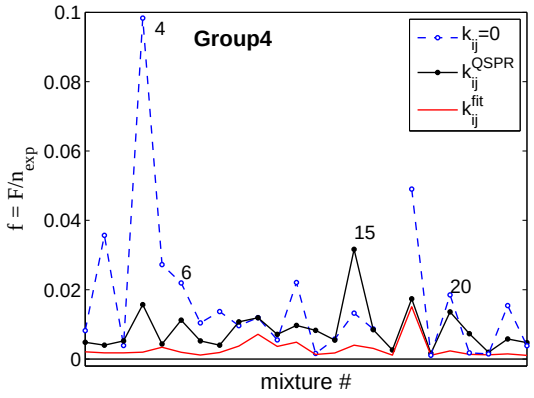


Figure 3.5: Deviations of PC-SAFT from experimental binary phase equilibrium data of one non-associating, dipolar and one non-associating, quadrupolar component (group 4). The deviations are defined as $f = F/n_{\text{exp}}$ according to Eq.(3.23). The red line serves as a reference and is obtained for individually optimized k_{ij}^{fit} values. The symbols (connected by black line) are obtained using estimated values k_{ij}^{QSPR}. The blue dashed line represent phase equilibrium calculations, when the PC-SAFT EoS is not corrected ($k_{ij} = 0$).

pared to phase equilibrium calculations with the estimated values from London's dispersive theory $k_{ij}(\sigma, I)$. The results obtained with the estimated value k_{ij}^{QSPR} are

in most cases better than the results obtained with $k_{ij}(\sigma, I)$. For example, mixture #7 (dimethyl ether-carbon dioxide) requires a negative k_{ij} value. Negative k_{ij} values cannot be calculated using Eq.(3.5). As discussed in Section 4.1, this limitation has a higher impact on mixtures containing polar components. For mixture #7, the estimated k_{ij}^{QSPR} is negative. With k_{ij}^{QSPR}, the PC-SAFT EoS is corrected and phase equilibrium calculations are considerably better. Moreover, for mixtures that require high positive k_{ij} values, the predicted value k_{ij}^{QSPR} leads to significantly lower residuals in the phase equilibrium calculations. Representative examples are mixture #2 (nitrogen-propylbenzene) and mixture #17 (nitrogen-toluene). The

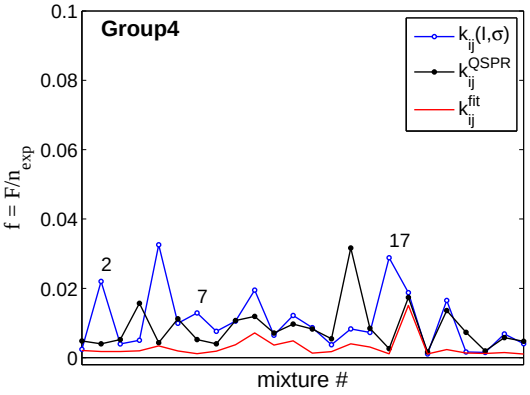
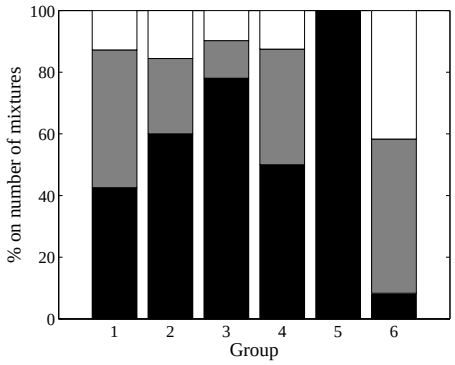
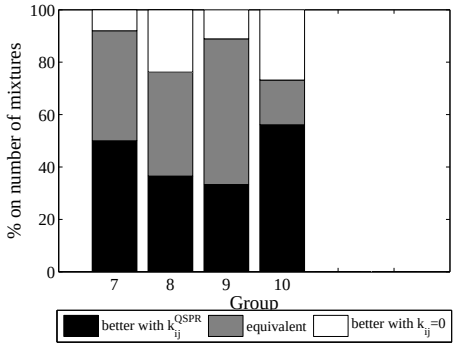


Figure 3.6: Deviations of PC-SAFT from experimental binary phase equilibrium data of one non-associating, dipolar and one non-associating, quadrupolar component (group 4). The deviations are defined as $f = F/n_{\text{exp}}$ according to Eq.(3.23). The red line serves as a reference and is obtained for individually optimized k_{ij}^{fit} values. The black line represents phase equilibrium calculations, when the PC-SAFT EoS is corrected with k_{ij}^{QSPR}. The blue solid line is for results using $k_{ij}(\sigma, I)$ from London's dispersive theory.

k_{ij} values predicted by the proposed QSPR model (with the descriptors defined in Eq.(3.8) to (3.17) and the regressed coefficients given in Tables 3.2 and 3.3) are used in phase equilibrium calculations for all mixtures considered in our QSPR database. In Figure 3.7, a general overview is given of the improvement potential in phase equilibrium calculations, using the k_{ij} values estimated with the proposed QSPR model. For each group of mixtures, Figure 3.7 displays the percentage of cases in which phase equilibrium calculations are more accurate with the estimated value k_{ij}^{QSPR}. The percentage of cases when phase equilibrium calculations with $k_{ij} = 0$ are more accurate and the percentage of cases when phase equilibrium calculations are equally good for k_{ij}^{QSPR} and $k_{ij} = 0$ are given as well. For all groups of mixtures (associating and non-associating) the proposed QSPR model provides k_{ij} estimates that generally improve the accuracy of PC-SAFT in phase equilibrium calculations. According to Figure 3.7, the most significant improvement is achieved for mixtures of group 5. However, the small size of group 5 (Table 3.1) and its'



(a)



(b)

Figure 3.7: Assessing phase equilibrium calculations: Black surface represents the percentage of cases for which calculations with estimated k_{ij}^{QSPR} values are more accurate than with $k_{ij} = 0$. The gray shaded area represents the percentage of cases when both cases are equivalent.

limited diversity (Table E.7) do not allow us to draw a generalized conclusion about the performance of the method for mixtures of two non-associating, quadrupolar components. Further, in the group with the least improvement, for mixtures of two dipolar, non-associating components (group 6), the individually adjusted k_{ij} values are close to zero (see Appendix E). Calculation results for $k_{ij} = 0$ are thus close to optimal, so that the use of estimated k_{ij}^{QSPR} cannot lead to significant improvements for group 6.

3.5. Conclusions

We developed and analyzed a multivariate regression model for the estimation of the binary interaction parameter k_{ij} of the PC-SAFT EoS using a QSPR method. The descriptors of the regression model are based on the pure component parameters of

the PC-SAFT model. The regression models were developed separately for mixtures with non-associating components and for mixtures with at least one associating component.

For mixtures with non-associating components, values of k_{ij} were also estimated from London's dispersive theory, using the expression of Hudson and McCoubrey [50]. Values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}) and values of k_{ij} estimated from London's dispersive theory ($k_{ij}(\sigma, I)$) were compared to values of k_{ij} individually fitted on experimental data. For mixtures containing non-associating, polar components the QSPR regression model allows for negative values of k_{ij} and leads to more accurate predictions than the expression of Hudson and McCoubrey. The QSPR regression model considers additional contributions to the value of k_{ij} than just the asymmetry in the dispersive intermolecular forces.

The coefficient of multiple determination R^2 of the QSPR regression model for mixtures with non-associating components is 88.4%. For mixtures with at least one associating component R^2 is 70.1%. Thus, for both regression cases the QSPR model can, in the jargon of QSPR methods, be characterized as stable and robust and with good predictive capacity.

The k_{ij} values estimated as function of the pure component PC-SAFT parameters, with the proposed QSPR regression model, can be used to enhance the accuracy of phase equilibrium calculations with PC-SAFT. With the proposed QSPR model for the estimation of k_{ij}, no prior knowledge of the real behavior of the examined binary mixture is demanded. The QSPR approach can be implemented in a CAMD framework that uses the pure component parameters of the optimized fluid directly as optimization variables. In that case, we expect that phase equilibrium calculations and thus the accuracy of CAMD results can be enhanced.

References

[1] E. A. Muller and K. E. Gubbins, *Molecular-based equations of state for associating fluids: A review of SAFT and related approaches,* Ind. Eng. Chem. Res. **40**, 2193 (2001).

[2] I. G. Economou, *Statistical Associating Fluid Theory: A successful model for the calculation of thermodynamic and phase equilibrium properties of complex fluid mixtures,* Ind. Eng. Chem. Res. **41**, 953 (2002).

[3] S. P. Tan, H. Adidharma, and M. Radosz, *Recent advances and applications of statistical associating fluid theory,* Ind. Eng. Chem. Res. **47**, 8063 (2008).

[4] C. McCabe and A. Galindo, *Chapter 8 SAFT associating fluids and fluid mixtures,* in *Applied Thermodynamics of Fluids* (The Royal Society of Chemistry, 2010) pp. 215–279.

[5] W. G. Chapman, S. G. Sauer, D. Ting, and A. Ghosh, *Phase behavior applications of SAFT based equations of state: from associating fluids to polydisperse, polar copolymers,* Fluid Phase Equilib. **217**, 137 (2004).

[6] G. Sadowski, *Thermodynamics of polymer systems,* Macromolecular Symposia **206**, 333 (2004).

[7] F. Tumakaka, J. Gross, and G. Sadowski, *Thermodynamic modeling of complex systems using PC-SAFT,* Fluid Phase Equilib. **228-229**, 89 (2005).

[8] B. Liebergesell, S. Kaminski, C. Pauls, T. W. de Loos, T. Vlugt, K. Leonhard, and A. Bardow, *High-pressure vapor-liquid equilibria of the second generation biofuel blends (2-methylfuran+iso-octane) and (2-methyltetrahydrofuran+di-n-butyl ether): Experiments and PCP-SAFT modeling,* Fluid Phase Equilib. **400**, 95 (2015).

[9] R. Gani and J. P. O'Connell, *Properties and CAPE: from present uses to future challenges,* Comput. Chem. Eng. **25**, 3 (2001).

[10] J. P. O'Connell, R. Gani, P. M. Mathias, G. Maurer, J. D. Olson, and P. A. Crafts, *Thermodynamic property modeling for chemical process and product engineering: Some perspectives,* Ind. Eng. Chem. Res. **48**, 4619 (2009).

[11] R. Gani, B. Nielsen, and A. Fredenslund, *A group contribution approach to computer-aided molecular design,* AIChE J. **37**, 1318 (1991).

[12] L. Y. Ng, F. K. Chong, and N. G. Chemmangattuvalappil, *Challenges and opportunities in Computer Aided Molecular Design,* Comput. Chem. Eng. **81**, 115 (2015).

[13] F. E. Pereira, E. Keskes, A. Galindo, G. Jackson, and C. S. Adjiman, *Integrated design of* CO_2 *capture processes from natural gas,* in *Process Systems Engineering* (Wiley-VCH Verlag GmbH & Co. KGaA, 2008) pp. 231–248.

[14] A. Bardow, K. Steur, and J. Gross, *Continuous-Molecular Targeting for integrated solvent and process design,* Ind. Eng. Chem. Res. **49**, 2834 (2010).

[15] F. Pereira, E. Keskes, A. Galindo, G. Jackson, and C. Adjiman, *Integrated solvent and process design using a SAFT-VR thermodynamic description: High-pressure separation of carbon dioxide and methane,* Comput. Chem. Eng. **35**, 474 (2011).

[16] M. Lampe, M. Stavrou, H. M. Bucker, J. Gross, and A. Bardow, *Simultaneous optimization of working fluid and process for organic rankine cycles using PC-SAFT,* Ind. Eng. Chem. Res. **53**, 8821 (2014).

[17] A. Papadopoulos, S. Bard, A. Chremos, E. Forte, T. Zarogiannis, P. Seferlis, S. Papadokonstantakis, C. Adjiman, A. Galindo, and G. Jackson, *Efficient screening and selection of post-combustion* CO_2 *capture solvents,* Chem. Eng. Trans. **39**, 211 (2014).

[18] M. Stavrou, M. Lampe, A. Bardow, and J. Gross, *Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for* CO_2 *capture,* Ind. Eng. Chem. Res. **53**, 18029 (2014).

[19] M. Lampe, M. Stavrou, J. Schilling, E. Sauer, J. Gross, and A. Bardow, *Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT,* Comput. Chem. Eng. **81**, 278 (2015).

[20] C. S. Adjiman, A. Galindo, and G. Jackson, *Molecules matter: The expanding envelope of process design,* Comput. Chem. Eng. **34**, 55 (2014).

[21] R. Gani, *Chapter 14 Case studies in chemical product design use of CAMD techniques,* in *Chemical Product Design: Toward a Perspective Through Case Studies*, Comput. Aided Chem. Eng., Vol. 23, edited by R. G. Ka M. Ng and K. Dam-Johansen (Elsevier, 2007) pp. 435 – 458.

[22] A. I. Papadopoulos and P. Linke, *Integrated solvent and process selection for separation and reactive separation systems,* Chem. Eng. Process. **48**, 1047 (2009).

[23] M. Harini, J. Adhikari, and K. Y. Rani, *A review on property estimation methods and computational schemes for rational solvent design: A focus on pharmaceuticals,* Ind. Eng. Chem. Res. **52**, 6869 (2013).

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

[25] J. Burger, V. Papaioannou, S. Gopinath, G. Jackson, A. Galindo, and C. S. Adjiman, *A hierarchical method to integrated solvent and process design of physical CO_2 absorption using the SAFT-γ Mie approach,* AIChE J. **61**, 3249 (2015).

[26] V. Papaioannou, T. Lafitte, C. Avendano, C. S. Adjiman, G. Jackson, E. A. Muller, and A. Galindo, *Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments,* J. Chem. Phys. **140**, (2014).

[27] 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**, 1244 (2001).

[28] J. Gross and G. Sadowski, *Application of the Perturbed-Chain SAFT equation of state to associating systems,* Ind. Eng. Chem. Res. **41**, 5510 (2002).

[29] J. Gross, *An equation-of-state contribution for polar components: Quadrupolar molecules,* AIChE J. **51**, 2556 (2005).

[30] J. Gross and J. Vrabec, *An equation-of-state contribution for polar components: Dipolar molecules,* AIChE J. **52**, 1194 (2006).

[31] P. T. Sikora, *Combining rules for spherically symmetric intermolecular potentials,* J. Phys. B: At., Mol. Opt. Phys **3**, 1475 (1970).

[32] M. J. Hiza and A. G. Duncan, *A correlation for the prediction of interaction energy parameters for mixtures of small molecules,* AIChE J. **16**, 733 (1970).

[33] 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 (1973).

[34] F. Kohler, J. Fischer, and E. Wilhelm, *Intermolecular force parameters for unlike pairs,* J. Mol. Struct. **84**, 245 (1982).

[35] K. Tang and J. Toennies, *New combining rules for well parameters and shapes of the van der Waals potential of mixed rare gas systems,* Z Phys D Atom Mol Cl **1**, 91 (1986).

[36] T. A. Halgren, *The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters,* J. Am. Chem. Soc. **114**, 7827 (1992).

[37] M. Waldman and A. Hagler, *New combining rules for rare gas van der Waals parameters,* J. Comput. Chem. **14**, 1077 (1993).

[38] T. Schnabel, J. Vrabec, and H. Hasse, *Unlike Lennard-Jones parameters for vapor-liquid equilibria,* J. Mol. Liq. **135**, 170 (2007).

[39] A. J. Haslam, A. Galindo, and G. Jackson, *Prediction of binary intermolecular potential parameters for use in modelling fluid mixtures,* Fluid Phase Equilib. **266**, 105 (2008).

[40] M. Singh, K. Leonhard, and K. Lucas, *Making equation of state models predictive: Part 1: Quantum chemical computation of molecular properties,* Fluid Phase Equilib. **258**, 16 (2007).

[41] K. Leonhard, N. V. Nhu, and K. Lucas, *Making equation of state models predictive: Part 2: An improved PCP-SAFT equation of state,* Fluid Phase Equilib. **258**, 41 (2007).

[42] F. T. Peters, F. S. Laube, and G. Sadowski, *PC-SAFT based group contribution method for binary interaction parameters of polymer/solvent systems,* Fluid Phase Equilib. **358**, 137 (2013).

[43] D. Nguyen-Huynh, J.-P. Passarello, P. Tobaly, and J.-C. de Hemptinne, *Modeling phase equilibria of asymmetric mixtures using a group-contribution SAFT (GC-SAFT) with a k_{ij} correlation method based on London's Theory. 1. application to CO_2 + n-alkane, methane + n-alkane, and ethane + n-alkane systems,* Ind. Eng. Chem. Res. **47**, 8847 (2008).

[44] D. Nguyen-Huynh, T. K. S. Tran, S. Tamouza, J.-P. Passarello, P. Tobaly, and J.-C. de Hemptinne, *Modeling phase equilibria of asymmetric mixtures using a group-contribution SAFT (GC-SAFT) with a k_{ij} correlation method based on London's theory. 2. application to binary mixtures containing aromatic hydrocarbons, n-alkanes, CO_2, N_2, and H_2S,* Ind. Eng. Chem. Res. **47**, 8859 (2008).

[45] S. Tamouza, J.-P. Passarello, P. Tobaly, and J.-C. de Hemptinne, *Group contribution method with SAFT EOS applied to vapor liquid equilibria of various hydrocarbon series,* Fluid Phase Equilib. **222-223**, 67 (2004).

[46] J. A. P. Coutinho, P. M. Vlamos, and G. M. Kontogeorgis, *General form of the cross-energy parameter of equations of state,* Ind. Eng. Chem. Res. **39**, 3076 (2000).

[47] G. M. Kontogeorgis, E. C. Voutsas, I. V. Yakoumis, and D. P. Tassios, *An equation of state for associating fluids,* Ind. Eng. Chem. Res. **35**, 4310 (1996).

[48] M. Shacham, G. S. Cholakov, R. P. Stateva, and N. Brauner, *Quantitative structure-property relationships for prediction of phase equilibrium related properties,* Ind. Eng. Chem. Res. **49**, 900 (2010).

[49] A. M. Abudour, S. A. Mohammad, R. L. R. Jr., and K. A. Gasem, *Generalized binary interaction parameters for the Peng-Robinson equation of state,* Fluid Phase Equilib. **383**, 156 (2014).

[50] G. H. Hudson and J. McCoubrey, *Intermolecular forces between unlike molecules,* Trans. Faraday Soc. **56**, 761 (1959).

[51] G. M. Kontogeorgis, *Association theories for complex thermodynamics,* Chem. Eng. Res. Des. **91**, 1840 (2013).

[52] A. Katritzky, M. Karelson, and R. Petrukhin, *CODESSA PRO project,* (2005).

[53] TALETE-SRL, *Dragon 6,* (2013).

[54] A. Kazakov, C. D. Muzny, V. Diky, R. D. Chirico, and M. Frenkel, *Predictive correlations based on large experimental datasets: Critical constants for pure compounds,* Fluid Phase Equilib. **298**, 131 (2010).

[55] M. A. Sobati and D. Abooali, *Molecular based models for estimation of critical properties of pure refrigerants: Quantitative structure property relationship (QSPR) approach,* Thermochim. Acta **602**, 53 (2015).

[56] E. N. Muratov, E. V. Varlamova, A. G. Artemenko, P. G. Polishchuk, and V. E. Kuzmin, *Existing and developing approaches for QSAR analysis of mixtures,* Mol. Inform. **31**, 202 (2012).

[57] M. Kleiner and G. Sadowski, *Modeling of polar systems using PCP-SAFT: An approach to account for induced-association interactions,* J. Phys. Chem. C **111**, 15544 (2007).

[58] G. M. Kontogeorgis and G. K. Folas, *Thermodynamic models for industrial applications,* in *Thermodynamic Models for Industrial Applications* (John Wiley & Sons, Ltd, 2009).

[59] J. P. Wolbach, , and S. I. Sandler, *Using molecular orbital calculations to describe the phase behavior of cross-associating mixtures,* Ind. Eng. Chem. Res. **37**, 2917 (1998).

[60] N. von Solms, M. L. Michelsen, and G. M. Kontogeorgis, *Applying association theories to polar fluids,* Ind. Eng. Chem. Res. **43**, 1803 (2004).

[61] M. Kleiner, F. Tumakaka, and G. Sadowski, *Thermodynamic modeling of complex systems,* (Springer Berlin Heidelberg, 2008) pp. 1–34.

[62] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular theory of gases* (John Wiley & Sons, 1987).

[63] S. H. Huang and M. Radosz, *Equation of state for small, large, polydisperse, and associating molecules,* Ind. Eng. Chem. Res. **29**, 2284 (1990).

[64] F. Ruether and G. Sadowski, *Modeling the solubility of pharmaceuticals in pure solvents and solvent mixtures for drug process design,* J. Pharm. Sci. **98**, 4205 (2009).

[65] AIChE-DECHEMA, *Pure Component Database DIPPR-801*, Tech. Rep. (AIChE-DECHEMA, 2011).

[66] *Dortmund Data Bank Software,* (2015).

[67] J. Gmehling, B. Kolbe, M. Kleiber, and J. Rarey, *Chemical thermodynamics for process simulation* (Wiley-VCH Verlag & Co. KGaA, 2012).

[68] G. Soave, S. Gamba, and L. A. Pellegrini, *SRK equation of state: Predicting binary interaction parameters of hydrocarbons and related compounds,* Fluid Phase Equilib. **299**, 285 (2010).

[69] Mathworks, *Matlab(r), mathworks - documentation center,* (2015).

[70] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 5th ed. (John Wiley & Sons, 2012).

[71] P. Gramatica, P. Pilutti, and E. Papa, *Validated QSAR prediction of OH tropospheric degradation of VOCs: Splitting into training-test sets and consensus modeling,* J. Chem. Inf. Comput. Sci. **44**, 1794 (2004).

[72] *OECD Guidance document on the validation of Quantitative Structure-Activity Relationships QSAR models 2007,* .

[73] A. Tropsha, *Best practices for QSAR model development, validation, and exploitation,* Mol. Inform. **29**, 476 (2010).

[74] M. Daszykowski, B. Walczak, and D. Massart, *Representative subset selection,* Anal. Chim. Acta **468**, 91 (2002).

[75] R. W. Kennard and L. Stone, *Computer aided design of experiments,* Technometrics **11**, 137 (1969).

[76] P. de Groot, G. Postma, W. Melssen, and L. Buydens, *Selecting a representative training set for the classification of demolition waste using remote NIR sensing,* Anal. Chim. Acta **392**, 67 (1999).

[77] B. Bourguignon, P. F. de Agular, M. S. Khots, and D. L. Massart, *Optimization in irregularly shaped regions: pH and solvent strength in reversed-phase high-performance liquid chromatography separations,* Anal. Chem. **66**, 893 (1994).

[78] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics* (Pearson, 2013).

[79] A. R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn, and D. A. Dobchev, *Quantitative correlation of physical and chemical properties with chemical structure: Utility for prediction,* Chem. Rev. **110**, 5714 (2010).

[80] V. Prana, P. Rotureau, G. Fayet, D. Andre, S. Hub, P. Vicot, L. Rao, and C. Adamo, *Prediction of the thermal decomposition of organic peroxides by validated QSPR models,* J. Hazard. Mater. **276**, 216 (2014).

[81] V. Consonni, D. Ballabio, and R. Todeschini, *Evaluation of model predictive ability by external validation techniques,* J. Chem. **24**, 194 (2010).

[82] A. Golbraikh and A. Tropsha, *Beware of q2!* Mol. Graphics Modell. **20**, 269 (2002).

[83] *CRC Handbook of Chemistry and Physics,* Internet Version 2005 (2005).

[84] M. Hechinger, *Model-based Identification of Promising Biofuel Candidates for Spark-ignited Engines*, Ph.D. thesis, Rheinisch-Westfalischen Technischen Hochschule Aachen (2014).

4

CoMT-CAMD for CO_2-Absorption: The Impact of Binary Interaction Parameters Predictions

The CoMT-CAMD approach relies on the predictive capabilities of the chosen thermodynamic model. Separation processes, in particular, require the estimation of mixture properties. Predictions of mixture phase equilibria or other properties that require a sufficiently accurate representation of excess properties are particularly demanding for thermodynamic models. This Chapter investigates how two modeling approaches for the PC-SAFT equation of state impact the resulting list of the proposed solvents for the CO_2-capture process. The two models are:

1. PC-SAFT, applied with k_{ij}=0 for all pairs of substances involving the target solvent
2. PC-SAFT, applied with the QSPR-relation for k_{ij} developed in Chapter 3 for pairs of substances involving the hypothetical target solvent

Both modeling approaches are assessed by comparison to results obtained with individually optimized k_{ij}-parameters and with fully optimized process settings.

4.1. Case study: Process and problem description

The examined case study preserves the topology of the pre-combustion CO_2 capture process described in Chapter 2 (Figure 2.3, Appendix B.1). The process specifications are given in Table 4.1 and differ slightly from Chapter 2. The efficiency of the pump for lean solvent recirculation has been changed according to ref. [1]. Besides, the compressor is now modeled with 3 compression stages for the (CO_2-rich) vapor stream exiting the desorption flash to the pipeline pressure ($N_{\text{comp-des}} = 3$). A 3-staged compression is a more realistic design option, leading to adequate compression ratios r [1], in the range between 2.9 and 4.8[1].

The simultaneous process and solvent optimization problem is formulated as in problem (2.2), for the cost objective function described in Chapter 2 and Appendix B.2. In the current chapter, only the case of polar solvents was examined. The implementation of CoMT-CAMD to polar solvents is a more demanding problem than the implementation to non-polar solvents due to the additional degree of freedom of dipole moment. In order to examine the limits of the proposed implementation of k_{ij}-predictions, the problem of simultaneous process and polar solvent optimization is therefore seen as a more suitable case-study.

The degrees of freedom are the process variables $\overline{x} = \{P^{\text{des}}, P^{\text{rg}}, T^{\text{lean}}\}$ and the molecular parameters of the solvent $\overline{p} = \{m, \sigma, \varepsilon/k, \mu/(m\sigma^3(\varepsilon/k))^{0.5}\}$. The lower and upper bounds of the variables during optimization remain also unchanged. The database of candidate solvents used in the mapping step is the database of polar, non-associating organic components (Appendix C), same as for the problem described in Chapter 2.

Residual caloric properties and phase equilibria are calculated with the PC-SAFT EoS. The ideal gas heat capacity c_p^{ig} and the molar mass M of the solvent are estimated with QSPR models, as functions of the PC-SAFT pure component parameters. Binary interaction parameters k_{ij} for mixtures of the optimized solvent with CO_2 and H_2 either are predicted using QSPR models (Chapter 3), or k_{ij} is set to zero for these binary pairs. The CoMT-CAMD results are compared. The k_{ij} values for the binary mixtures CO_2 - H_2O, CO_2 - H_2 and H_2 - H_2O were adjusted to experimental mixture data. The PC-SAFT pure component parameters for CO_2, H_2 and H_2O and the k_{ij} values of their binary mixtures used in this case study[2] are given in Tables 4.2 and 4.3.

[1] $r_{lb} = \left(\frac{P^{\text{ppl}}}{P^{\text{des}}_{\text{lb}}}\right)^{\left(1/N_{\text{comp-des}}\right)}$ and $r_{ub} = \left(\frac{P^{\text{ppl}}}{P^{\text{des}}_{\text{ub}}}\right)^{\left(1/N_{\text{comp-des}}\right)}$ with $P^{\text{des}}_{\text{lb}} = 0.1013\,\text{MPa}$ and $P^{\text{des}}_{\text{ub}} = 0.4500\,\text{MPa}$.

[2] The PC-SAFT pure component parameters for CO_2 and H_2 are the same with those used in Chapter 2. The PC-SAFT pure component parameters for H_2O are slightly different, since the dipole moment of H_2O is additionally considered. The k_{ij} values for the binary mixtures of CO_2, H_2 and H_2O were adjusted to a higher number of experimental data and using the PC-SAFT pure component parameters given in Table 4.2.

Table 4.1: Stream, process and equipment specifications adopted from the Buggenum pilot plant [2]. The pipeline pressure is taken from ref. [3]; efficiencies from ref. ([4]); the efficiency of the lean solvent pump and the number of compression stages for the pipeline compressor estimated according to ref. [1].

Stream specifications		
Syngas feed composition	$x^{sg}_{H_2}$ / mol/mol	0.5480
	$x^{sg}_{CO_2}$ / mol/mol	0.4485
	$x^{sg}_{H_2O}$ / mol/mol	0.0035
Syngas feed temperature	T^{sg} / K	313.00
Rich gas recycle temperature	T_{rg} / K	313.00
Pipeline gas temperature	T^{ppl} / K	313.00
Pipeline gas pressure	P^{ppl} / MPa	11.00
Solvent make-up temperature	T^{MU} / K	298.15
Solvent make-up pressure	P^{MU} / MPa	0.1013
Process specifications		
CO_2 capture rate in absorber	α_{CO_2} / –	0.90
Absorption pressure	P_{abs} / MPa	2.00
Equipment specifications		
Absorber - number of stages	N_{abs} / –	7
Rich gas compressor - number of stages	$N_{comp-rg}$ / –	1
Pipeline compressor - number of stages	$N_{comp-des}$ / –	3
Compressors - efficiency	η_{comp} / –	0.82
Solvent recirculating pump - efficiency	η_{pump} / –	0.90
Auxiliary pumps - efficiency	$\eta_{aux.pump}$ / –	0.75

Table 4.2: PC-SAFT pure component parameters for CO_2, H_2 and H_2O used in the current study.

Substance	M	m	σ	ε/k	μ	Q	κ^{AB}	ε^{AB}/k	Ref.
	$g \cdot mol^{-1}$	–	Å	K	D	DÅ	–	K	
CO_2	44.01	1.5131	3.1869	163.33	-	4.4	-	-	[5]
H_2	2.016	1.0000	2.9280	37.00	-	-	-	-	[6]
H_2O	18.02	1.3886	2.7139	339.14	1.85	-	0.03	1687.1	this work

Table 4.3: Binary interaction parameters k_{ij} for the mixtures CO_2-H_2O, CO_2-H_2 and H_2-H_2O adjusted to experimental VLE data.

Binary mixture	k_{ij}	$\%AAD(x^L)$	$\%AAD(x^V)$	Temp. range	Number of experimental data points	Ref.
CO_2-H_2O	-0.0766	0.16	0.70	323 K – 473 K	56	[7–10]
CO_2-H_2	-0.0623	1.45	2.23	225 K – 280 K	92	[11–13]
H_2-H_2O	-0.3945	0.09	-	273 K – 373 K	40	[14, 15]

4.2. Prediction of k_{ij} values for the solvent binary mixtures

Values of the binary interaction parameter k_{ij} for mixtures of the optimized solvent with CO_2 and H_2 are estimated using the descriptors of the QSPR model presented in Eq.(3.6) to (3.17) (Chapter 3). The limited accuracy of the QSPR model in the prediction of k_{ij} values for binary mixtures of one non-associating, dipolar and one associating component has been taken into consideration. Therefore, the values of k_{ij} for the binary mixtures of polar solvents with H_2O have been set equal to zero (no correction is implemented).

The QSPR model presented in Chapter 3 leads, generally, to satisfactory correlations of k_{ij} and better predictions of phase equilibria with PC-SAFT. Nevertheless, the training set of the QSPR method presented in Chapter 3 does not include mixtures with H_2. Binary mixtures with H_2 require unusual k_{ij}-values and mixtures with H_2 were deliberately excluded from the training set in Chapter 3. Here, values of k_{ij} for mixtures of the solvent with H_2 are estimated using the descriptors of the QSPR model presented in Chapter 3 (Eq.(3.6) to (3.17)), with coefficients adjusted only for binary mixtures containing H_2. The training set and the corresponding coefficients of the QSPR model for the estimation of k_{ij} values for binary mixtures with H_2 are given in Appendix E.4. Additionally, the training set for binary mixtures with one non-associating, dipolar and one non-associating, quadrupolar

component was extended here to include additional mixtures with CO_2. In order to provide more accurate predictions of k_{ij} for mixtures including CO_2, the QSPR model coefficients were adjusted individually for this group. The new model coefficients and the extended training set are given in Appendix E.5. The performance of both QSPR models, evaluated using the internal and external validation techniques of the QSPR method, is given in Table 4.4.
With multiple correlation coefficients $R^2 \geq 80\,\%$, both models can be considered to be predictive. The high values of Q^2_{ext} indicate high predictive power of the QSPR models. (For a more detailed discussion on the meaning of the QSPR-measures for internal and external model validation see Chapter 3).

Table 4.4: Results of internal and external validation of the QSPR models (Eq.(3.19)) with coefficients adjusted for binary mixtures with H_2 and CO_2 (Appendix E.4 and E.5).

k_{ij} QSPR model	$R^2(\%)$	$Q^2_{\text{ext}}(\%)$
mixtures with H_2	83.9	88.3
mixtures with CO_2	79.5	99.2

4.3. CoMT-CAMD results

The PC-SAFT parameters of the optimal (hypothetical) polar solvent $\overline{p}^{opt}$ at optimal process conditions $\overline{x}^{opt}$ are calculated in the CoMT step. The calculations in the CoMT step were carried out using k_{ij}-estimations (k_{ij}^{pred}) and with k_{ij}=0, separately. The results of the CoMT step for both cases are presented in Table 4.5. The PC-SAFT parameters and the optimal process conditions for the best identified real molecule (propylene carbonate) are given as a reference: The value of k_{ij} for the binary mixture of propylene carbonate with CO_2 is directly adjusted to experimental data (k_{ij}^{fit}). The estimation of k_{ij} relies on the PC-SAFT parameters. For the set of PC-SAFT parameters of the hypothetical molecule, the estimated value of k_{ij} for the mixture hypothetical solvent-CO_2 is very close to zero ($k_{ij,(CO_2-\text{solv.})} = 0.0004$). Thus, the results of the CoMT step with k_{ij}-estimations and with $k_{ij} = 0$ in Table 4.5 are very similar.

The best performing polar components are identified using the mapping procedure described in Chapter 2: First, the process inequality constraint (solvent emission $\mathrm{m_S}$) of 444 polar, non-associating real molecules is approximated using a first-order Taylor series expansion (Eq.(2.4)); then, the performance of the real molecules, that are feasible according to the approximation of the constraint ($\mathrm{m_S}^{appr} \leq 20\,\mathrm{mg/Nm^3}$), is evaluated using a second-order Taylor approximation of the cost objective function f^{appr} (Eq.(2.3)); finally, a ranking of the feasible real molecules regarding their performance measured by f^{appr} delivers the mapping list. The set of the ten best performing components identified in the mapping step is presented in Table 4.6. The results in Table 4.6 confirm the findings of Chapter 2. The mapping procedure successfully identifies the real molecule that lies in the

Table 4.5: Comparison of the result of the CoMT step with k_{ij}-estimations for polar solvents to the result of the CoMT step with $k_{ij} = 0$ and to the result of individual process optimization for the best identified real molecule (propylene carbonate). The value of k_{ij} for the binary mixture of propylene carbonate with CO_2 is adjusted to experimental data.

	optimal polar molecule	optimal polar molecule	best polar real molecule
	hypothetical ($k_{ij} = 0$)	hypothetical ($k_{ij}^{\text{pred}} = 0.0004$)	propylene carbonate ($k_{ij}^{\text{fit}} = 0.0314$)
$\mathrm{M/g/mol}$	99.57	98.72	102.09
$m/-$	3.3116	3.3067	3.3386
σ/Å	3.3689	3.3728	3.3506
ε/k/K	309.80	308.00	312.75
μ/D	4.9669	4.9629	4.9765
		optimal process conditions	
P^{des}/MPa	0.139	0.139	0.131
P^{rg}/MPa	0.893	0.892	0.898
T^{lean}/K	277.85	277.18	278.78
$\mathrm{m_S/mg/Nm^3}$	3.577	3.828	3.006
$f_{\text{electr.demand}}/€/\mathrm{T_{CO_2\,capt.}}$	5.612	5.624	5.966
$f_{\text{H2 loss}}/€/\mathrm{T_{CO_2\,capt.}}$	0.068	0.101	0.131
$f_{\text{solvent loss}}/€/\mathrm{T_{CO_2\,capt.}}$	0.112	0.118	0.090
$f_{\text{utilities}}/€/\mathrm{T_{CO_2\,capt.}}$	0.745	0.767	0.894
$f/€/\mathrm{T_{CO_2\,capt.}}$	6.537	6.610	7.081

immediate vicinity of the optimal (hypothetical) one, with respect to the molecular parameters and optimal process conditions. Results from individual process optimizations ('real ranking') confirm the validity of the mapping step ('mapping ranking' of the 10 best performing components according to the approximation of the objective function).

Table 4.6: Best polar candidate solvents identified by CoMT-CAMD. The ranking predicted in the mapping step ('mapping ranking') is compared to the ranking after individual process optimization ('real ranking'). The optimal process conditions $\{P^{\mathrm{des}}, P^{\mathrm{rg}}, T^{\mathrm{lean}}\}^{opt}$ and the value of the objective function $f(k_{ij}^{\mathrm{pred}})$ from individual process optimizations are given. The value of $\mathrm{m_S}$ indicates a components' feasibility against the inequality process constraint ${\mathrm{m_S}}^{opt} < 20\,\mathrm{mg/Nm^3}$. Components' entries with active constraints at optimum (infeasible components) are noted with a star (*). The melting point temperature T_{mp} is the lower bound of temperature allowed in individual process optimizations to avoid solidification of the solvent.

IUPAC Name	ranking	
	mapping	real (k_{ij}^{pred})
propylene carbonate	1	1
diethyl sulfate	2	2
γ-valerolactone	3	5
methyl maleic anhydride	4	*
N-methyl-2-pyrrolidone	5	4
butyric anhydride	6	3
γ-butyrolactone	7	8
isobutyric anhydride	8	7
propionic anhydride	9	11
diethyl malonate	10	12

IUPAC Name	$f(k_{ij}^{\mathrm{pred}})$	${\mathrm{m_S}}^{opt}$	P_{des}^{opt}	P_{rg}^{opt}	T_{lean}^{opt}	$\mathrm{T_{mp}}$
	€/$\mathrm{T_{CO_2\,capt.}}$	mg/Nm³	MPa	MPa	K	K
propylene carbonate	6.837	1.962	0.144	0.927	274.6	224.9
diethyl sulfate	7.138	2.842	0.148	0.934	260.0	248.2
γ-valerolactone	7.322	10.952	0.144	0.927	269.4	242.2
methyl maleic anhydride	*	*	*	*	*	281.1
N-methyl-2-pyrrolidone	7.284	8.077	0.142	0.927	268.1	249.2
butyric anhydride	7.254	4.003	0.155	0.938	255.4	199.9
γ-butyrolactone	7.668	15.694	0.135	0.929	275.1	229.8
isobutyric anhydride	7.589	7.189	0.160	0.940	251.6	219.7
propionic anhydride	7.816	12.623	0.151	0.932	254.1	228.2
diethyl malonate	8.309	1.570	0.142	0.948	253.6	224.3

In Table 4.7, the 'mapping list' (10 best performing solvents) obtained by CoMT-CAMD with the k_{ij}-estimation models is compared to the ranking of the same components in the 'mapping list' obtained by CoMT-CAMD with $k_{ij} = 0$. Propylene carbonate, N-methyl-2-pyrrolidone and γ-valerolactone are among the best performing solvents in both mapping lists. The difference in the ranking of the components can be attributed to the implementation of the k_{ij}-estimation models.

Table 4.7: Influence of k_{ij}-estimations on the mapping ranking. The mapping ranking of the 10 best performing solvents obtained by CoMT-CAMD with k_{ij}-estimations (k_{ij}^{pred}) is compared to the mapping ranking of the same components obtained by CoMT-CAMD with $k_{ij} = 0$.

	mapping ranking	
IUPAC Name	# f^{appr} (k_{ij}-estimations)	# f^{appr} ($k_{ij} = 0$)
propylene carbonate	1	1
diethyl sulfate	2	5
γ-valerolactone	3	2
methyl maleic anhydride	4	3
N-methyl-2-pyrrolidone	5	4
butyric anhydride	6	8
γ-butyrolactone	7	6
isobutyric anhydride	8	10
propionic anhydride	9	13
diethyl malonate	10	21

In the following, results from individual process optimizations with predicted k_{ij} values for the binary mixture of the solvent with CO_2 are compared to results with k_{ij} values independently adjusted to experimental data. Due to the scarcity of experimental data for binary mixtures with H_2, the k_{ij} values for binary mixtures of the solvent with H_2 were in both cases set equal to the predicted values. A further comparison with results from individual process optimizations with $k_{ij} = 0$ for the solvent binary mixtures intends to show in which extend the accuracy of the CoMT-CAMD result is improved by the implementation of the k_{ij}-estimation method. Besides, for the sake of consistency, k_{ij} values for binary mixtures of the solvent with H_2O are defined equal to zero throughout. A matrix that summarizes the k_{ij} estimation method for each case is given in Table 4.8. The results of the individual process optimizations are shown in Table 4.9.

According to the results in Table 4.9, the predicted performance with estimated k_{ij} values (case 1) is closer to the actual one (reference: case 2). When no correction is used in phase equilibrium calculations ($k_{ij} = 0$), the value of the objective function is consistently underestimated. The phase equilibrium diagrams presented in Figure 4.1 support this observation. Compared to experimental data, PC-SAFT predictions with $k_{ij} = 0$ overestimate the concentration of CO_2 in the liquid phase. For the binary mixture of propylene carbonate with CO_2, no significant correction

Table 4.8: Comparison matrix for results evaluation, according to the k_{ij} estimation method.

	k_{ij} evaluation method		
binary mixture	case 1: $f(k_{ij}^{\text{pred}})$	case 2: $f(k_{ij}^{\text{fit}})$	case 3: $f(k_{ij}=0)$
solvent-CO2	QSPR model	exp. data	0
solvent-H2	QSPR model	QSPR model	0
solvent-H2O	0	0	0

Table 4.9: Influence of k_{ij} estimation method to the calculated optimal performance of the system. Results from individual optimizations ('real f') and the resulting ranking of the components ('real ranking') are presented for the 3 cases of k_{ij} estimation methods in Table 4.8. The value of the objective function f is given in units €/τ_{CO_2}.

mapping list (CoMT-CAMD with k_{ij}-estimations)	real f (individual process optimizations)			real ranking (individual process optimizations)		
	case 1 $f(k_{ij}^{\text{pred}})$	case 2 $f(k_{ij}^{\text{fit}})$	case 3 $f(k_{ij}=0)$	case 1 $f(k_{ij}^{\text{pred}})$	case 2 $f(k_{ij}^{\text{fit}})$	case 3 $f(k_{ij}=0)$
propylene carbonate	6.837	7.081	6.545	1	1	1
diethyl sulfate						
γ-valerolactone	7.322	7.433	6.923	3	3	3
methyl maleic anhydride						
N-methyl-2-pyrrolidone	7.284	7.162	6.963	2	2	4
butyric anhydride						
γ-butyrolactone	7.668	7.479	6.722	4	4	2
isobutyric anhydride						
propionic anhydride						
diethyl malonate						

is achieved with the estimated k_{ij} value. As can be seen in Table 4.10, the estimated k_{ij} value for this mixture is too low, closer to 0 than to the correction actually necessary. However, the evaluation of the results generally shows that phase equilibrium calculations with estimated k_{ij} values describe the real performance of the examined systems (experimental data) with higher accuracy (Table 4.8).

A final point should be commented, namely the decision not to implement a binary interaction correction for the mixtures of the solvent with H_2O. Mixtures of H_2O with dipolar, non-associating components are difficult to predict for thermodynamic models and they require correction ($k_{ij} \neq 0$) for more accurate descriptions of phase equilibria. As already discussed in Chapter 3, mixtures with one dipolar, non-associating component and one associating component are most demanding, due to the effects of induced association [21]. In order to avoid introducing unnecessary uncertainties, more accurate correlations of k_{ij} are required for this type

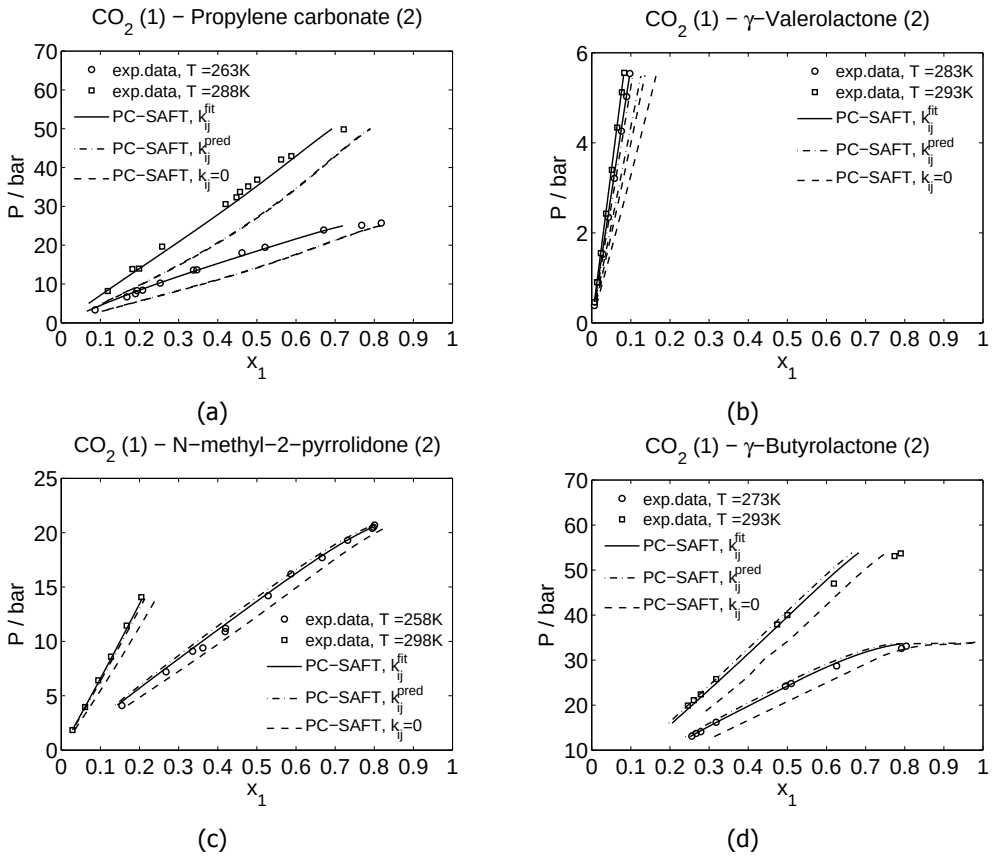


Figure 4.1: Vapor-liquid equilibria for binary mixtures of best performing polar solvents with CO_2: a) CO_2 - propylene carbonate [16], b) CO_2 - γ-valerolactone [17], c) CO_2 - N-methyl-2-pyrrolidone [18, 19] and d) CO_2 - γ-butyrolactone [20]. Predictions of PC-SAFT with estimated k_{ij} values (k_{ij}^{pred}) are compared to isothermal experimental data. Correlations with k_{ij} values adjusted to experimental data (k_{ij}^{fit}) and without correction ($k_{ij} = 0$) are also included. The values of the binary interaction parameter (k_{ij}^{pred} and k_{ij}^{fit}) used in the calculations for each binary mixture are given in Table 4.10.

of mixtures. Generally, an improved description of mixtures with H_2O is necessary. In this case an extension of the process model to include water separation units and the implementation of robust methods for liquid-liquid equilibrium calculations might be needed.

Table 4.10: Binary interaction parameters k_{ij} for the binary mixtures of CO_2 with polar solvents. k_{ij}^{fit} was estimated using a QSPR model (Appendix E.5). k_{ij}^{fit} was adjusted to experimental data, which were obtained from the Dortmund DataBase [22] and passed the thermodynamic consistency tests [23].

Binary mixture	k_{ij}^{pred}	k_{ij}^{fit}	Temp. range	n	Ref.
CO_2 - propylene carbonate	0.0010	0.0314	228 K – 463 K	245	DDB [22]
CO_2 - γ-valerolactone	0.0243	0.0466	283 K – 323 K	40	DDB [22]
CO_2 - N-methyl-2-pyrrolidone	0.0131	0.0163	258 K – 373 K	139	DDB [22]
CO_2 - γ-butyrolactone	0.0221	0.0184	273 K – 308 K	62	DDB [22]

4.4. Conclusions

k_{ij}-estimation models were implemented in the CoMT-CAMD framework for the binary mixtures of the solvent with CO_2 and H_2. The implementation of k_{ij}-estimation models in CoMT-CAMD has an impact on the mapping ranking. QSPR relations for the prediction of k_{ij} are embedded into the PC-SAFT model. Thereby, the derivatives calculated at optimum for the approximation of the objective function $f^{\text{appr}}(\overline{x}, \overline{y})$ and of the constraint $m_S^{\text{appr}}(\overline{x}, \overline{y})$ are different than the derivatives with $k_{ij} = 0$. Although the optimal parameters of the hypothetical molecule (result of the CoMT step) with k_{ij}-estimations and with $k_{ij} = 0$ are similar, the mapping ranking is different, because the same change in the molecular parameters results to a different change in the objective function.

A comparison of phase equilibria calculations with predicted k_{ij}-values and with $k_{ij} = 0$ to experimental data of vapor-liquid equilibria was carried out for best performing real solvents. The results of phase equilibria calculations with predicted values of k_{ij} are closer to the experimental data than the results obtained with $k_{ij} = 0$. Considering values of the objective function obtained by calculations with k_{ij} adjusted to experimental data as a reference, the implementation of the k_{ij}-estimations gives more representative results than when $k_{ij} = 0$ is used.

The results of this chapter show, that the prediction of k_{ij}-values with the QSPR models developed in this thesis enhances the accuracy in phase equilibria calculations and thereby the accuracy of CoMT-CAMD results.

References

[1] R. Smith, *Chemical process design and integration* (John Wiley, 2005).

[2] K. Damen, R. Gnutek, J. Kaptein, N. R. Nannan, B. Oyarzun, C. Trapp, P. Colonna, E. van Dijk, J. Gross, and A. Bardow, *Developments in the pre-combustion CO_2 capture pilot plant at the Buggenum IGCC,* Energy Procedia **4**, 1214 (2011).

[3] J. Davison, *Performance and costs of power plants with capture and storage of CO_2,* Energy **32**, 1163 (2007).

[4] P. Chiesa, S. Consonni, T. Kreutz, and R. Williams, *Co-production of hydrogen, electricity and CO_2 from coal with commercially ready technology. Part A: Performance and emissions,* Int. J. Hydrogen Energy **30**, 747 (2005).

[5] J. Gross, *An equation-of-state contribution for polar components: Quadrupolar molecules,* AIChE J. **51**, 2556 (2005).

[6] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular theory of gases* (John Wiley & Sons, 1987).

[7] J. A. Briones, J. C. Mullins, M. C. Thies, and B. U. Kim, *Ternary phase equilibria for acetic acid-water mixtures with supercritical carbon dioxide,* Fluid Phase Equilib. (1987).

[8] A. Bamberger, G. Sieder, and G. Maurer, *High-pressure (vapor + liquid) equilibrium in binary mixtures of carbon dioxide + water or acetic acid at temperatures from 313 to 353 K,* J. Supercrit. Fluids **17**, 97 (2000).

[9] S.-X. Hou, G. C. Maitland, and J. M. Trusler, *Measurement and modeling of the phase behavior of the (carbon dioxide + water) mixture at temperatures from 298.15 K to 448.15 K,* J. Supercrit. Fluids **73**, 87 (2013).

[10] G. Mueller, *Master thesis, universitaet kaiserslautern,* (1983).

[11] C. Tsang and W. Street, *Phase equilibria in the H_2/CO_2 system at temperatures from 220 to 290 K and pressures to 172 MPa,* Chem. Eng. Sci. **36**, 993 (1981).

[12] P. Barrick, C. Heck, and J. Spano, *Tech.Rep., Rep.No. AFML TR 66-390, 1-67,* (1966).

[13] K. Bezanehtak, G. B. Combes, F. Dehghani, , N. R. Foster, and D. L. Tomasko, *Vapor-liquid equilibrium for binary systems of carbon dioxide + methanol, hydrogen + methanol, and hydrogen + carbon dioxide at high pressures,* J. Chem. Eng. Data **47**, 161 (2002).

[14] P. C. Gillespie and G. M. Wilson, *GPA Research Report, Rep.No. RR-41*, Tech. Rep. (1980).

[15] R. Wiebe and V. L. Gaddy, *The solubility of hydrogen in water at 0, 50, 75 and 100° C from 25 to 1000 atmospheres,* J. Am. Chem. Soc. **56**, 76 (1934).

[16] Y. P. Zubchenko, S. F. Shakhova, T. Wei, L. I. Titelman, and L. K. Kaplan, J. Appl. Chem. USSR **44(9)**, 2078 (1971).

[17] D. Deng, G. Han, Y. Jiang, and N. Ai, *Solubilities of carbon dioxide in five biobased solvents,* J. Chem. Eng. Data **60**, 104 (2015).

[18] Y. P. Zubchenko, S. Shakhova, and O. Ladygina, Soviet Chem. Ind. **17(9)**, 1077 (1982).

[19] F. Murrieta-Guevara, A. Romero-Martinez, and A. Trejo, *Solubilities of carbon dioxide and hydrogen sulfide in propylene carbonate, N-methylpyrrolidone and sulfolane,* Fluid Phase Equilib. **44**, 105 (1988).

[20] V. Matvienko and N. Yarym-Agaev, Zh. Prikl. Khim. **72(7)**, 1085 (1999).

[21] M. Kleiner and G. Sadowski, *Modeling of polar systems using PCP-SAFT: An approach to account for induced-association interactions,* J. Phys. Chem. C **111**, 15544 (2007).

[22] *Dortmund Data Bank Software,* (2015).

[23] J. Gmehling, B. Kolbe, M. Kleiber, and J. Rarey, *Chemical thermodynamics for process simulation* (Wiley-VCH Verlag & Co. KGaA, 2012).

5

Conclusions and Perspectives

5.1. Conclusions

5.1.1. Further development of the CoMT-CAMD method

In this thesis, the framework of Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD) is further developed and implemented for the simultaneous process and solvent design for CO_2 capture.

The main performance trade-offs of the CO_2-capture system are evaluated using a single, cost-based, objective function. Limitations about the environmental performance of the solvent are, additionally, considered through a process inequality constraint. The performance of candidate solvents is evaluated by an approximation of the objective function using a Taylor series expansion around the optimal solution of the CoMT optimization problem (mapping step). Evaluation of inequality process constraints was included in the mapping step using a linear approximation. Thereby, real components that do not satisfy the process inequality constraints ("infeasible components"), are excluded from the set of best performing candidates. As proven by individual process optimizations (validation of the CoMT-CAMD results), the best performing components identified in the mapping step are indeed the feasible components with best achievable performance for the problem at hand.
The results of the case studies examined in chapters 2 and 4 showed that CoMT-CAMD can be successfully solved for rigorous (reasonably complex) process models and that minor changes in the process model do not alter significantly the set of best performing real solvents. Changes in the objective function are expected, however, to have higher influence on the result of the problem. Here lies also one of the main strengths of CoMT-CAMD, being a deterministic method. In this regard, the objective function should be carefully chosen and consider all major trade-offs of the design problem.

Further, predictions of the ideal gas specific heat capacity c_p^{ig} and the molar mass M of pure components, as function of the PC-SAFT pure component parameters, were developed. Reasonable correlations of c_p^{ig} and M are established using QSPR-models, with correlation coefficients R^2 higher than 0.98 and 0.93, respectively. In CoMT-CAMD, the PC-SAFT pure component parameters of the optimized fluid (i.e. solvent) are molecular degrees of freedom for the optimization problem. The proposed QSPR models incorporated in CoMT-CAMD enable the calculation of full caloric properties and mass specific quantities.

Prediction of mixture properties and phase equilibria are particularly demanding for thermodynamic models. Often, binary interaction parameters adjusted to experimental data have to be applied, in order to allow for more accurate calculations of excess properties. In CAMD problems, binary interaction parameters cannot be adjusted to experimental mixture data. Therefore, in the course of this thesis, a QSPR-model, that relates the binary interaction parameter k_{ij} to the PC-SAFT pure component parameters, has been developed.
Good predictions of k_{ij} were obtained for binary mixtures of two non-associating substances, of one non-associating and one associating substance and of two associating substances. For these types of binary mixtures, k_{ij} can be successfully correlated to the PC-SAFT parameters of the mixture components. k_{ij} predictions

with the proposed QSPR-model for binary mixtures of one dipolar, non-associating and one associating substance were, however, only partially successful. The effects of induced association present in this type of mixtures are not trivial and their influence should be further examined.

In the last chapter of this thesis, the role of binary interaction parameters for an improved prediction of mixture properties has been investigated. CoMT-CAMD was implemented for the simultaneous process and solvent optimization for CO_2-capture, with estimations of k_{ij} in binary mixtures of the optimized solvent. The evaluation of the results showed that the accuracy of the CoMT-CAMD approach can be further enhanced by using the developed QSPR-models for k_{ij}.

5.1.2. Optimal solvents for pre-combustion CO_2 capture

The process examined in this thesis resembles in its' key features the process of CO_2 capture of the pre-combustion capture pilot plant at the IGCC power plant in Buggenum. The applied process model considers the main steps of absorption, desorption and CO_2 compression. The performance of the system is measured by a single, process-based cost objective function. The objective function is written as the sum of operating costs covering the primary energy consumption, consumption of raw materials and product loss in the major stages of the process. In addition to an economic penalty for solvent loss, a maximum value for solvent loss was considered as an inequality constraint, reflecting an emission threshold. Thereby the optimization of the overall performance of the system is attained, subject to environmental constraints.

Non-polar and polar solvents were examined separately. The set of best performing real components includes state-of-the-art solvents, like propylene carbonate and N-methyl-2-pyrrolidone. Because the optimization and the mapping procedure are fully deterministic, independent from pre-selection of candidates and heuristics, the presence of state-of-the-art solvents gives some additional confidence regarding the plausibility of the entire list of best performing solvents. Similarities in the molecular structure of the best performing components are also observed. Carbonylic oxygen and phenyl rings are present in the molecule of almost all best performing components. The favorable interactions of carbonylic oxygen and phenyl rings with CO_2 have been already studied by other authors [1–3]. The best components generally correspond to small molecules (low values of the segment number and segment diameter in the PC-SAFT representation), with high dispersive energy and dipole moment.

The best identified polar solvents show a higher potential to improve the overall performance of the process. The best polar solvents outperform DEPGs. DEPG-blends have been successfully employed in applications for pre-combustion CO_2 capture. It should be taken into account, though, that such applications (e.g. the Selexol process) are designed for the sequential removal of H_2S and CO_2. Here, the removal of H_2S is not evaluated by the objective function.

A promising result of this work is the identification of γ-valerolactone among the best performing polar solvents. γ-Valerolactone has been classified by some authors [4, 5] as a green solvent for bio-refineries and appears to be a promis-

ing solvent for CO_2 capture as well. For the binary mixture γ-valerolactone - CO_2 only scarce experimental data are available, for low pressure vapor-liquid equilibria. Additional VLE measurements at elevated pressures for the binary systems of γ-valerolactone with CO_2, H_2 and H_2O are suggested. A more thorough experimental research may confirm the result of CoMT-CAMD and reveal γ-valerolactone as a prospective, green solvent for pre-combustion CO_2 capture.

5.2. Perspectives

Founded on the use of physically based thermodynamic models, CoMT-CAMD grows fast, together with the PC-SAFT model. Developments in the PC-SAFT model support CoMT-CAMD and broaden its' application range. On the other side, CoMT-CAMD poses constantly challenging questions regarding the prediction of thermodynamic properties with PC-SAFT.

A prospective future extension of the CoMT-CAMD framework is to consider investment costs in the objective function. Estimating investment costs requires sizing (i.e. dimensioning) of key equipment, which in turn necessitates transport properties, like viscosity, heat transfer and diffusion coefficients. In CoMT-CAMD, the transport properties of the optimized fluid have to be expressed as functions of the PC-SAFT pure component parameters. The prediction of viscosity, heat transfer and diffusion coefficients for pure substances and mixtures with PC-SAFT is an ongoing work [6, 7], that will allow, for example, the design of heat exchangers and rate-based models for absorption.

CoMT-CAMD has been implemented for the design of new solvent molecules, beyond an existing substance-database. [8] The procedure for the identification of the optimal molecule at optimal process conditions (CoMT optimization step) remains unchanged. For the design of new solvent molecules, a group-contribution approach of PC-SAFT (GPC-SAFT [9]) is employed and together with feasibility constraints on the molecular structure in the mapping step, as described in ref. [8] by the present author and co-workers. The PC-SAFT pure component parameters of the designed solvent are calculated from the parameters of functional groups, using GPC-SAFT [9]. So far, only molecular structures with a single polar group have been considered. The parameterization of GPC-SAFT to molecules with multiple polar groups will increase accuracy and thus allow for the design of more powerful fluids. Further, the implementation of a hetero-segmented group-contribution approach for PC-SAFT (ref. [9]) may allow for the distinction between isomers.

A first step towards the extension of CoMT-CAMD to the design of fluid mixtures has already been taken by the implementation of the method to working fluid mixtures for ORC applications [10]. For the design of binary fluid mixtures, the pure component parameters of the second fluid are introduced as additional molecular degrees of freedom to the optimization problem. In the mapping step, the implementation of a CAMD approach, with predictions of k_{ij} values based on GPC-SAFT, seems to be a promising approach.

The presence of water is for many problems a limiting design factor. The implementation of robust methods for the prediction of liquid-liquid phase equilibria with PC-SAFT would help to consider the influence of water in mixture streams with the

optimized fluid. Finally, robust liquid-liquid equilibrium calculations will support the implementation of CoMT-CAMD to the optimization of liquid extraction process.

A perspective of CoMT-CAMD on a somewhat longer time scale is given for the modeling of (solid) materials and products determined by structure including interfaces.

CoMT-CAMD is a platform, on which molecular thermodynamics, process engineering and modern algorithmic techniques combine their strengths, in order to approach the demanding problem of integrated process and molecular design. Heuristic decisions give their place to deterministic solutions and CoMT-CAMD allows for a new generation of optimal process and fluid designs.

References

[1] S. G. Kazarian, M. F. Vincent, F. V. Bright, C. L. Liotta, and C. A. Eckert, *Specific intermolecular interaction of carbon dioxide with polymers,* J. Am. Chem. Soc. **118**, 1729 (1996).

[2] M. R. Nelson and R. F. Borkman, *Ab initio calculations on CO_2 binding to carbonyl groups,* J. Phys. Chem. A **102**, 7860 (1998).

[3] M. B. Miller, D. R. Luebke, and R. M. Enick, *CO_2-philic oligomers as novel solvents for CO_2 absorption,* Energy Fuels **24**, 6214 (2010).

[4] J. H. Clark, F. E. I. Deswarte, and T. J. Farmer, *The integration of green chemistry into future biorefineries,* Biofuels, Bioprod., Biorefin. **3**, 72 (2009).

[5] D. M. Alonso, S. G. Wettstein, and J. A. Dumesic, *Gamma-valerolactone, a sustainable platform molecule derived from lignocellulosic biomass,* Green Chem. **15**, 584 (2013).

[6] O. Lötgering-Lin and J. Gross, *Group contribution method for viscosities based on entropy scaling using the perturbed-chain polar statistical associating fluid theory,* Ind. Eng. Chem. Res. **54**, 7942 (2015).

[7] M. Hopp, O. Lötgering-Lin, and J. Gross, *Transportgrössen aus der Entropieskalierung mit PC-SAFT: Viskosität und Wärmeleitfähigkeit,* in *Thermodynamik Kolloquium 2015* (2015).

[8] M. Lampe, M. Stavrou, J. Schilling, E. Sauer, J. Gross, and A. Bardow, *Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT,* Comput. Chem. Eng. **81**, 278 (2015).

[9] E. Sauer, M. Stavrou, and J. Gross, *Comparison between a Homo- and a Heterosegmented Group Contribution approach based on the Perturbed-Chain Polar Statistical Associating Fluid Theory Equation of State,* Ind. Eng. Chem. Res. **53**, 14854 (2014).

[10] M. Lampe, P. Edel, J. Schilling, J. Gross, and A. Bardow, *Integrated design of working fluid mixtures and Organic Rankine Cycles (ORC) in the Continuous Molecular Targeting (CoMT) framework,* 3rd International Seminar on ORC Power Systems, Brussels, Belgium (2015).

A

The PC-SAFT EoS

This appendix summarizes the equations used for calculating thermodynamic properties from the PC-SAFT equation-of-state (EoS). The equations presented bellow are directly adopted from ref. [1–5].

A.1. Helmholtz energy

Hard-chain (reference fluid) contribution. The contribution of the hard-chain reference fluid is given as

$$\tilde{a}^{hc} = \overline{m}\tilde{a}^{hs} - \sum_i x_i \,(m_i - 1)\ln g_{ii}^{hs}(\sigma_{ii}) \tag{A.1}$$

with $\overline{m} = \sum_i x_i m_i$ the mean segment number in the mixture.
The contribution of the hard-sphere reference $\tilde{a}^{hs}$ is given by

$$\tilde{a}^{hs} = \frac{1}{\zeta_0}\cdot\left[\frac{3\zeta_1\zeta_2}{(1-\zeta_3)} + \frac{\zeta_2^3}{\zeta_3\,(1-\zeta_3)^2} + \left(\frac{\zeta_2^3}{\zeta_3^2} - \zeta_0\right)\ln(1-\zeta_3)\right] \tag{A.2}$$

where ζ_n is defined as

$$\zeta_n = \frac{\pi}{6}\rho\sum_i x_i m_i d_i^n \quad n \in \{0,1,2,3\} \tag{A.3}$$

ρ is the number density of molecules and d_i is the temperature-dependent diameter of component i, given by

$$d_i = \sigma_i\left[1 - 0.12\exp\left(-3\frac{\varepsilon_{ii}}{kT}\right)\right] \tag{A.4}$$

The radial distribution function at contact distance of two hard spheres [6, 7] is given by

$$g_{ij}^{hs} = \frac{1}{(1-\zeta_3)} + \frac{d_i d_j}{d_i + d_j}\cdot\frac{3\zeta_2}{(1-\zeta_3)^2} + \left(\frac{d_i d_j}{d_i + d_j}\right)^2\cdot\frac{2\zeta_2^2}{(1-\zeta_3)^3} \tag{A.5}$$

Dispersive contribution. The contribution due to dispersive intermolecular potentials for chain molecules is given as a sum of first- and second-order contributions [8]:

$$\tilde{a}^{disp} = \tilde{a}_1 + \tilde{a}_2 \tag{A.6}$$

$$\tilde{a}_1 = -2\pi\rho \cdot I_1(\eta, \overline{m}) \cdot \overline{m^2\varepsilon\sigma^3} \tag{A.7}$$

$$\tilde{a}_2 = -\pi\rho C_1 \cdot I_2(\eta, \overline{m}) \cdot \overline{m^2\varepsilon^2\sigma^3} \tag{A.8}$$

where η is the packing fraction defined as $\eta = \zeta_3$ and the integrals of the perturbation theory $I_1(\eta, \overline{m})$ and $I_2(\eta, \overline{m})$ are substituted by power series in density [1]:

$$I_1(\eta, \overline{m}) = \sum_{i=0}^{6} a_i(\overline{m})\eta^i \tag{A.9}$$

$$I_2(\eta, \overline{m}) = \sum_{i=0}^{6} b_i(\overline{m})\eta^i \tag{A.10}$$

The coefficients of Eq.(A.9) and (A.10) are adjusted to experimental pure component data of elongated molecules i.e. the series of n-alkanes and they are given in the work of Gross and Sadowski [1]. The abbreviated terms $\overline{m^2\varepsilon\sigma^3}$, $\overline{m^2\varepsilon^2\sigma^3}$ and C_1 are given as

$$\overline{m^2\varepsilon\sigma^3} = \sum_i \sum_j x_i x_j m_i m_j \left(\frac{\varepsilon_{ij}}{kT}\right) \sigma_{ij}^3 \tag{A.11}$$

$$\overline{m^2\varepsilon^2\sigma^3} = \sum_i \sum_j x_i x_j m_i m_j \left(\frac{\varepsilon_{ij}}{kT}\right)^2 \sigma_{ij}^3 \tag{A.12}$$

$$C_1 = kT\left(\frac{\partial\rho}{\partial P}\right)^{hc} = \overline{m} \cdot \left(1 + Z^{hc} + \rho\frac{\partial Z^{hc}}{\partial\rho}\right)^{-1} = \left(1 + \overline{m} \cdot \frac{8\eta - 2\eta^2}{(1-\eta)^4} + (1-\overline{m}) \cdot \frac{20\eta - 27\eta^2 + 12\eta^3 - 2\eta^4}{[(1-\eta)(2-\eta)]^2}\right)^{-1} \tag{A.13}$$

The dispersive parameters of a pair of unlike segments are given by the Lorentz-Berthelot combining rules:

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \tag{A.14}$$

$$\varepsilon_{ij} = (1 - k_{ij})\sqrt{\varepsilon_i\varepsilon_j} \tag{A.15}$$

where k_{ij} is the binary interaction parameter that corrects the equation-of-state for unlike intermolecular potentials.

Association contribution. The term of the residual Helmholtz energy due to associating intermolecular potentials $\tilde{a}^{assoc}$ is calculated as described in the work of Chapman et al. [5]. The radial distribution function g_{ij}^{hs} is here the one defined in Eq.(A.5). For mixtures

$$\tilde{a}^{assoc} = \sum_i x_i \left[\sum_{A_i} \left[\ln X^{A_i} - \frac{X^{A_i}}{2} \right] + \frac{1}{2} M_i \right] \tag{A.16}$$

where M_i is the number of associating sites of the molecule of component i in the mixture and summation runs over all different types of association sites (A, B, C, ...) in each molecule i. x_i is the mole fraction of all molecules of component i and the mole fraction X^{A_i} of molecules not bonded at site A_i is given by

$$X^{A_i} = \left[1 + N \cdot \sum_j \sum_{B_j} \rho_j X^{B_j} \Delta^{A_i B_j} \right]^{-1} \tag{A.17}$$

where $\sum_j$ runs over all mixture components, $\sum_{B_j}$ runs over all different sites on molecule j $(A_j, B_j, C_j, \ldots)$ and $\rho_j = x_j \cdot \rho$ is the molar density of component j. The association strength $\Delta^{A_i B_j}$ between a site of type A in molecule i and a site of type B in molecule j is given by

$$\Delta^{A_i B_j} = \sigma_{ij} g^{hs}(\sigma_{ij}) \kappa^{A_i B_j} \left[exp\left(\frac{\varepsilon^{A_i B_j}}{kT} \right) - 1 \right] \tag{A.18}$$

The cross-association parameters are calculated from pure-component association parameters using the Wolbach and Sandler [9] combining rules:

$$\varepsilon^{A_i B_j} = \frac{1}{2} \left(\varepsilon^{A_i B_i} + \varepsilon^{A_j B_j} \right) \tag{A.19}$$

$$\kappa^{A_i B_j} = \sqrt{\kappa^{A_i B_i} \kappa^{A_j B_j}} \cdot \left[\frac{2 \left(\sigma_{ii} \sigma_{jj} \right)}{\sigma_{ii} + \sigma_{jj}} \right]^3 \tag{A.20}$$

In the current work the PC-SAFT EoS for associating components is implemented using the 2B-association scheme, according to Huang and Radosz [10]. The assumptions of the 2B-association scheme for mixtures can be summarized as:

$$\begin{aligned}
&\Delta^{A_i A_i} = \Delta^{B_i B_i} = \Delta^{A_j A_j} = \Delta^{B_j B_j} = 0 \\
&\Delta^{A_i B_i} = \Delta^{A_i B_i} \neq 0 \\
&\Delta^{A_i B_j} = \Delta^{A_j B_i} \\
&X^{A_i} = X^{B_i} \\
&X^{A_i} \neq X^{A_j}
\end{aligned} \tag{A.21}$$

Contribution of quadrupole-quadrupole,dipole-dipole and dipole-quadrupole interactions. The term of the residual Helmholtz energy due to quadrupolar interactions is based on a perturbation theory of third-order, written in the Padè approximation [3, 11]:

$$\tilde{a}^{\mathrm{qq}} = \frac{\tilde{a}_2^{\mathrm{qq}}}{1 - \tilde{a}_3^{\mathrm{qq}}/\tilde{a}_2^{\mathrm{qq}}} \tag{A.22}$$

with

$$\tilde{a}_2^{\mathrm{qq}} = -\pi \left(\frac{3}{4}\right)^2 \rho \sum_i \sum_j x_i x_j \frac{\varepsilon_{ii}}{kT} \cdot \frac{\varepsilon_{jj}}{kT} \cdot \frac{\sigma_{ii}^5 \sigma_{jj}^5}{\sigma_{ij}^7} \cdot Q_i^{*2} \cdot Q_j^{*2} \cdot J_{2,ij}^{\mathrm{qq}} \tag{A.23}$$

$$\begin{aligned} \tilde{a}_3^{\mathrm{qq}} = & \frac{\pi}{3} \left(\frac{3}{4}\right)^3 \rho \sum_i \sum_j x_i x_j \left(\frac{\varepsilon_{ii}}{kT}\right)^{3/2} \left(\frac{\varepsilon_{jj}}{kT}\right)^{3/2} \cdot \frac{\sigma_{ii}^{15/2} \sigma_{jj}^{15/2}}{\sigma_{ij}^1 2} \cdot Q_i^{*3} \cdot Q_j^{*3} \cdot J_{3,ij}^{\mathrm{qq}} + \\ & \frac{4\pi^2}{3} \left(\frac{3}{4}\right)^3 \rho^2 \sum_i \sum_j \sum_k x_i x_j x_k \cdot \frac{\varepsilon_{ii}}{kT} \cdot \frac{\varepsilon_{jj}}{kT} \cdot \frac{\varepsilon_{kk}}{kT} \cdot \frac{\sigma_{ii}^5 \sigma_{jj}^5 \sigma_{kk}^5}{\sigma_{ij}^3 \sigma_{ik}^3 \sigma_{jk}^3} \cdot \\ & \cdot Q_i^{*2} \cdot Q_j^{*2} \cdot Q_k^{*2} \cdot J_{3,ijk}^{\mathrm{qq}} \end{aligned} \tag{A.24}$$

where $Q_i^{*2} = Q_i^2 / \left(m_i \sigma_{ii}^5 \varepsilon_{ii}\right)$ is the dimensionless squared quadrupole moment. $J_{2,ij}^{\mathrm{qq}}$ and $J_{3,ij}^{\mathrm{qq}}$ are the integrals over the pair correlation function of the reference fluid and $J_{3,ijk}^{\mathrm{qq}}$ is the integral over three-body correlation functions. As described in the work of Gross [3] the integral $J_{3,ij}^{\mathrm{qq}}$ is assumed to be equal to zero. The integral $J_{2,ij}^{\mathrm{qq}}$ for a pure fluid is assumed to be only a function of T and ρ and a functional of elongation (or segment number, m):

$$J_{2,ij}^{\mathrm{qq}} = \sum_{n=0}^{4} \left(a_{n,ij} + b_{n,ij} \cdot \frac{\varepsilon_{ij}}{kT}\right) \cdot \eta^n \tag{A.25}$$

The integral $J_{3,ijk}^{\mathrm{qq}}$ is assumed to be temperature-independent and it is approximated as

$$J_{3,ijk}^{\mathrm{qq}} = \sum_{n=0}^{4} c_{n,ijk} \cdot \eta^n \tag{A.26}$$

Gross [3] adjusted the model constants of Eq.(A.25) and (A.26) to molecular simulation data [12] of spherical and non-spherical molecules respectively. The adjustment of the model constants, that was carried out for varying elongations of a two-centers Lennard-Jones (2CLJ) fluid, allows an expression of quadrupolar interactions, which is valid for varying chain lengths [3].

In an analogous manner the terms of the residual Helmholtz energy due to dipole-dipole and dipole-quadrupole interactions are derived [4, 13].

References

[1] 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**, 1244 (2001).

[2] J. Gross and G. Sadowski, *Application of the Perturbed-Chain SAFT equation of state to associating systems,* Ind. Eng. Chem. Res. **41**, 5510 (2002).

[3] J. Gross, *An equation-of-state contribution for polar components: Quadrupolar molecules,* AIChE J. **51**, 2556 (2005).

[4] J. Gross and J. Vrabec, *An equation-of-state contribution for polar components: Dipolar molecules,* AIChE J. **52**, 1194 (2006).

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

[6] T. Boublik, *Hard–Sphere Equation of State,* J. Chem. Phys. **53**, 471 (1970).

[7] G. A. Mansoori, N. F. Carnahan, K. E. Starling, and T. W. Leland, *Equilibrium thermodynamic properties of the mixture of Hard Spheres,* J. Chem. Phys. **54**, 1523 (1971).

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

[9] J. P. Wolbach, , and S. I. Sandler, *Using molecular orbital calculations to describe the phase behavior of cross-associating mixtures,* Ind. Eng. Chem. Res. **37**, 2917 (1998).

[10] S. H. Huang and M. Radosz, *Equation of state for small, large, polydisperse, and associating molecules,* Ind. Eng. Chem. Res. **29**, 2284 (1990).

[11] G. Stell, J. Rasiah, and H. Narang, *Thermodynamic perturbation theory for simple polar fluids.II,* Mol. Phys. **27**, 1393 (1974).

[12] J. Stoll, J. Vrabec, H. Hasse, and J. Fischer, *Comprehensive study of the vapour-liquid equilibria of the pure two-centre Lennard-Jones plus pointquadrupole fluid,* Fluid Phase Equilib. **179**, 339 (2001).

[13] J. Vrabec and J. Gross, *Vapor–Liquid Equilibria simulation and an equation of state contribution for dipole–quadrupole interactions,* J. Phys. Chem. B **112**, 51 (2008).

B

Process model and objective function

B.1. Process model for the pre-combustion CO_2 capture

The process topology for the pre-combustion CO_2 capture (Figure B.1) considers the main stages of absorption, desorption and CO_2 compression to pipeline pressure. The number of stages in the absorber has been defined using a sensitivity analysis as described in section 2.3.1 and detailed in B.1.1. Desorption includes two sequential pressure reduction flash stages. The absorber and the two desorption flashes are modeled using a phase equilibrium stage model. Two recycle streams are considered; the H_2 rich gas stream from the first (rich gas) desorption flash ('rg') and the liquid stream of the lean solvent ('lean'). The process simulation problem is formulated as a system of non-linear equations

$$\mathbf{r}(\overline{y}) = 0 \tag{B.1}$$

where $\overline{y}$ is the vector of the iterated process variables. The equations in (B.1) are formulated on the residuals between the iterated and calculated values of the process variables for: a) the internal streams of the absorber $\mathbf{r}^{\mathrm{abs}}(\overline{y})$ and b) the process variables of the two recycle streams $\mathbf{r}^{\mathrm{rec,rg}}(\overline{y})$ and $\mathbf{r}^{\mathrm{lean}}(\overline{y})$. Finally, (B.1) includes the equation of the process specification for constant CO_2 capture rate in the absorber $\mathbf{r}^{\mathrm{spec}}(\overline{y})$.

$$\mathbf{r}(\overline{y}) := \{\mathbf{r}^{abs}(\overline{y}), \quad \mathbf{r}^{rec,rg}(\overline{y}), \quad \mathbf{r}^{rec,lean}(\overline{y}), \quad \mathbf{r}^{spec}(\overline{y})\} \tag{B.2}$$

In process optimization as well as in simultaneous process and solvent optimization, the equations in (B.2) represent the equality process constraints of the optimization problem.

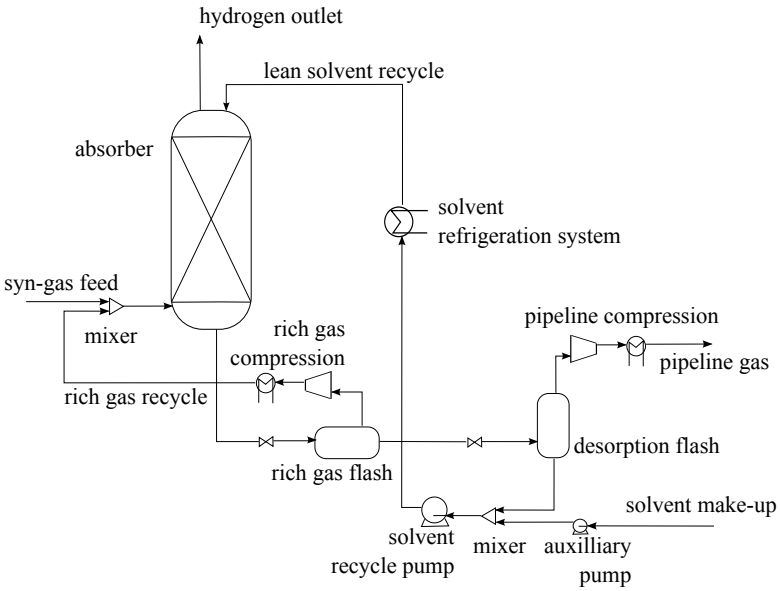


Figure B.1: Process topology of pre-combustion CO_2 capture system (adopted from Stavrou et al. [1]).

B.1.1. Absorption

For the absorber an equilibrium stage model is considered. Figure B.2 gives a general schematic representation of the j-th stage of the absorber. The so-called

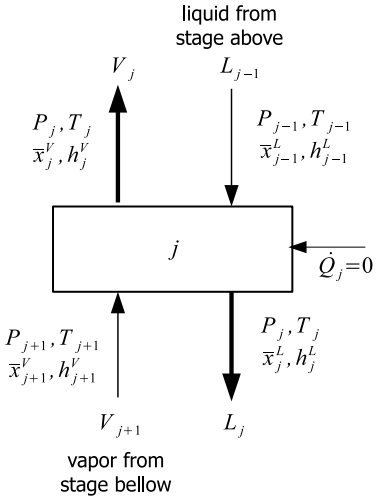


Figure B.2: General representation of an adiabatic equilibrium stage of the absorber model.

MESH equations used to model the equilibrium stages j are: the component balance equations (M), the equilibrium relations (E), the summation equations (S) and the energy balance equations (H). [2]
The summation equations for the vapor and liquid streams are

$$\sum_{i=1}^{N_c} x_{i,j}^V = 1 \quad \text{and} \quad \sum_{i=1}^{N_c} x_{i,j}^L = 1 \tag{B.3}$$

where $x_{i,j}^V$ and $x_{i,j}^L$ are the mole fractions of each system component i in the vapor V_j and the liquid L_j stream exiting stage j and N_c is the total number of components. The component balance equation for component i, for the j-th stage of the absorber writes

$$x_{i,j+1}^V \cdot V_{j+1} + x_{i,j-1}^L \cdot L_{j-1} - x_{i,j}^V \cdot V_j - x_{i,j}^L \cdot L_j = 0 \tag{B.4}$$

When Eq. (B.4) is formulated for all N_c components, the equation of the total mass balance is redundant.
For the j-th stage of the adiabatic absorber, where the heat stream $\dot{Q}_j = 0$, the energy balance equation writes

$$h_{j+1}^V \cdot V_{j+1} + h_{j-1}^L \cdot L_{j-1} - h_j^V \cdot V_j - h_j^L \cdot L_j = 0 \tag{B.5}$$

where h_j^V and h_j^L denote the molar enthalpy of the vapor and the liquid stream, respectively.
The phase equilibrium condition is solved for all components i using the isofugacity condition

$$x_{i,j}^V \cdot \varphi_{i,j}^V - x_{i,j}^L \cdot \varphi_{i,j}^L = 0 \tag{B.6}$$

where $\varphi_{i,j}^V$ and $\varphi_{i,j}^L$ are the fugacity coefficients of component i in the vapor and liquid phase, respectively.

Calculation of internal streams using a simultaneous correction method

Equilibrium stage models for multi-stage, multi-component separation problems have been extensively studied in the literature. A summary of the various solution methods for equilibrium stage models can be found in ref. [3] and [2].
In the current work, the absorber model is solved using a simultaneous correction method [2]. Simultaneous correction methods are computationally efficient and are considered as robust regarding the convergence in physical property calculations. [2] The MESH equations are solved simultaneously.
The variables used in the current implementation are slightly different from the variables shown in Figure B.2. The iterated variables are the component vapor and liquid molar flows and the temperature of the internal streams of the absorber. More specifically, for the j-th stage of the absorber (with $j = 2, \dots, N-1$) the temperatures T_j^V and T_j^L of the exiting vapor and liquid streams are iterated alongside with the component molar flows $\overline{v}_j := \{v_{1,j}, \dots, v_{N_c,j}\}$ and $\overline{l}_j := \{l_{1,j}, \dots, l_{N_c,j}\}$. The

molar enthalpies of the exiting streams are calculated as function of the pressure P_j, the iterated temperature and component molar flows:

$$h_j^V = h_j^V\left(P_j, T_j^V, \overline{x}_j^V\right) \quad \text{and} \quad h_j^L = h_j^L\left(P_j, T_j^L, \overline{x}_j^L\right) \tag{B.7}$$

$$\text{with} \quad x_{i,j}^V = \frac{v_{i,j}}{\sum_{i=1}^{N_c} v_{i,j}} \quad \text{and} \quad x_{i,j}^L = \frac{l_{i,j}}{\sum_{i=1}^{N_c} l_{i,j}}$$

For a countercurrent multi-equilibrium-stage model, such as an absorption column, one needs to iteratively solve internal variables. For the 1st stage of the absorber $(j = 1)$ we define as iterated variables the temperature T_1^L and component molar flows of the exiting liquid stream $\overline{l}_1$; for the last stage of the absorber $(j = N)$ only the temperature T_N^V and component molar flows of the exiting vapor stream $\overline{v}_N$ are iterated.
The pressure drop along the absorption column is neglected, thus

$$P_j = P^{\text{abs}} = P. \tag{B.8}$$

A schematic representation of the equilibrium stage model with the modified simulation variables is given in Figure B.3.

Phase equilibrium calculations are carried out using an adiabatic-isobaric flash algorithm (PQ-flash). For every stage j (with $j = 1, \dots, N$) a pseudo-feed stream with total molar flow F_j, composition $\overline{x}_j^{\text{F}}$ and molar enthalpy h_j^{F} is defined. The pseudo-feed stream is calculated from the iterated streams entering stage j, namely the liquid stream exiting stage $j-1$ and the vapor stream exiting stage $j+1$:

$$F_j = \sum_{i=1}^{N_c} l_{i,j-1} + \sum_{i=1}^{N_c} v_{i,j+1} \tag{B.9}$$

$$x_{i,j}^{\text{F}} = \frac{l_{i,j+1} + v_{i,j-1}}{F_j} \tag{B.10}$$

$$\text{and} \quad h_j^{\text{F}} = \frac{\sum_{i=1}^{N_c} l_{i,j-1}}{F_j} \cdot h_{j-1}^L + \frac{\sum_{i=1}^{N_c} v_{i,j+1}}{F_j} \cdot h_{j-1}^V. \tag{B.11}$$

For the 1st stage the pseudo-feed stream is calculated using Eq.(B.9) to (B.11) with the iterated component molar flows $\overline{v}_2$ of the vapor stream exiting stage 2 and the iterated component molar flows $\overline{l}^{\text{lean}}$ of the lean solvent recycle that enters the absorber in stage 1. The pseudo-feed stream for the last stage N is calculated using Eq.(B.9) to (B.11) with the iterated component molar flows $\overline{l}_{N-1}$ of the liquid stream exiting stage $N-1$ and the iterated component molar flows of the rich gas recycle stream $\overline{v}^{\text{rg}}$ entering the absorber in stage N.
Using the pseudo-feed stream defined in Eq.(B.9) to (B.11), phase equilibrium calculations deliver for every stage j the calculated values of the component molar flows $\overline{\hat{v}}_j := \{\hat{v}_{1,j}, \dots, \hat{v}_{N_c,j}\}$ and $\overline{\hat{l}}_j := \{\hat{l}_{1,j}, \dots, \hat{l}_{N_c,j}\}$, the temperatures $\hat{T}_j^V = \hat{T}_j^L = \hat{T}_j$

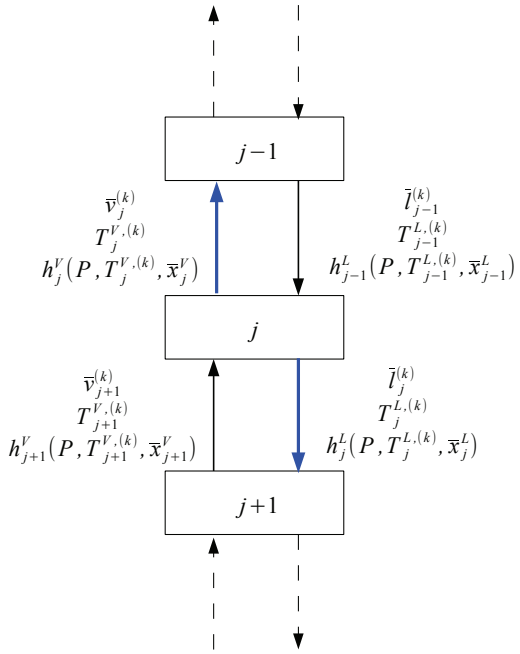


Figure B.3: In the k-th iteration for stage j of the absorber (with $j = 2, \dots, N-1$) the component molar flows $\bar{v}_j^{(k)}$ and $\bar{l}_j^{(k)}$ and the temperatures $T_j^{V,(k)}$ and $T_j^{L,(k)}$ of the vapor and the liquid exiting streams are iterated, respectively.

and thereby the molar enthalpies $\hat{h}_j^V$ and $\hat{h}_j^L$ of the exiting vapor and liquid streams (Figure B.4). For the j-th stage of the absorber (with $j = 2, \dots, N-1$) the equations of the residuals between iterated variables and their calculated values then write

$$\mathbf{r}_j^{\text{abs}} = \begin{pmatrix} h_j^V - \hat{h}_j^V \\ h_j^L - \hat{h}_j^L \\ v_{1,j} - \hat{v}_{1,j} \\ v_{2,j} - \hat{v}_{2,j} \\ \vdots \\ v_{N_c,j} - \hat{v}_{N_c,j} \\ l_{1,j} - \hat{l}_{1,j} \\ l_{2,j} - \hat{l}_{2,j} \\ \vdots \\ l_{N_c,j} - \hat{l}_{N_c,j} \end{pmatrix}, \quad \text{for} \quad j \neq 1 \quad \text{and} \quad j \neq N \tag{B.12}$$

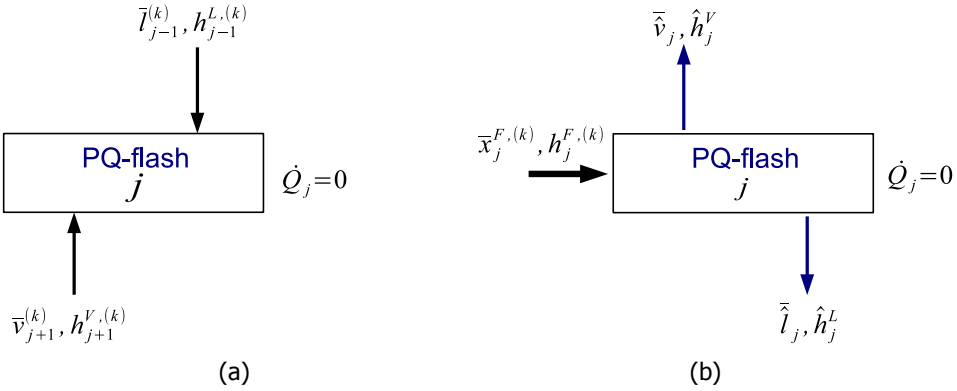


Figure B.4: Generic representation of the input and output variables of the adiabatic-isobaric flash (PQ-Flash) calculations for stage j. (a): The actual input variables for PQ-flash are the iterated component molar flows and molar enthalpies (function of the iterated temperature) of the streams entering the equilibrium stage. Stage j is considered adiabatic ($\dot{Q}_j = 0$). (b): A pseudo-feed stream is defined for stage j as function of the iterated variables of the liquid and vapor streams entering stage j, as presented in (a). Phase equilibrium calculations with PQ-flash give the component molar flows and the molar enthalpies of the vapor and liquid streams exiting stage j in equilibrium.

For the 1st stage

$$\mathbf{r}_1^{\text{abs}} = \begin{pmatrix} h_1^L - \hat{h}_1^L \\ l_{1,1} - \hat{l}_{1,1} \\ l_{2,1} - \hat{l}_{2,1} \\ \vdots \\ l_{N_c,1} - \hat{l}_{N_c,1} \end{pmatrix} \tag{B.13}$$

while for the last stage N

$$\mathbf{r}_N^{\text{abs}} = \begin{pmatrix} h_N^V - \hat{h}_N^V \\ v_{1,N} - \hat{v}_{1,N} \\ v_{2,N} - \hat{v}_{2,N} \\ \vdots \\ v_{N_c,N} - \hat{v}_{N_c,N} \end{pmatrix} \tag{B.14}$$

The residual equations for the internal streams of the absorber formulate the square $2(N_c + 1)(N - 1)$ equation system

$$\mathbf{r}^{\text{abs}}(\bar{y}) := \{\mathbf{r}_1^{\text{abs}}, \quad \mathbf{r}_2^{\text{abs}}, \quad \dots, \quad \mathbf{r}_{N-1}^{\text{abs}}, \quad \mathbf{r}_N^{\text{abs}}\} \tag{B.15}$$

$\mathbf{r}^{\text{abs}}(\bar{y}^{\text{abs}})$ is a subset of the equation system in (B.2).

Phase equilibrium calculations

For phase equilibrium calculations in the absorber the adiabatic-isobaric flash problem (PQ-flash) is solved for every stage j (Figure B.4). The isothermal-isobaric two-phase flash problem (PT-flash), which is a sub-problem of PQ-flash, is solved using the successive substitution Rachford-Rice algorithm (according to Michelsen and Mollerup) [4]. The required caloric properties (residual part) and fugacities in the PT-flash and PQ-flash problems are calculated with the PC-SAFT EoS.

Process specifications

For the examined pre-combustion CO_2 capture system a constant CO_2 capture rate in the absorber of 90% is specified. In this work, the CO_2 capture rate α_{CO_2} is defined as CO_2 captured in the absorber with respect to the CO_2 molar flow in the syngas stream $v_{\mathrm{CO}_2}^{\mathrm{sg}}$

$$\alpha_{\mathrm{CO}_2} = 1 - \frac{\hat{v}_{\mathrm{CO}_2,1}}{v_{\mathrm{CO}_2}^{\mathrm{sg}}} \tag{B.16}$$

where $\hat{v}_{\mathrm{CO}_2,1}$ is the (calculated) value of the CO_2 molar flow in the gas stream exiting the 1^{st} stage of the absorber. The residual equation $\mathbf{r}^{spec}(\overline{y})$ then writes

$$\mathbf{r}^{spec}(\overline{y}) = \left(1 - \alpha_{\mathrm{CO}_2}\right) \cdot v_{\mathrm{CO}_2}^{\mathrm{sg}} - \hat{v}_{\mathrm{CO}_2,1} \tag{B.17}$$

The independent variable for the specification in Eq.(B.17) is the molar flow of the solvent make-up stream $l_{\mathrm{s}}^{\mathrm{MU}}$.

Estimation of the number of absorption stages

Note: Parts of this section have been published as Supporting Information to the work of Stavrou et al. [1].

For the simultaneous process and solvent optimization in the CoMT step an economic objective function is minimized. The degrees of freedom for the optimization are the pressure levels in the rich gas and desorption flash P^{rg} and P^{des}, the temperature of the lean solvent T^{lean} and the PC-SAFT molecular parameters of the solvent $\overline{p} = \{m, \sigma, \varepsilon/k, \mu\}$. In the current work the number of stages in the absorber is not considered as degree of freedom for the CoMT optimization problem (Eq. (2.2)), because such an optimization would require consideration of investment costs as part of the objective function. In order to define a reasonable number of equilibrium stages, a sensitivity analysis was carried out and the influence of the number of absorption stages on the CoMT optimization results was examined. Figures B.5 and B.6 show that the number of absorption stages has a low impact on the optimal molecular parameters of the solvent calculated in the CoMT step. Figures B.7 and B.8 show that increasing the number of absorption stages above 7 will only have a moderate effect on the optimal process conditions and the optimal value of the objective function. The absorber is therefore defined with a fixed number of $N_{abs} = 7$ equilibrium stages.

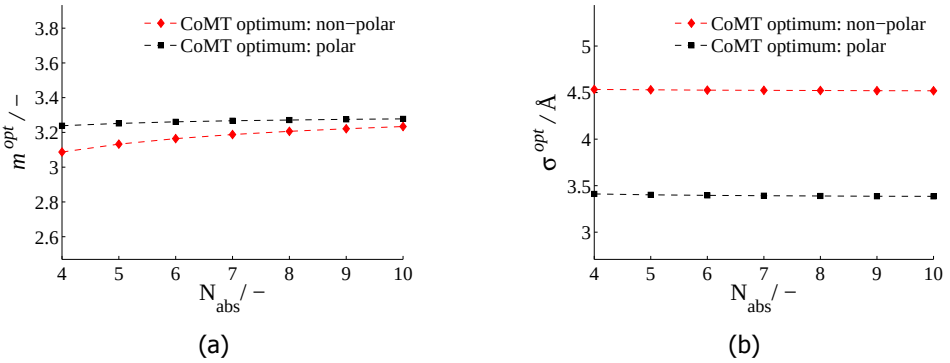


Figure B.5: Sensitivity of the optimal molecular parameters of the solvent: (a) segment number m^{opt} and (b) segment diameter σ^{opt} for varying number of absorption stages N_{abs} for non-polar (red diamond symbols) and polar molecules (black square symbols).

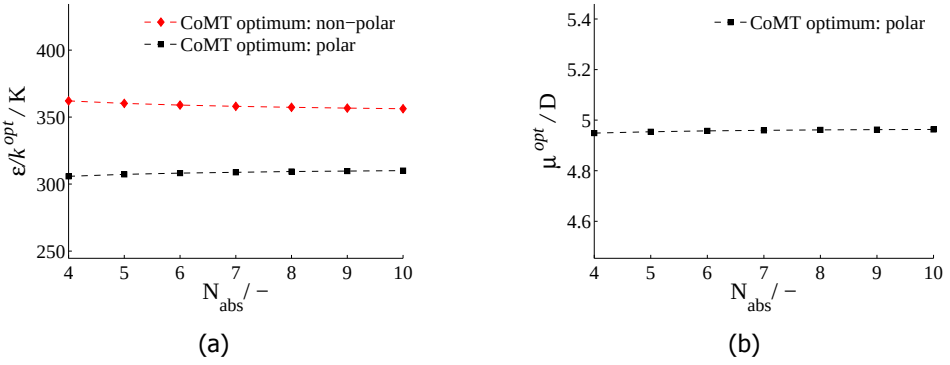


Figure B.6: Sensitivity of the optimal molecular parameters of the solvent: (a) dispersion energy ε/k^{opt} and (b) dipole moment μ^{opt} for varying number of absorption stages N_{abs} for non-polar (red diamond symbols) and polar molecules (black square symbols).

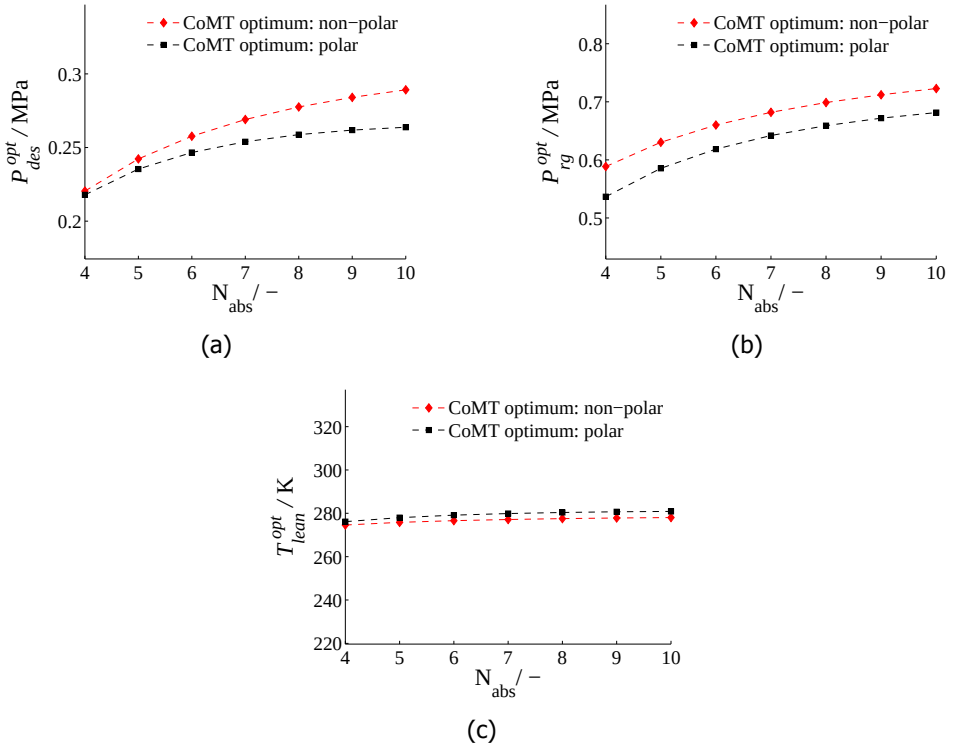


Figure B.7: Sensitivity of the optimal process conditions: (a) optimal pressure of the second desorption flash P_{des}^{opt}, (b) optimal pressure of the hydrogen-rich gas desorption flash P_{rg}^{opt} and (c) optimal temperature of the lean solvent T_{lean}^{opt} for varying number of absorption stages N_{abs} for non-polar (red diamond symbols) and polar molecules (black square symbols).

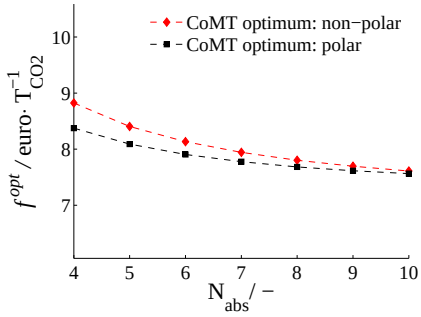


Figure B.8: Sensitivity of the optimal value of the objective function for varying number of absorption stages N_{abs} in the absorber for non-polar (red diamond symbols) and polar molecules (black square symbols).

B.1.2. Desorption

The desorption flash units are modeled using the equilibrium stage model described in Section B.1.1. For each flash unit the phase equilibrium problem is solved for the adiabatic-isobaric flash using the PQ-flash algorithm (Figure B.4). The solution of the phase equilibrium problem for the desorption flashes is, however, more simple than in the case of the absorber. The phase equilibrium problems for the two desorption flashes are solved independently and do not require the simultaneous convergence of internal variables.

B.1.3. Recycle streams

The examined process topology includes two recycle streams entering the absorber: the (vapor) recycle stream returning from the first desorption stage (rich gas flash) after compression and cooling (hereafter noted with the superscript 'rg') and the (liquid) recycle stream of the lean solvent after pumping and cooling (hereafter noted with the superscript 'lean') (Figure B.1). For both recycle streams the iterated variables are the component molar flows and the temperature

$$\overline{y}^{\mathrm{rg}} := \{T^{rg}, \overline{v}^{rg}\} \quad \text{and} \quad \overline{y}^{\mathrm{lean}} := \{T^{lean}, \overline{l}^{lean}\}$$

The molar enthalpies of the recycle streams are calculated as function of the pressure and the iterated temperature and component molar flows

$$h^{\mathrm{rg}} = h^{\mathrm{rg}}\left(P^{\mathrm{abs}}, T^{\mathrm{rg}}, \overline{v}^{\mathrm{rg}}\right) \quad \text{and} \quad h^{\mathrm{lean}} = h^{\mathrm{lean}}\left(P^{\mathrm{abs}}, T^{\mathrm{lean}}, \overline{v}^{\mathrm{lean}}\right) \tag{B.18}$$

In addition, calculated values for the component molar flows and the molar enthalpy of the rich gas recycle are obtained through process calculations for the units following the absorber. The equations of the residuals between the iterated and calculated values of the variables form the equation subsystem $\mathbf{r}^{\mathrm{rec,rg}}(\overline{y})$ in (B.2)

$$\mathbf{r}^{\mathrm{rec,rg}} = \begin{pmatrix} h^{\mathrm{rg}} - \hat{h}^{\mathrm{rg}} \\ v_1^{\mathrm{rg}} - \hat{v}_1^{\mathrm{rg}} \\ v_2^{\mathrm{rg}} - \hat{v}_2^{\mathrm{rg}} \\ \vdots \\ v_{N_c}^{\mathrm{rg}} - \hat{v}_{N_c}^{\mathrm{rg}} \end{pmatrix} \tag{B.19}$$

In a similar fashion, the equation subsystem $\mathbf{r}^{\mathrm{rec,lean}}(\overline{y})$ is defined for the recycle stream of the lean solvent entering the absorber at stage 1 after pumping and cooling

$$\mathbf{r}^{\mathrm{rec,lean}} = \begin{pmatrix} h^{\mathrm{lean}} - \hat{h}^{\mathrm{lean}} \\ l_1^{\mathrm{lean}} - \hat{l}_1^{\mathrm{lean}} \\ l_2^{\mathrm{lean}} - \hat{l}_2^{\mathrm{lean}} \\ \vdots \\ l_{N_c}^{\mathrm{lean}} - \hat{l}_{N_c}^{\mathrm{lean}} \end{pmatrix} \tag{B.20}$$

B.1.4. Pressure change units

Compressors

The vapor stream exiting the rich gas flash is recycled to the absorber and is compressed to the absorption pressure P^{abs}. The vapor stream that exits the desorption flash is being compressed to the pipeline pressure P^{ppl}.

In both cases polytropic, staged compressors are considered according to the model described in ref. [5]. The duty required for the gas compression $\dot{W}_{\text{comp}}$ (here defined as a positive stream) is

$$\dot{W}_{\text{comp}} = \left|\frac{1}{\eta_{\text{comp}}} N_{\text{comp}} \cdot P_{\text{in}} \cdot \dot{V}_{\text{in}} \cdot \left[1 - r_{\text{comp}}^{\left(\frac{n}{n-1}\right)}\right] \cdot \left(\frac{n}{n-1}\right)\right| \tag{B.21}$$

with η_{comp} the efficiency of the compressor, n the polytropic exponent, N_{comp} the number of compression stages, P_{in} the inlet pressure, $\dot{V}_{\text{in}}$ the volumetric molar flow of the stream entering the compressor, and r_{comp} the pressure ratio for minimum work for N_{comp} stages.

Regarding compression both vapor streams are treated as ideal gas streams. Thereby, the polytropic exponent n is assumed to be approximately equal to the heat capacity ratio $\gamma = \frac{c_p}{c_v}$. The heat capacity ratio γ is held constant equal to 1.3, which is a fair assumption for gas streams containing mainly CO_2 and H_2 in the temperature range between 250 and $400\,\text{K}$ [6].

The pressure ratio r_{comp} is given by:

$$r_{\text{comp}} = \left(\frac{P_{\text{out}}}{P_{\text{in}}}\right)^{1/N_{\text{comp}}} \tag{B.22}$$

where P_{out} is the outlet pressure of the compressed stream (P^{abs} and P^{ppl} respectively). The inlet pressures P_{in} of both compressors (P^{rg} for the compressor after the rich gas flash and P^{des} for compressor after the desorption flash) are degrees of freedom in the simultaneous process and solvent optimization problem.

Pumps

The lean solvent stream is recycled to the absorber with a recirculating pump. The duty required by the pump $\dot{W}_{\text{pump}}$ is

$$\dot{W}_{\text{pump}} = \frac{1}{\eta_{\text{pump}}} \cdot \dot{V}^{\text{RL}} \cdot (P_{\text{out}} - P_{\text{in}}) \tag{B.23}$$

with η_{pump} the efficiency of the pump, $\dot{V}^{\text{lean}}$ the volumetric molar flow of the lean solvent stream, P_{out} the pressure in the absorber P^{abs} and P_{in} the pressure of the desorption flash P^{des}.

The liquid stream exiting the desorption flash is mixed with the fresh solvent make-up to the lean solvent stream. The work required by the auxiliary pump of the make-up stream (Figure B.1) is calculated according to Eq.(B.23) for $P_{\text{in}} = 0.1013\,\text{MPa}$ and $P_{\text{out}} = P^{\text{des}}$.

B.1.5. Heat exchange units

In the examined process topology heat exchangers are considered for the cooling of gas and liquid streams after compression and pumping respectively (Figure B.1).

Cooling of compressed gas streams

First, the rich gas recycle is compressed to the absorption pressure P^{abs} and mixed with the syngas stream. After compression and before mixing, the rich gas recycle has to be cooled to the temperature of the syngas stream $T^{\text{sg}} = 313\,\text{K}$. Second, the vapor stream exiting the desorption flash is compressed to the pipeline pressure and has to be cooled to the specified temperature for the pipeline gas T^{ppl} (here $T^{\text{ppl}} = 313\,\text{K}$). In both cases simple coolers with water as cooling medium are considered. No pressure drop is assumed. The required cooling duty $\dot{Q}_{\text{cool,comp}}$ is

$$\dot{Q}_{\text{cool,comp}} = \dot{n} \cdot (h_{\text{cold}} - h_{\text{hot}}) \tag{B.24}$$

with $\dot{n}$ as the molar flow of the compressed gas stream.
For the cooling of the compressed rich gas stream

$$h_{\text{cold}} = h_{\text{cold}}\left(T^{\text{sg}}, P^{\text{abs}}, \overline{x}^{\text{v,RG}}\right)$$
$$h_{\text{hot}} = h_{\text{hot}}\left(T^{\text{comp}}_{\text{out}}, P^{\text{abs}}, \overline{x}^{\text{v,RG}}\right)$$

while for the cooling of the compressed gas stream after desorption

$$h_{\text{cold}} = h_{\text{cold}}\left(T^{\text{ppl}}, P^{\text{ppl}}, \overline{x}^{\text{v,des}}\right)$$
$$h_{\text{hot}} = h_{\text{hot}}\left(T^{\text{comp}}_{\text{out}}, P^{\text{ppl}}, \overline{x}^{\text{v,des}}\right).$$

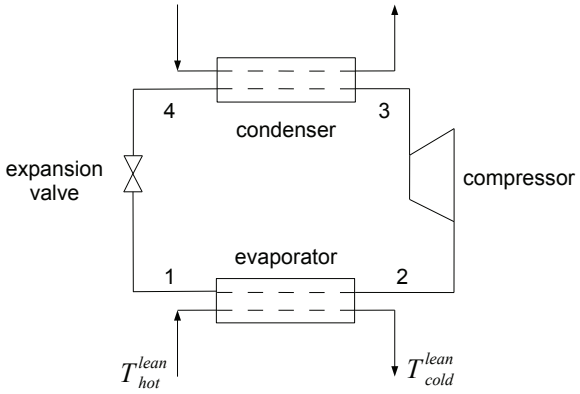
The required molar flow of the cooling medium $\dot{n}^{w}_{\text{cool}}$ (here liquid water) can be calculated as

$$\dot{n}^{w}_{\text{cool}} = \frac{\dot{Q}_{\text{cool,comp}}}{\overline{c}^{w}_{p} \cdot \Delta T} \tag{B.25}$$

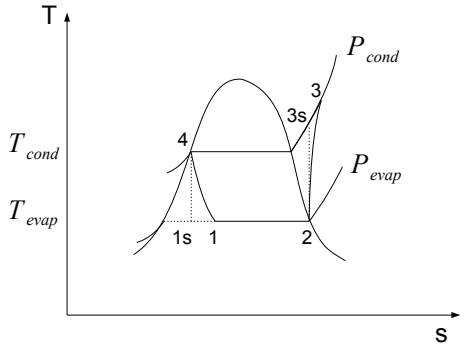
considering a minimum temperature difference ΔT of $10\,\text{K}$ for the heat transfer and an average specific heat capacity for water $\overline{c}^{w}_{p} = 0.5 \cdot \left(c^{w}_{p}(283\,\text{K}) + c^{w}_{p}(293\,\text{K})\right)$. In the current implementation, the values of $c^{w}_{p}(T)$ were obtained from the DIPPR database [6].

Lean solvent cooling

The temperature of the lean solvent entering the absorber T^{lean} is an additional degree of freedom in the process and solvent optimization problem. For the cooling of the lean solvent to the temperature T^{lean} a vapor-compression refrigeration (VCR) system is considered. Here a simple VCR cycle is implemented (Figure B.9). Simple VCR cycles can provide cooling as low as typically $233\,\text{K}$ [5] and possess an adequate operating range for the requirements of the optimization problem ($T^{\text{lean}}_{lb} = 253\,\text{K}$). The cooling of the lean solvent is provided in the evaporator. The required duty



(a)



(b)

Figure B.9: Vapor-compression refrigeration cycle (VCR) [5, 7]. (a): Schematic of a simple VCR system. (b): Temperature-entropy diagram for the VCR system presented in (a). The vapor-liquid mixture of the refrigerant is evaporated at T_{evap} (process '1-2') and provides the required duty $\dot{Q}_{\mathrm{cool,pump}}$ to cool the lean solvent from $T_{\mathrm{hot}}^{\mathrm{lean}}$ to $T_{\mathrm{cold}}^{\mathrm{lean}} = T^{\mathrm{lean}}$. The refrigerant (saturated vapor at state '2') is compressed from P_{evap} to P_{cond} (process '2-3'). The refrigerant enters the condenser as superheated vapor (state '3'), where it is cooled and condensed at T_{cond} (process '3-4'). The (saturated) liquid refrigerant (state '4') is isenthalpicaly expanded to the pressure P_{evap} (process '4-1').

$\dot{Q}_{\mathrm{cool,pump}}$ is

$$\dot{Q}_{\mathrm{cool,pump}} = \dot{L}^{\mathrm{lean}} \cdot \left(h_{\mathrm{cold}}^{\mathrm{lean}} - h_{\mathrm{hot}}^{\mathrm{lean}}\right) \tag{B.26}$$

where $\dot{L}^{\mathrm{lean}}$ is the molar flow of the lean solvent stream, $h_{\mathrm{hot}}^{\mathrm{lean}}$ is the molar enthalpy of the lean solvent stream entering the refrigeration system and $h_{\mathrm{cold}}^{\mathrm{lean}}$ is the molar enthalpy of the cooled lean solvent stream. $h_{\mathrm{hot}}^{\mathrm{lean}}$ is given by the energy balance for the pump

$$h_{\mathrm{hot}}^{\mathrm{lean}} = h_{\mathrm{pump,in}}^{\mathrm{lean}} + \frac{\dot{W}_{\mathrm{pump}}}{\dot{L}^{\mathrm{lean}}}. \tag{B.27}$$

$h_{\text{cold}}^{\text{lean}}$ is calculated using the PC-SAFT EoS (with QSPR estimations for the ideal gas specific heat capacity of the solvent $c_{p,s}^{ig}$) as

$$h_{\text{cold}}^{\text{lean}} = h^{\text{lean}}\left(T_{\text{cold}}^{\text{lean}}, P^{\text{abs}}, \overline{x}^{\text{lean}}\right)$$

with $T_{\text{cold}}^{\text{lean}} = T^{\text{lean}}$ and $\overline{x}^{\text{lean}}$ the vector of component mole fractions of the lean solvent stream exiting the pump.
The compression duty $\dot{W}_{\text{refr}}^{\text{rev}}$ for the reversible refrigeration process (process '1s-2-3s-4' in Figure B.9b) is

$$\dot{W}_{\text{refr}}^{\text{rev}} = \dot{Q}_{\text{cool,pump}}\left(\frac{T_{\text{cond}}}{T_{\text{evap}}} - 1\right) \tag{B.28}$$

where T_{cond} and T_{evap} are the condensation and evaporation temperatures of the refrigerant. The coefficient of performance of the reversible process COP_{rev} [5, 7] is

$$\text{COP}_{\text{rev}} = \frac{\dot{Q}_{\text{cool,pump}}}{\dot{W}_{\text{refr}}^{\text{rev}}}. \tag{B.29}$$

Accordingly, for the real process (process '1-2-3-4' in Figure B.9b) the coefficient of performance COP is

$$\text{COP} = \frac{\dot{Q}_{\text{cool,pump}}}{\dot{W}_{\text{refr}}} \tag{B.30}$$

where $\dot{W}_{\text{refr}}$ is the real compression duty. Here a constant ratio $\frac{\text{COP}}{\text{COP}_{\text{rev}}}$ equal to 0.25 [8] is assumed. The real compression duty $\dot{W}_{\text{refr}}$ is then calculated as

$$\dot{W}_{\text{refr}} = \dot{W}_{\text{refr}}^{\text{rev}} \cdot \left(\frac{\text{COP}}{\text{COP}_{\text{rev}}}\right)^{-1}. \tag{B.31}$$

The temperature T_{evap} in Eq.(B.28) is defined by the temperature of the lean solvent T^{lean}, considering a minimum temperature difference ΔT_{min} of $10\,\text{K}$ in the evaporator:

$$T_{\text{evap}} = T^{\text{lean}} - \Delta T_{\text{min}}. \tag{B.32}$$

In the condenser, the temperature T_{cond} at which heat is transferred from the refrigerant is set to $323.15\,\text{K}$. The constant value of T_{cond} is estimated as follows:
For $T_{\text{cond}} \geq T_{\text{evap}}$ should hold:

$$T_{\text{cond}} \geq T^{\text{lean}} - \Delta T_{\text{min}}. \tag{B.33}$$

The temperature T^{lean} is a degree of freedom in the process and solvent optimization problem. Thus,

$$(T_{\text{cond}})_{\text{min}} = T_{\text{ub}}^{\text{lean}} - \Delta T_{\text{min}} \tag{B.34}$$

where $T_{\text{ub}}^{\text{lean}}$ is the upper bound of T^{lean} during the optimization, equal to $333.15\,\text{K}$.

B.2. Definition of the objective function

Note: Parts of this Appendix have been published as Supporting Information to the work of Stavrou et al. [1].

The total specific cost per ton of captured CO_2 accounts for the specific cost contributions as: electricity demand, hydrogen loss, fresh solvent consumption and demand in cooling utilities, according to

$$f_{\text{tot}} = f_{\text{electr.demand}} + f_{\text{H2 loss}} + f_{\text{solvent loss}} + f_{\text{utilities}} \tag{B.35}$$

Cost term for the electric energy demand, $f_{\text{electr.demand}}$. The specific cost associated with the electricity demand $f_{\text{electr.demand}}$ accounts for the costs of: the pump with duty $\dot{W}_{\text{pump}}$, the CO_2 compression to pipeline pressure with duty $\dot{W}_{\text{comp,ppl}}$ and the rich gas compression with duty $\dot{W}_{\text{comp,rg}}$ (in MW):

$$f_{\text{electr.demand}} = \frac{\left(\dot{W}_{\text{comp,ppl}} + \dot{W}_{\text{comp,rg}} + \dot{W}_{\text{pump}}\right) \cdot c_{\text{el}}}{\dot{m}_{\text{CO}_2,\text{capt}}} \tag{B.36}$$

where c_{el} is the electricity production price [9] equal to $55\,€/\text{MWh}$ and $\dot{m}_{\text{CO}_2,\text{capt}}$ is the mass flow rate of CO_2 captured in absorber (in T/s).
The mass flow of captured CO_2 $\dot{m}_{\text{CO}_2,\text{capt}}$ is expressed in terms of the CO_2-capture rate α_{CO_2} (Eq.(B.16)) from the absorber and the molar flow rate of CO_2 in the syngas $\dot{n}_{\text{CO}_2,\text{sg}}$, as

$$\dot{m}_{\text{CO}_2,\text{capt}} = \alpha_{\text{CO}_2} \cdot \dot{n}_{\text{CO}_2,\text{sg}} \cdot M_{\text{CO}_2}. \tag{B.37}$$

Cost term for hydrogen loss, $f_{\text{H}_2\text{loss}}$. The specific cost associated with the hydrogen unintentionally captured and stored with the CO_2-rich stream is evaluated according to lost opportunity to produce electricity from the hydrogen with

$$f_{\text{H}_2\text{loss}} = \frac{\dot{n}_{\text{H2}}^{\text{ppl}} \cdot \Delta H_{H_2}^{\text{comb}}}{\dot{m}_{\text{CO}_2,\text{capt}}} \cdot \eta_{\text{el}} \cdot c_{\text{el}} \tag{B.38}$$

where $\Delta H_{H_2}^{\text{comb}}$ is the enthalpy of combustion for hydrogen, $\dot{n}_{\text{H2}}^{\text{ppl}}$ is the molar flow of hydrogen in the vapor stream exiting the desorption flash (in kmol/sec) and η_{el} is the efficiency of power production for the combined cycle, once hydrogen is present at high pressure, set equal to $65\,\%$.[1]

Cost term for solvent loss, $f_{\text{solventloss}}$. The specific cost of solvent loss $f_{\text{solventloss}}$ measures the required fresh solvent make-up stream $\dot{n}^{\text{MU}}$ (in kmol/sec) and accounts for the solvent loss both in the absorption and desorption stage:

$$f_{\text{solventloss}} = \frac{\dot{n}^{\text{MU}} \cdot \text{v}_{\text{S}}}{\dot{m}_{\text{CO}_2,\text{capt}}} \cdot c_{\text{bulk}} \tag{B.39}$$

[1] Note that η_{el} is set equal to $61\,\%$ for the case study presented in Chapter 4.

The specific molar volume of the solvent v_S is calculated with the PC-SAFT EoS at constant temperature ($298.15\,\mathrm{K}$) and pressure ($0.1013\,\mathrm{MPa}$). For the fresh solvent cost we consider an average price of high molecular weight chemicals of $4{,}000\,€/\mathrm{T}$. We estimate the price for bulk chemicals c_{bulk} equal to $4{,}040\,€/\mathrm{m^3}$ using the molar volume of DEPG-5 by $298.15\,\mathrm{K}$ and $0.1013\,\mathrm{MPa}$.

Cost term for the utilities demand, $f_{\mathrm{utilities}}$. The specific cost of utilities $f_{\mathrm{utilities}}$ takes into account the electric energy demand for the operation of the cooling systems [5], with three contributions from the duties $\dot{W}_{\mathrm{aux-pump,rg}}$, $\dot{W}_{\mathrm{aux-pump,ppl}}$ and $\dot{W}_{\mathrm{refr}}$ according to

$$f_{\mathrm{utilities}} = \frac{\left(\dot{W}_{\mathrm{aux-pump,rg}} + \dot{W}_{\mathrm{aux-pump,ppl}} + \dot{W}_{\mathrm{refr}}\right)}{\dot{m}_{\mathrm{CO_2,capt}}} \cdot c_{\mathrm{el}} \tag{B.40}$$

The coolers, after the rich-gas compression and after the pipeline compression (section B.1.5) are operated with cooling water, so that the cost contribution is due to the electrical duty of the cooling-water pumps, $\dot{W}_{\mathrm{aux-pump,rg}}$ and $\dot{W}_{\mathrm{aux-pump,ppl}}$, respectively. The required molar flow of cooling water is given by Eq.(B.25). The duty $\dot{W}_{\mathrm{refr}}$ required by the vapor-compression system for the lean solvent cooling is given by Eq.(B.26) to (B.31).

References

[1] M. Stavrou, M. Lampe, A. Bardow, and J. Gross, *Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for CO_2 capture,* Ind. Eng. Chem. Res. **53**, 18029 (2014).

[2] H. A. Kooijman and R. Taylor, *The ChemSep book* (public domain, 2006).

[3] J. D. Seader and E. J. Henley, *Separation process principles* (John Wiley & Sons, Inc., 2006).

[4] M. L. Michelsen and J. M. Möllerup, *Thermodynamic models: Fundamentals & Computational aspects* (Tie-Line Publications, 2007).

[5] R. Smith, *Chemical process design and integration* (John Wiley, 2005).

[6] AIChE-DECHEMA, *Pure Component Database DIPPR-801*, Tech. Rep. (AIChE-DECHEMA, 2011).

[7] J. M. Smith, H. C. Van Ness, and M. M. Abbott, *Introduction to chemical Engineering thermodynamics* (McGraw-Hill, 2005).

[8] A. Arora and S. Kaushik, *Theoretical analysis of a vapour compression refrigeration system with R502, R404A and R507A,* Int. J. Refrig. **31**, 998 (2008).

[9] *Second Strategic Energy Review: Energy Sources, production costs and performance of technologies for power generation, heating and transport* (Commission of the European Communities, Brussels, 2008).

C

Database of candidate components for solvent selection

This Appendix has been published as Supporting Information to the work of Stavrou et al. [1].

Table C.1 presents the 168 non-polar, non-associating components considered in our study as the mapping database of non-polar components. Table C.2 presents the 444 polar, non-associating components considered as the mapping database of polar components. For each component in the list we give its CAS number and name according to the DIPPR database [2].

Table C.1: List of the 168 non-polar, non-associating components comprising the database of non-polar components. Each component is denoted with its CAS number and name [2].

CAS Number	Name	CAS Number	Name
109660	n-PENTANE	291645	CYCLOHEPTANE
110543	n-HEXANE	292648	CYCLOOCTANE
107835	2-METHYLPENTANE	1072055	2,6-DIMETHYLHEPTANE
96140	3-METHYLPENTANE	6747323	2,2-DIMETHYL-3-ETHYLPENTANE
75832	2,2-DIMETHYLBUTANE	1068877	2,4-DIMETHYL-3-ETHYLPENTANE
79298	2,3-DIMETHYLBUTANE	646048	trans-2-PENTENE
142825	n-HEPTANE	4050457	trans-2-HEXENE
591764	2-METHYLHEXANE	3269528	trans-3-HEXENE
589344	3-METHYLHEXANE	3769231	4-METHYL-1-HEXENE
617787	3-ETHYLPENTANE	760214	2-ETHYL-1-BUTENE
590352	2,2-DIMETHYLPENTANE	563780	2,3-DIMETHYL-1-BUTENE
108087	2,4-DIMETHYLPENTANE	558372	3,3-DIMETHYL-1-BUTENE
562492	3,3-DIMETHYLPENTANE	563791	2,3-DIMETHYL-2-BUTENE
464062	2,2,3-TRIMETHYLBUTANE	3404715	2-ETHYL-1-PENTENE
111659	n-OCTANE	4686136	trans-2-HEPTENE

592278	2-METHYLHEPTANE
589811	3-METHYLHEPTANE
589537	4-METHYLHEPTANE
590738	2,2-DIMETHYLHEXANE
592132	2,5-DIMETHYLHEXANE
563166	3,3-DIMETHYLHEXANE
609267	2-METHYL-3-ETHYLPENTANE
1067089	3-METHYL-3-ETHYLPENTANE
564023	2,2,3-TRIMETHYLPENTANE
540841	2,2,4-TRIMETHYLPENTANE
560214	2,3,3-TRIMETHYLPENTANE
565753	2,3,4-TRIMETHYLPENTANE
111842	n-NONANE
3522949	2,2,5-TRIMETHYLHEXANE
7154805	3,3,5-TRIMETHYLHEPTANE
6747301	2,4,4-TRIMETHYLHEXANE
1067205	3,3-DIETHYLPENTANE
7154792	2,2,3,3-TETRAMETHYLPENTANE
1186534	2,2,3,4-TETRAMETHYLPENTANE
1070877	2,2,4,4-TETRAMETHYLPENTANE
6747389	2,3,3,4-TETRAMETHYLPENTANE
111013	SQUALANE
124185	n-DECANE
3475815	2,2,3,3-TETRAMETHYLHEXANE
1071814	2,2,5,5-TETRAMETHYLHEXANE
3178221	tert-BUTYLCYCLOHEXANE
1120214	n-UNDECANE
112403	n-DODECANE
629505	n-TRIDECANE
629594	n-TETRADECANE
629629	n-PENTADECANE
544763	n-HEXADECANE
629787	n-HEPTADECANE
5911046	3-METHYLNONANE
871830	2-METHYLNONANE
7301949	4-METHYLNONANE
5869859	5-METHYLNONANE
3221612	2-METHYLOCTANE
2216333	3-METHYLOCTANE
2216344	4-METHYLOCTANE
5869804	3-ETHYLHEPTANE
1071267	2,2-DIMETHYLHEPTANE
287923	CYCLOPENTANE
96377	METHYLCYCLOPENTANE
1640897	ETHYLCYCLOPENTANE
1638262	1,1-DIMETHYLCYCLOPENTANE
1192183	cis-1,2-DIMETHYLCYCLOPENTANE
822504	trans-1,2-DIMETHYLCYCLOPENTANE
2532583	cis-1,3-DIMETHYLCYCLOPENTANE
2040962	n-PROPYLCYCLOPENTANE
3875512	ISOPROPYLCYCLOPENTANE
6747505	1-METHYL-1-ETHYLCYCLOPENTANE
4686147	trans-3-HEPTENE
4038044	3-ETHYL-1-PENTENE
3404613	3-METHYL-1-HEXENE
594569	2,3,3-TRIMETHYL-1-BUTENE
3389429	trans-2-OCTENE
107391	2,4,4-TRIMETHYL-1-PENTENE
107404	2,4,4-TRIMETHYL-2-PENTENE
1632162	2-ETHYL-1-HEXENE
5026766	6-METHYL-1-HEPTENE
4919018	trans-3-OCTENE
4850238	trans-4-OCTENE
693890	1-METHYLCYCLOPENTENE
6746864	2,3-DIMETHYL-1-HEXENE
592450	1,4-HEXADIENE
706310	1,5,9-CYCLODODECATRIENE
764136	2,5-DIMETHYL-2,4-HEXADIENE
629209	1,3,5,7-CYCLOOCTATETRAENE
928494	3-HEXYNE
71432	BENZENE
106423	p-XYLENE
622968	p-ETHYLTOLUENE
108678	MESITYLENE
99876	p-CYMENE
105055	p-DIETHYLBENZENE
877441	1,2,4-TRIETHYLBENZENE
1078713	n-HEPTYLBENZENE
104723	n-DECYLBENZENE
538681	n-PENTYLBENZENE
1077163	n-HEXYLBENZENE
2189608	n-OCTYLBENZENE
1081772	n-NONYLBENZENE
6742547	n-UNDECYLBENZENE
123024	n-TRIDECYLBENZENE
1459105	n-TETRADECYLBENZENE
123013	n-DODECYLBENZENE
1074551	1-METHYL-4-n-PROPYLBENZENE
2765186	1-n-PROPYLNAPHTHALENE
939275	2-ETHYLNAPHTHALENE
7058017	sec-BUTYLCYCLOHEXANE
123911	1,4-DIOXANE
93969	BIS(alpha-METHYLBENZYL) ETHER
56235	CARBON TETRACHLORIDE
110576	1,4-DICHLORO-trans-2-BUTENE
127184	TETRACHLOROETHYLENE
156605	trans-1,2-DICHLOROETHYLENE
355420	PERFLUORO-n-HEXANE
509148	TETRANITROMETHANE
26047	TETRACHLOROSILANE
75150	CARBON DISULFIDE
7146603	2,3-DIMETHYLOCTANE
4032944	2,4-DIMETHYLOCTANE
5869893	2,5-DIMETHYLOCTANE

2040951	n-BUTYLCYCLOPENTANE	2051301	2,6-DIMETHYLOCTANE
110827	CYCLOHEXANE	1072168	2,7-DIMETHYLOCTANE
108872	METHYLCYCLOHEXANE	3524730	5-METHYL-1-HEXENE
1678917	ETHYLCYCLOHEXANE	5870107	2-METHYL-1-HEPTENE
590669	1,1-DIMETHYLCYCLOHEXANE	2131182	n-PENTADECYLBENZENE
6876239	trans-1,2-DIMETHYLCYCLOHEXANE	678262	PERFLUORO-n-PENTANE
638040	cis-1,3-DIMETHYLCYCLOHEXANE	335579	PERFLUORO-n-HEPTANE
2207036	trans-1,3-DIMETHYLCYCLOHEXANE	7550450	TITANIUM TETRACHLORIDE
624293	cis-1,4-DIMETHYLCYCLOHEXANE	2980714	2-METHYL-1-NONENE
1795262	1-trans-3,5-TRIMETHYLCYCLOHEXANE	111966	DIETHYLENE GLYCOL DIMETHYL ETHER
1678928	n-PROPYLCYCLOHEXANE	· 112492	TRIETHYLENE GLYCOL DIMETHYL ETHER
696297	ISOPROPYLCYCLOHEXANE	· 143248	TETRAETHYLENE GLYCOL DIMETHYL ETHER
1678939	n-BUTYLCYCLOHEXANE	1191873	PENTAETHYLENE GLYCOL DIMETHYL ETHER
493016	cis-DECAHYDRONAPHTHALENE	1072408	HEXAETHYLENE GLYCOL DIMETHYL ETHER
493027	trans-DECAHYDRONAPHTHALENE	1191919	HEPTAETHYLENE GLYCOL DIMETHYL ETHER
92513	BICYCLOHEXYL	n.a	OCTAETHYLENE GLYCOL DIMETHYL ETHER
1795160	n-DECYLCYCLOHEXANE	n.a.	NONAETHYLENE GLYCOL DIMETHYL ETHER

Table C.2: List of the 444 polar, non-associating components comprising the database of polar components. Each component is denoted with its CAS number and name [2].

CAS Number	Name	CAS Number	Name
78784	ISOPENTANE	110009	FURAN
1678984	iso-BUTYLCYCLOHEXANE	109999	TETRAHYDROFURAN
109671	1-PENTENE	61015	cis-1,3-DICHLOROPROPENE
627203	cis-2-PENTENE	61026	trans-1,3-DICHLOROPROPENE
563462	2-METHYL-1-BUTENE	628762	1,5-DICHLOROPENTANE
513359	2-METHYL-2-BUTENE	75092	DICHLOROMETHANE
592416	1-HEXENE	67663	CHLOROFORM
7688213	cis-2-HEXENE	75343	1,1-DICHLOROETHANE
7642093	cis-3-HEXENE	107062	1,2-DICHLOROETHANE
763291	2-METHYL-1-PENTENE	79005	1,1,2-TRICHLOROETHANE
760203	3-METHYL-1-PENTENE	78875	1,2-DICHLOROPROPANE
691372	4-METHYL-1-PENTENE	71556	1,1,1-TRICHLOROETHANE
625274	2-METHYL-2-PENTENE	630206	1,1,1,2-TETRACHLOROETHANE
922623	3-METHYL-cis-2-PENTENE	79345	1,1,2,2-TETRACHLOROETHANE
691383	4-METHYL-cis-2-PENTENE	75296	ISOPROPYL CHLORIDE
674760	4-METHYL-trans-2-PENTENE	96184	1,2,3-TRICHLOROPROPANE
592767	1-HEPTENE	513360	ISOBUTYL CHLORIDE
6443921	cis-2-HEPTENE	507200	tert-BUTYL CHLORIDE
6094026	2-METHYL-1-HEXENE	616217	1,2-DICHLOROBUTANE
7642106	cis-3-HEPTENE	79016	TRICHLOROETHYLENE

111660	1-OCTENE
124118	1-NONENE
872059	1-DECENE
821954	1-UNDECENE
112414	1-DODECENE
2437561	1-TRIDECENE
1120361	1-TETRADECENE
3360617	1-PENTADECENE
629732	1-HEXADECENE
112889	1-OCTADECENE
142290	CYCLOPENTENE
110838	CYCLOHEXENE
628922	CYCLOHEPTENE
931884	CYCLOOCTENE
7642048	cis-2-OCTENE
7642151	cis-4-OCTENE
4850227	cis-3-OCTENE
6765395	1-HEPTADECENE
5989275	d-LIMONENE
1574410	cis-1,3-PENTADIENE
2004708	trans-1,3-PENTADIENE
591935	1,4-PENTADIENE
78795	ISOPRENE
598254	3-METHYL-1,2-BUTADIENE
5194514	trans,trans-2,4-HEXADIENE
542927	CYCLOPENTADIENE
513815	2,3-DIMETHYL-1,3-BUTADIENE
592574	1,3-CYCLOHEXADIENE
628411	1,4-CYCLOHEXADIENE
111784	1,5-CYCLOOCTADIENE
237347	trans-1,3-HEXADIENE
1647161	1,9-DECADIENE
503173	DIMETHYLACETYLENE
627190	1-PENTYNE
693027	1-HEXYNE
629050	1-OCTYNE
598232	3-METHYL-1-BUTYNE
108883	TOLUENE
100414	ETHYLBENZENE
95476	o-XYLENE
108383	m-XYLENE
103651	n-PROPYLBENZENE
98828	CUMENE
611143	o-ETHYLTOLUENE
620144	m-ETHYLTOLUENE
526738	1,2,3-TRIMETHYLBENZENE
95636	1,2,4-TRIMETHYLBENZENE
104518	n-BUTYLBENZENE
538932	ISOBUTYLBENZENE
107051	3-CHLOROPROPENE
87683	HEXACHLORO-1,3-BUTADIENE
100447	BENZYL CHLORIDE
108907	MONOCHLOROBENZENE
95501	o-DICHLOROBENZENE
541731	m-DICHLOROBENZENE
95498	o-CHLOROTOLUENE
106434	p-CHLOROTOLUENE
156592	cis-1,2-DICHLOROETHYLENE
126998	CHLOROPRENE
540545	n-PROPYL CHLORIDE
109693	n-BUTYL CHLORIDE
78864	sec-BUTYL CHLORIDE
543599	1-CHLOROPENTANE
76017	PENTACHLOROETHANE
75354	1,1-DICHLOROETHYLENE
124732	1,2-DIBROMO-TETRAFLUOROETHANE
74964	BROMOETHANE
79276	1,1,2,2-TETRABROMOETHANE
106945	1-BROMOPROPANE
75263	2-BROMOPROPANE
109659	1-BROMOBUTANE
629049	1-BROMOHEPTANE
557915	1,1-DIBROMOETHANE
354585	1,1,1-TRICHLORO-TRIFLUOROETHANE
108861	BROMOBENZENE
74884	METHYL IODIDE
75036	ETHYL IODIDE
107084	n-PROPYL IODIDE
75309	ISOPROPYL IODIDE
306832	2,2-DICHLORO-1,1,1-TRIFLUOROETHANE
91667	N,N-DIETHYLANILINE
96548	N-METHYLPYRROLE
120945	N-METHYLPYRROLIDINE
110861	PYRIDINE
121697	N,N-DIMETHYLANILINE
109068	2-METHYLPYRIDINE
75081	ETHYL MERCAPTAN
107039	n-PROPYL MERCAPTAN
75661	tert-BUTYL MERCAPTAN
513440	ISOBUTYL MERCAPTAN
513531	sec-BUTYL MERCAPTAN
111319	n-HEXYL MERCAPTAN
1455216	n-NONYL MERCAPTAN
111886	n-OCTYL MERCAPTAN
75332	ISOPROPYL MERCAPTAN
1569693	CYCLOHEXYL MERCAPTAN
100538	BENZYL MERCAPTAN
624895	METHYL ETHYL SULFIDE

135988	sec-BUTYLBENZENE
98066	tert-BUTYLBENZENE
527844	o-CYMENE
535773	m-CYMENE
141935	m-DIETHYLBENZENE
488233	1,2,3,4-TETRAMETHYLBENZENE
7364194	p-tert-BUTYL ETHYLBENZENE
717748	1,3,5-TRIISOPROPYLBENZENE
605016	PENTAETHYLBENZENE
934747	5-ETHYL-m-XYLENE
1074175	1-METHYL-2-n-PROPYLBENZENE
1827870	1,2,4-TRIMETHYL-3-ETHYLBENZENE
7851273	1,2,4-TRIMETHYL-5-ETHYLBENZENE
100425	STYRENE
611154	o-METHYLSTYRENE
100801	m-METHYLSTYRENE
7564638	o-ETHYLSTYRENE
7525624	m-ETHYLSTYRENE
622979	p-METHYLSTYRENE
98839	alpha-METHYLSTYRENE
536743	ETHYNYLBENZENE
766905	cis-1-PROPENYLBENZENE
1746232	p-tert-BUTYLSTYRENE
90120	1-METHYLNAPHTHALENE
119642	1,2,3,4-TETRAHYDRONAPHTHALENE
95136	INDENE
496117	INDANE
123386	PROPANAL
123728	BUTANAL
78842	2-METHYLPROPANAL
110623	PENTANAL
111717	HEPTANAL
66251	HEXANAL
124130	OCTANAL
124196	NONANAL
623369	2-METHYL-2-PENTENAL
112312	DECANAL
112447	UNDECANAL
112549	DODECANAL
486198	TRIDECANAL
96173	2-METHYLBUTYRALDEHYDE
590863	3-METHYLBUTYRALDEHYDE
107028	ACROLEIN
123739	trans-CROTONALDEHYDE
93538	2-PHENYLPROPIONALDEHYDE
100527	BENZALDEHYDE
123637	PARALDEHYDE
67641	ACETONE
78933	METHYL ETHYL KETONE
96220	3-PENTANONE
3877154	METHYL n-PROPYL SULFIDE
6163640	METHYL t-BUTYL SULFIDE
111477	DI-n-PROPYL SULFIDE
352932	DIETHYL SULFIDE
3698940	ETHYL n-OCTYL SULFIDE
75183	DIMETHYL SULFIDE
110021	THIOPHENE
110816	DIETHYL DISULFIDE
5332525	UNDECYL MERCAPTAN
143102	n-DECYL MERCAPTAN
110667	n-PENTYL MERCAPTAN
624920	DIMETHYL DISULFIDE
629196	DI-n-PROPYL DISULFIDE
112550	n-DODECYL MERCAPTAN
1639094	n-HEPTYL MERCAPTAN
109795	n-BUTYL MERCAPTAN
108985	PHENYL MERCAPTAN
110010	TETRAHYDROTHIOPHENE
67685	DIMETHYL SULFOXIDE
75365	ACETYL CHLORIDE
79049	CHLOROACETYL CHLORIDE
76028	TRICHLOROACETYL CHLORIDE
98884	BENZOYL CHLORIDE
98566	p-CHLOROBENZOTRIFLUORIDE
462066	FLUOROBENZENE
107302	CHLOROMETHYL METHYL ETHER
392563	HEXAFLUOROBENZENE
111444	DI(2-CHLOROETHYL)ETHER
97938	TRIETHYL ALUMINUM
3268493	3-(METHYLMERCAPTO)PROPANAL
68122	N,N-DIMETHYLFORMAMIDE
106898	alpha-EPICHLOROHYDRIN
98011	FURFURAL
108656	PROPYLENE GLYCOL MONOMETHYL ETHER ACETATE
7719122	PHOSPHORUS TRICHLORIDE
25873	PHOSPHORUS OXYCHLORIDE
25782	TRICHLOROSILANE
7791255	SULFUR CHLORIDE
7719097	THIONIL CHLORID
107460	HEXAMETHYLDISILOXANE
141639	DODECAMETHYLPENTASILOXANE
124709	METHYL VINYL DICHLOROSILANE
4814096	[3-(MERCAPTO)PROPYL] TRIETHOXYSILANE
556672	OCTAMETHYLCYCLOTETRASILOXANE
7789200	DEUTERIUM OXIDE
925780	3-NONANONE
616126	3-METHYL-trans-2-PENTENE
616024	METHYL MALEIC ANHYDRIDE
589402	sec-BUTYL FORMATE
540885	tert-BUTYL ACETATE

108101	METHYL ISOBUTYL KETONE
565617	3-METHYL-2-PENTANONE
109499	5-HEXEN-2-ONE
106354	3-HEPTANONE
123193	4-HEPTANONE
589388	3-HEXANONE
107879	2-PENTANONE
563804	METHYL ISOPROPYL KETONE
591786	2-HEXANONE
110430	2-HEPTANONE
141797	MESITYL OXIDE
75978	3,3-DIMETHYL-2-BUTANONE
108838	DIISOBUTYL KETONE
565800	DIISOPROPYL KETONE
872504	N-METHYL-2-PYRROLIDONE
624420	ETHYL ISOAMYL KETONE
502567	5-NONANONE
821556	2-NONANONE
120923	CYCLOPENTANONE
108941	CYCLOHEXANONE
111137	2-OCTANONE
98862	ACETOPHENONE
57578	beta-PROPIOLACTONE
108292	gamma-VALEROLACTONE
565695	ETHYL ISOPROPYL KETONE
814788	METHYL ISOPROPENYL KETONE
108247	ACETIC ANHYDRIDE
123626	PROPIONIC ANHYDRIDE
106310	BUTYRIC ANHYDRIDE
107313	METHYL FORMATE
109944	ETHYL FORMATE
110747	n-PROPYL FORMATE
592847	n-BUTYL FORMATE
542552	ISOBUTYL FORMATE
638493	n-PENTYL FORMATE
79209	METHYL ACETATE
141786	ETHYL ACETATE
109604	n-PROPYL ACETATE
123864	n-BUTYL ACETATE
110190	ISOBUTYL ACETATE
123922	ISOPENTYL ACETATE
108214	ISOPROPYL ACETATE
105464	sec-BUTYL ACETATE
108054	VINYL ACETATE
554121	METHYL PROPIONATE
105373	ETHYL PROPIONATE
106365	n-PROPYL PROPIONATE
590012	n-BUTYL PROPIONATE
105668	n-PROPYL n-BUTYRATE
547637	METHYL ISOBUTYRATE
97621	ETHYL ISOBUTYRATE
622457	CYCLOHEXYL ACETATE
108327	PROPYLENE CARBONATE
84662	DIETHYL PHTHALATE
131113	DIMETHYL PHTHALATE
123251	DIETHYL SUCCINATE
111820	METHYL DODECANOATE
141059	DIETHYL MALEATE
624486	DIMETHYL MALEATE
2998085	sec-BUTYL ACRYLATE
616386	DIMETHYL CARBONATE
2315686	n-PROPYL BENZOATE
628557	DIISOBUTYL ETHER
628717	1-HEPTYNE
462953	ETHYLAL
629141	1,2-DIETHOXYETHANE
1191997	2,3-DIHYDROFURAN
8970440	1-ETHYL-2-ISOPROPYLBENZENE
142289	1,3-DICHLOROPROPANE
78886	2,3-DICHLOROPROPENE
98088	BENZOTRIFLUORIDE
74953	DIBROMOMETHANE
74975	BROMOCHLOROMETHANE
151677	HALOTHANE
75627	BROMOTRICHLOROMETHANE
354234	1,2-DICHLORO-1,1,2-TRIFLUOROETHANE
1717006	1,1-DICHLORO-1-FLUOROETHANE
76120	1,1,2,2-TETRACHLORO-DIFLUOROETHANE
638459	n-HEXYL IODIDE
542698	n-BUTYL IODIDE
108758	2,4,6-TRIMETHYLPYRIDINE
108485	2,6-DIMETHYLPYRIDINE
108996	3-METHYLPYRIDINE
108894	4-METHYLPYRIDINE
420122	THIACYCLOPROPANE
628295	METHYL n-BUTYL SULFIDE
4290927	ETHYL t-BUTYL SULFIDE
287274	TRIMETHYLENE SULFIDE
554143	2-METHYLTHIOPHENE
616444	3-METHYLTHIOPHENE
127195	N,N-DIMETHYLACETAMIDE
105395	ETHYLCHLOROACETATE
4394858	4-FORMYLMORPHOLINE
1184583	DIMETHYLALUMINUM CHLORIDE
107517	OCTAMETHYLTRISILOXANE
556694	OCTADECAMETHYLOCTASILOXANE
631367	TETRAETHYL SILANE
541015	HEXADECAMETHYLHEPTASILOXANE

623427	METHYL n-BUTYRATE
105544	ETHYL n-BUTYRATE
96333	METHYL ACRYLATE
140885	ETHYL ACRYLATE
141322	n-BUTYL ACRYLATE
108645	ETHYL ISOVALERATE
80626	METHYL METHACRYLATE
628637	n-PENTYL ACETATE
103093	2-ETHYLHEXYL ACETATE
140114	BENZYL ACETATE
659701	ISOPENTYL ISOVALERATE
142927	n-HEXYL ACETATE
120514	BENZYL BENZOATE
136607	n-BUTYL BENZOATE
112061	n-HEPTYL ACETATE
112141	n-OCTYL ACETATE
112174	n-DECYL ACETATE
131168	DI-n-PROPYL PHTHALATE
109217	n-BUTYL n-BUTYRATE
103117	2-ETHYLHEXYL ACRYLATE
97869	ISOBUTYL METHACRYLATE
97881	n-BUTYL METHACRYLATE
93583	METHYL BENZOATE
93890	ETHYL BENZOATE
105588	DIETHYL CARBONATE
95921	DIETHYL OXALATE
105533	DIETHYL MALONATE
110429	METHYL DECANOATE
110270	ISOPROPYL MYRISTATE
60297	DIETHYL ETHER
108203	DIISOPROPYL ETHER
142961	DI-n-BUTYL ETHER
1634044	METHYL tert-BUTYL ETHER
557175	METHYL n-PROPYL ETHER
598538	METHYL ISOPROPYL ETHER
628284	METHYL n-BUTYL ETHER
628320	ETHYL PROPYL ETHER
104461	ANETHOLE
693652	DI-n-PENTYL ETHER
637923	tert-BUTYL ETHYL ETHER
919948	ETHYL tert-PENTYL ETHER
109875	METHYLAL
105577	ACETAL
75569	1,2-PROPYLENE OXIDE
109922	ETHYL VINYL ETHER
111433	DI-n-PROPYL ETHER
628819	n-BUTYL ETHYL ETHER
110714	1,2-DIMETHOXYETHANE
100663	ANISOLE
103731	PHENETOLE
103504	DIBENZYL ETHER
96106	DIETHYLALUMINUM CHLORIDE
97723	ISOBUTYRIC ANHYDRIDE
6737117	sec-BUTENYL ACETATE
3938952	ETHYL TRIMETHYL ACETATE
106650	DIMETHYL SUCCINATE
1795091	2-METHYLTHIACYCLOPENTANE
1551322	2-ETHYL-TETRAHYDROTHIOPHENE
872559	2-ETHYLTHIOPHENE
1551275	2-n-PROPYLTHIOPHENE
638028	2,5-DIMETHYLTHIOPHENE
453317	ETHYL PROPYL DISULFIDE
100685	METHYL PHENYL SULFIDE
544401	DI-n-BUTYL SULFIDE
4110503	ETHYL PROPYL SULFIDE
143226	2-(2-(2-BUTOXYETHOXY)ETHOXY) ETHANOL
124630	METHANESULFONYL CHLORIDE
585079	tert-BUTYL METHACRYLATE
7784341	ARSENIC (III) CHLORIDE
78104	TETRAETHOXYSILANE
75547	METHYL DICHLOROSILANE
75796	METHYL TRICHLOROSILANE
75945	VINYLTRICHLOROSILANE
1067534	TRIS(2-METHOXYETHOXY)VINYLSILANE
681845	METHYL SILICATE
545990	SULFUR DICHLORIDE
75241	TRIMETHYLALUMINUM
78002	TETRAETHYL LEAD
1066359	DIMETHYLCHLOROSILANE
75774	TRIMETHYLCHLOROSILANE
75785	DIMETHYLDICHLOROSILANE
541026	DECAMETHYLCYCLOPENTASILOXANE
540976	DODECAMETHYLCYCLOHEXASILOXANE
80104	DIPHENYLDICHLOROSILANE
115219	ETHYLTRICHLOROSILANE
1719535	DICHLORODIETHYLSILANE
75876	TRICHLOROACETALDEHYDE
78400	TRIETHYL PHOSPHATE
2768027	VINYLTRIMETHOXYSILANE
2487903	TRIMETHOXYSILANE
1112396	DIMETHYLDIMETHOXYSILANE
98135	PHENYLTRICHLOROSILANE
149746	PHENYLMETHYLDICHLOROSILANE
107528	TETRADECAMETHYLHEXASILOXANE
141628	DECAMETHYLTETRASILOXANE
625605	ETHYL THIOLACETATE
77781	DIMETHYL SULFATE
64675	DIETHYL SULFATE
111159	ETHOXYETHYL ACETATE
141979	ETHYL ACETOACETATE
594445	ETHANESULFONYL CHLORIDE
540636	1,2-ETHANEDITHIOL

9177049	1-METHYL-3-(METHYLETHOXY)BENZENE	763699	ETHYL-3-ETHOXYPROPIONATE
534156	1,1-DIMETHOXYETHANE	108225	METHYLVINYL ACETATE
1708298	2,5-DIHYDROFURAN	5878193	METHOXYACETONE

References

[1] M. Stavrou, M. Lampe, A. Bardow, and J. Gross, *Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for CO_2 capture,* Ind. Eng. Chem. Res. **53**, 18029 (2014).

[2] AIChE-DECHEMA, *Pure Component Database DIPPR-801*, Tech. Rep. (AIChE-DECHEMA, 2011).

D

QSPR models for the prediction of c_p^{ig} and molar mass

The size of the training sets and the squared correlation coefficients R^2 of the QSPR multivariable regression model for the prediction of c_p^{ig} and of the molar mass M are given in Table D.1.

Table D.1: QSPR multivariable regression model for the prediction of molar mass and c_p^{ig}: Size of the training set (number of components) used to adjust the coefficients of the QSPR-model and the resulting squared correlation coefficient R^2 for non-polar and polar components.

	non-polar components	polar components
Ideal gas specific heat capacity c_p^{ig} by $T = 300\,\mathrm{K}$		
training set size	156	525
R^2	0.997	0.987
Molar mass M		
training set size	168	445
R^2	0.966	0.932

Tables D.2 and D.3 present the QSPR model coefficients used in Eq.(2.8) and Eq.(2.7) of Chapter 2 for the prediction of the ideal gas heat capacity c_p^{ig} and of the molar mass M, respectively. The coefficients of the QSPR models were adjusted for the PC-SAFT pure component database, as described in Chapter 2.

Table D.2: Coefficients of the QSPR model for the prediction of the ideal gas heat capacity c_p^{ig} given by Eq.(2.8).

QSPR coefficients - Ideal gas specific heat capacity c_p^{ig} by $T = 300\,\mathrm{K}$		
	non-polar components	polar components
γ_0	-16955.243	-20257.909
γ_1	-3829.1124	13051.796
γ_2	1279.9093	914.33815
γ_3	-327.68470	-227.59750

Table D.3: Coefficients of the QSPR model for the prediction of the molar mass M given by Eq.(2.7).

QSPR coefficients - Molar mass M		
	non-polar components	polar components
β_1	0.0978500	0.0011995
β_2	-0.0020034	0.5673782
β_3	0.0006309	0.0001554
β_4	0	0.0308482

E

QSPR models for the prediction of k_{ij}

The contents of Appendices E.1 to E.3 are published as Supporting Information to the work of Stavrou et al. [1].

E.1. Comparison with k_{ij} predictions based on London's dispersive theory

Table E.1: Mixtures of two non-associating, non-polar components (group 1): For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the value of k_{ij} fitted to experimental data (k_{ij}^{fit}), the value of k_{ij} prediction with the QSPR model (k_{ij}^{QSPR}) and the value of k_{ij} prediction with the Hudson-McCoubrey expression[2] ($k_{ij}(I,\sigma)$).

Group 1 (nAnP - nAnP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	$k_{ij}(I,\sigma)$
1	3-METHYLPENTANE	OCTANE	96140	111659	-0.0003	0.0000	0.0001
2	CARBON-MONOXIDE	PROPANE	630080	74986	0.0517	0.0622	0.0161
3	CARBON-MONOXIDE	METHANE	630080	74828	0.0287	0.0301	0.0141
4	CARBON-MONOXIDE	ETHANE	630080	74840	0.0258	0.0462	0.0094
5	PROPANE	METHYLCYCLOHEXANE	74986	108872	0.0102	0.0153	0.0095
6	ETHANE	METHYLCYCLOHEXANE	74840	108872	0.0169	0.0226	0.0162
7	METHANE	PROPANE	74828	74986	0.0012	0.0082	0.0029
8	METHANE	METHYLCYCLOHEXANE	74828	108872	0.0338	0.0357	0.0133
9	METHANE	ETHANE	74828	74840	-0.0046	0.0041	0.0029
10	PROPANE	CYCLOHEXANE	74986	110827	0.0141	0.0119	0.0042
11	METHANE	HEXADECANE	74828	544763	0.0545	0.0259	0.0000
12	METHANE	OCTANE	74828	111659	0.0174	0.0198	0.0088
13	METHANE	ETHYLCYCLOHEXANE	74828	1678917	0.0278	0.0436	0.0174
14	METHANE	HEXANE	74828	110543	0.0135	0.0174	0.0064
15	PROPANE	ISOBUTANE	74986	75285	-0.0048	0.0022	0.0012

16	HEPTANE	METHYLCYCLOHEXANE	142825	108872	0.0021	0.0065	0.0020
17	ETHANE	HEXANE	74840	110543	0.0066	0.0095	0.0065
18	HEXANE	HEXADECANE	110543	544763	0.0011	0.0032	0.0000
19	HEXANE	OCTANE	110543	111659	-0.0030	0.0009	0.0002
20	ETHANE	OCTANE	74840	111659	0.0174	0.0112	0.0089
21	METHANE	BUTANE	74828	106978	0.0014	0.0126	0.0040
22	METHANE	NONANE	74828	111842	0.0148	0.0204	0.0095
23	CYCLOHEXANE	METHYLCYCLOHEXANE	110827	108872	0.0016	0.0016	0.0012
24	METHANE	HEPTANE	74828	142825	0.0112	0.0180	0.0076
25	HEXANE	CYCLOHEXANE	110543	110827	0.0016	0.0045	0.0002
26	2-METHYLPENTANE	OCTANE	107835	111659	-0.0019	0.0005	0.0001
27	METHANE	DECANE	74828	124185	0.0244	0.0201	0.0098
28	HEXANE	METHYLCYCLOPENTANE	110543	96377	0.0007	0.0029	0.0001
29	METHANE	PENTANE	74828	109660	0.0104	0.0156	0.0054
30	PROPANE	BUTANE	74986	106978	-0.0051	0.0024	0.0006
31	METHANE	ISOBUTANE	74828	75285	0.0062	0.0121	0.0040
32	ETHANE	DECANE	74840	124185	0.0196	0.0114	0.0096
33	BUTANE	DECANE	106978	124185	0.0071	0.0034	0.0018
34	ISOBUTANE	BUTANE	75285	106978	-0.0006	0.0002	0.0001
35	PENTANE	DECANE	109660	124185	0.0051	0.0018	0.0007
36	HEPTANE	OCTANE	142825	111659	0.0047	0.0007	0.0001
37	PROPANE	DECANE	74986	124185	0.0060	0.0066	0.0046
38	HEPTANE	p-XYLENE	142825	106423	0.0053	0.0040	0.0033
39	CYCLOHEXANE	HEPTANE	110827	142825	0.0012	0.0042	0.0001
40	2-3-DIMETHYLBUTANE	HEXANE	79298	110543	0.0005	0.0023	0.0012
41	PENTANE	HEXANE	109660	110543	0.0067	0.0007	0.0001
42	HEXADECANE	ETHANE	544763	74840	0.0064	0.0156	0.0000
43	2-METHYLPENTANE	HEPTANE	107835	142825	-0.0007	0.0002	0.0002
44	3-METHYLPENTANE	HEPTANE	96140	142825	-0.0001	0.0007	0.0002
45	HEXANE	HEPTANE	110543	142825	-0.0038	0.0002	0.0001
46	CYCLOHEXANE	OCTANE	110827	111659	0.0004	0.0034	0.0000
47	PENTANE	OCTANE	109660	111659	0.0047	0.0017	0.0005

Table E.2: Mixtures of a non-associating, dipolar and a non-associating, quadrupolar component (group 4): For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the value of k_{ij} fitted to experimental data (k_{ij}^{fit}), the value of k_{ij} prediction with the QSPR model (k_{ij}^{QSPR}) and the value of k_{ij} prediction with the Hudson-McCoubrey expression[2] ($k_{ij}(I,\sigma)$).

Group 4 (nAdP - nAqP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	$k_{ij}(I,\sigma)$
1	ACETONE	NITROGEN	67641	7727379	0.0336	0.0146	0.0284
2	NITROGEN	PROPYLBENZENE	7727379	103651	0.1658	0.1433	0.0527
3	BENZENE	1-BROMOBENZENE	71432	108861	-0.0082	0.0033	0.0002
4	CARBON-DIOXIDE	1-BROMOBENZENE	124389	108861	0.0462	0.0407	0.0446
5	SULFUR-DIOXIDE	BENZENE	7446095	71432	0.0494	0.0529	0.0864
6	CARBON-DIOXIDE	PROPYLBENZENE	124389	103651	0.0382	0.0533	0.0516
7	DIMETHYL-ETHER	CARBON-DIOXIDE	115106	124389	-0.0574	-0.0285	0.0132
8	PROPYLENE	NITROGEN	115071	7727379	0.0642	0.0466	0.0282
9	BUTYLBENZENE	BENZENE	104518	71432	-0.0101	0.0013	0.0011

10	CARBON-DIOXIDE	1-HEXENE	124389	592416	0.0125	0.0244	0.0346
11	VINYL-ACETATE	ETHYLENE	108054	74851	-0.0159	0.0071	0.0045
12	CARBON-DIOXIDE	o-XYLENE	124389	95476	0.0343	0.0445	0.0477
13	CARBON-DIOXIDE	ISOPROPYL-ETHANOATE	124389	108214	0.0019	-0.0158	0.0181
14	PROPYLENE	BENZENE	115071	71432	0.0087	0.0166	0.0035
15	DIETHYL-ETHER	CARBON-DISULFIDE	60297	75150	0.0057	0.0206	0.0026
16	CARBON-DIOXIDE	TOLUENE	124389	108883	0.0246	0.0461	0.0427
17	NITROGEN	TOLUENE	7727379	108883	0.1223	0.1304	0.0462
18	CARBON-DIOXIDE	CHLOROBENZENE	124389	108907	0.0351	0.03	0.0412
19	PROPYLENE	ETHYLENE	115071	74851	-0.0004	0.0054	0.001
20	1-HEXENE	ETHYLENE	592416	74851	0.0443	0.013	0.0054
21	1-HEXENE	BENZENE	592416	71432	0.0011	0.0076	0.0002
22	BENZENE	CHLOROBENZENE	71432	108907	-0.002	0.0016	0.0000
23	CARBON-DIOXIDE	ETHYLBENZENE	124389	100414	0.0345	0.0451	0.0471
24	BENZENE	TOLUENE	71432	108883	-0.0062	0.0015	0.0003

E.2. Binary mixture database of the QSPR method - Results and comparison with k_{ij} values adjusted to experimental VLE data

Table E.3: Regression results for mixtures of two non-associating, non-polar components (group 1). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted on experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

Group 1 (nAnP - nAnP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
Training set							
1	3-METHYLPENTANE	OCTANE	96140	111659	-0.0003	0	0
2	CARBON-MONOXIDE	PROPANE	630080	74986	0.0517	0.0622	0
3	CARBON-MONOXIDE	METHANE	630080	74828	0.0287	0.0301	0
4	CARBON-MONOXIDE	ETHANE	630080	74840	0.0258	0.0462	0
5	PROPANE	METHYLCYCLOHEXANE	74986	108872	0.0102	0.0153	0
6	ETHANE	METHYLCYCLOHEXANE	74840	108872	0.0169	0.0226	0
7	METHANE	PROPANE	74828	74986	0.0012	0.0082	0
8	METHANE	METHYLCYCLOHEXANE	74828	108872	0.0338	0.0357	0
9	METHANE	ETHANE	74828	74840	-0.0046	0.0041	0
10	PROPANE	CYCLOHEXANE	74986	110827	0.0141	0.0119	0
11	METHANE	HEXADECANE	74828	544763	0.0545	0.0259	0
12	METHANE	OCTANE	74828	111659	0.0174	0.0198	0
13	METHANE	ETHYLCYCLOHEXANE	74828	1678917	0.0278	0.0436	0
14	METHANE	HEXANE	74828	110543	0.0135	0.0174	0
15	PROPANE	ISOBUTANE	74986	75285	-0.0048	0.0022	0
16	HEPTANE	METHYLCYCLOHEXANE	142825	108872	0.0021	0.0065	0
17	ETHANE	HEXANE	74840	110543	0.0066	0.0095	0
18	HEXANE	HEXADECANE	110543	544763	0.0011	0.0032	0
19	HEXANE	OCTANE	110543	111659	-0.003	0.0009	0
20	ETHANE	OCTANE	74840	111659	0.0174	0.0112	0

21	METHANE	BUTANE	74828	106978	0.0014	0.0126	0
22	METHANE	NONANE	74828	111842	0.0148	0.0204	0
23	CYCLOHEXANE	METHYLCYCLOHEXANE	110827	108872	0.0016	0.0016	0
24	METHANE	HEPTANE	74828	142825	0.0112	0.018	0
25	HEXANE	CYCLOHEXANE	110543	110827	0.0016	0.0045	0
26	2-METHYLPENTANE	OCTANE	107835	111659	-0.0019	0.0005	0
27	METHANE	DECANE	74828	124185	0.0244	0.0201	0
28	HEXANE	METHYLCYCLOPENTANE	110543	96377	0.0007	0.0029	0
29	METHANE	PENTANE	74828	109660	0.0104	0.0156	0
30	PROPANE	BUTANE	74986	106978	-0.0051	0.0024	0
31	METHANE	ISOBUTANE	74828	75285	0.0062	0.0121	0
32	ETHANE	DECANE	74840	124185	0.0196	0.0114	0
33	BUTANE	DECANE	106978	124185	0.0071	0.0034	0
34	ISOBUTANE	BUTANE	75285	106978	-0.0006	0.0002	0
35	PENTANE	DECANE	109660	124185	0.0051	0.0018	0
36	HEPTANE	OCTANE	142825	111659	0.0047	0.0007	0
37	PROPANE	DECANE	74986	124185	0.006	0.0066	0
38	HEPTANE	p-XYLENE	142825	106423	0.0053	0.004	0
39	CYCLOHEXANE	HEPTANE	110827	142825	0.0012	0.0042	0
40	2-3-DIMETHYLBUTANE	HEXANE	79298	110543	0.0005	0.0023	0
41	PENTANE	HEXANE	109660	110543	0.0067	0.0007	0
			Test set				
42	HEXADECANE	ETHANE	544763	74840	0.0064	0.0156	0
43	2-METHYLPENTANE	HEPTANE	107835	142825	-0.0007	0.0002	0
44	3-METHYLPENTANE	HEPTANE	96140	142825	-0.0001	0.0007	0
45	HEXANE	HEPTANE	110543	142825	-0.0038	0.0002	0
46	CYCLOHEXANE	OCTANE	110827	111659	0.0004	0.0034	0
47	PENTANE	OCTANE	109660	111659	0.0047	0.0017	0

Table E.4: Regression results for mixtures of non-associating, non-polar with non-associating, dipolar components (group 2). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

Group 2 (nAnP - nAdP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
			Training set				
1	CARBON-MONOXIDE	1-OCTENE	630080	111660	0.0705	0.0944	0
2	CYCLOHEXANE	N-METHYL-2-PYRROLIDONE	110827	872504	0.0166	0.0203	0
3	METHANE	PROPYLBENZENE	74828	103651	0.0318	0.0341	0
4	CYCLOHEXANE	TOLUENE	110827	108883	0.0144	0.0019	0
5	ETHYL-ETHANOATE	METHYLCYCLOHEXANE	141786	108872	0.0508	0.0274	1
6	METHYLCYCLOHEXANE	CHLOROBENZENE	108872	108907	0.0121	0.0085	0
7	ETHANE	PROPYLBENZENE	74840	103651	0.0252	0.0219	0
8	CHLOROBUTANE	METHYLCYCLOHEXANE	109693	108872	0.0122	0.017	0
9	PROPANE	TOLUENE	74986	108883	0.0201	0.0135	0
10	HEXANE	CHLOROBENZENE	110543	108907	0.0075	0.0154	0

11	METHANE	PROPYLENE	74828	115071	-0.0046	0.0091	0
12	CHLOROBUTANE	HEPTANE	109693	142825	0.009	0.0115	0
13	CYCLOHEXANE	ISOPROPYL-ETHANOATE	110827	108214	0.0303	0.0199	1
14	NONANE	PROPYLBENZENE	111842	103651	0.0047	0.0057	1
15	ETHYL-ETHANOATE	CYCLOHEXANE	141786	110827	0.0501	0.0238	1
16	HEXANE	N-METHYL-2-PYRROLIDONE	110543	872504	0.0093	0.0239	0
17	PROPANE	DIMETHYL-ETHER	74986	115106	0.026	0.0136	1
18	HEPTANE	BUTYL-ETHANOATE	142825	123864	0.0289	0.0112	0
19	ETHANE	PROPYLENE	74840	115071	-0.0058	0.0044	0
20	CYCLOHEXANE	BUTYL-ETHANOATE	110827	123864	0.0265	0.0163	0
21	DIMETHYL-ETHER	ISOBUTANE	115106	75285	0.0281	0.0165	1
22	OCTANE	o-XYLENE	111659	95476	-0.0032	0.0068	0
23	CYCLOHEXANE	CHLOROBENZENE	110827	108907	0.0121	0.0101	0
24	METHYLCYCLOHEXANE	TOLUENE	108872	108883	0.012	0.0035	0
25	CHLOROBENZENE	DECANE	108907	124185	-0.0082	0.0139	0
26	1-HEXENE	HEXANE	592416	110543	0.0033	0.0034	0
27	METHYLCYCLOHEXANE	BUTYL-ETHANOATE	108872	123864	0.0276	0.0189	0
28	HEXANE	ETHYLBENZENE	110543	100414	0.0076	0.0077	0
29	TOLUENE	DECANE	108883	124185	0.0054	0.0049	0
30	ETHYLBENZENE	p-XYLENE	100414	106423	0.0003	0.0033	0
31	1-HEXENE	OCTANE	592416	111659	0.0006	0.0044	0
32	ETHYL-ETHANOATE	HEPTANE	141786	142825	0.0419	0.0169	1
33	1-BUTENE	BUTANE	106989	106978	0	0.0037	0
34	ETHYLBENZENE	NONANE	100414	111842	0.0072	0.0063	0
35	HEPTANE	TOLUENE	142825	108883	0.0046	0.0059	0
36	PROPYLENE	PROPANE	115071	74986	0.0069	0.0039	0
37	HEPTANE	o-XYLENE	142825	95476	0.0051	0.0077	0
38	HEPTANE	CHLOROBENZENE	142825	108907	0.0073	0.0151	0
39	OCTANE	PROPYLBENZENE	111659	103651	0.0066	0.006	0
40	HEPTANE	ETHYLBENZENE	142825	100414	0.0073	0.0075	0
		Test set					
41	HEXANE	CHLOROBUTANE	110543	109693	0.008	0.0117	1
42	1-HEXENE	HEPTANE	592416	142825	0.0039	0.0037	0
43	HEPTANE	N-METHYL-2-PYRROLIDONE	142825	872504	0.01	0.0236	0
44	TOLUENE	OCTANE	108883	111659	0.0072	0.005	0
45	CYCLOHEXANE	PROPYL-ETHANOATE	110827	109604	0.0355	0.0197	1

Table E.5: Regression results for mixtures of non-associating, non-polar with non-associating, quadrupolar components (group 3). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

Group 3 (nAnP - nAqP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
Training set							
1	NITROGEN	METHYLCYCLOHEXANE	7727379	108872	0.1245	0.1592	0
2	BENZENE	p-XYLENE	71432	106423	-0.0049	0.0017	0

3	CARBON-DIOXIDE	ETHYLCYCLOHEXANE	124389	1678917	0.0707	0.0797	0
4	NITROGEN	ISOBUTANE	7727379	75285	0.1001	0.0741	0
5	CARBON-DIOXIDE	BUTANE	124389	106978	0.0358	0.0312	0
6	NITROGEN	METHANE	7727379	74828	0.0226	0.0309	0
7	CARBON-DIOXIDE	PENTADECANE	124389	629629	0.0551	0.0528	0
8	NITROGEN	NONANE	7727379	111842	0.1232	0.1038	0
9	METHANE	BENZENE	74828	71432	0.0265	0.0323	1
10	CARBON-DIOXIDE	p-XYLENE	124389	106423	0.0282	0.0539	0
11	CARBON-DIOXIDE	ETHANE	124389	74840	0.0195	0.0174	1
12	NITROGEN	ETHANE	7727379	74840	0.039	0.0444	0
13	ETHYLENE	OCTANE	74851	111659	0.0182	0.0171	0
14	NITROGEN	PROPANE	7727379	74986	0.0623	0.0593	0
15	CARBON-DIOXIDE	METHYLCYCLOHEXANE	124389	108872	0.0573	0.067	0
16	CARBON-DIOXIDE	OCTANE	124389	111659	0.0451	0.0428	0
17	NITROGEN	PENTANE	7727379	109660	0.089	0.0865	0
18	CARBON-DIOXIDE	PROPANE	124389	74986	0.0327	0.0241	0
19	NITROGEN	HEPTANE	7727379	142825	0.1098	0.095	0
20	DODECANE	CARBON-DIOXIDE	112403	124389	0.0501	0.0479	0
21	CARBON-DIOXIDE	CYCLOHEXANE	124389	110827	0.0608	0.056	0
22	CARBON-DIOXIDE	PENTANE	124389	109660	0.0384	0.0362	0
23	CARBON-DISULFIDE	CYCLOPENTANE	75150	287923	0.0011	0.0078	1
24	CARBON-DIOXIDE	HEPTANE	124389	142825	0.0427	0.0399	0
25	ETHANE	BENZENE	74840	71432	0.0196	0.0208	0
26	BENZENE	OCTANE	71432	111659	0.0038	0.0052	0
27	BENZENE	METHYLCYCLOHEXANE	71432	108872	0.0089	0.0027	0
28	NITROGEN	OCTANE	7727379	111659	0.1222	0.1016	0
29	CARBON-DIOXIDE	OCTADECANE	124389	593453	0.0843	0.0544	0
30	ETHYLENE	ETHANE	74851	74840	0.0124	0.0031	0
31	BENZENE	DODECANE	71432	112403	-0.0005	0.0039	0
32	HEXANE	BENZENE	110543	71432	0.006	0.0063	1
33	CARBON-DIOXIDE	NONANE	124389	111842	0.0332	0.0437	0
34	METHYLCYCLOPENTANE	BENZENE	96377	71432	0.0076	0.003	1
35	NITROGEN	DECANE	7727379	124185	0.1109	0.1026	0
36	CARBON-DIOXIDE	UNDECANE	124389	1120214	0.0605	0.0474	0
			Test set				
37	CARBON-DISULFIDE	CYCLOHEXANE	75150	110827	-0.0029	0.0049	0
38	BENZENE	DECANE	71432	124185	0.0008	0.005	0
39	CARBON-DIOXIDE	DECANE	124389	124185	0.0518	0.0432	0
40	METHANE	ETHYLENE	74828	74851	0.0095	0.0029	0
41	BENZENE	HEPTANE	71432	142825	0.005	0.006	0

Table E.6: Regression results for mixtures of non-associating, dipolar with non-associating, quadrupolar components (group 4). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

Group 4 (nAdP - nAqP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo

			Training set				
1	ACETONE	NITROGEN	67641	7727379	0.0336	0.0146	0
2	NITROGEN	PROPYLBENZENE	7727379	103651	0.1658	0.1433	0
3	BENZENE	1-BROMOBENZENE	71432	108861	-0.0082	0.0033	0
4	CARBON-DIOXIDE	1-BROMOBENZENE	124389	108861	0.0462	0.0407	0
5	SULFUR-DIOXIDE	BENZENE	7446095	71432	0.0494	0.0529	0
6	CARBON-DIOXIDE	PROPYLBENZENE	124389	103651	0.0382	0.0533	0
7	DIMETHYL-ETHER	CARBON-DIOXIDE	115106	124389	-0.0574	-0.0285	0
8	PROPYLENE	NITROGEN	115071	7727379	0.0642	0.0466	0
9	BUTYLBENZENE	BENZENE	104518	71432	-0.0101	0.0013	0
10	CARBON-DIOXIDE	1-HEXENE	124389	592416	0.0125	0.0244	0
11	VINYL-ACETATE	ETHYLENE	108054	74851	-0.0159	0.0071	0
12	CARBON-DIOXIDE	o-XYLENE	124389	95476	0.0343	0.0445	0
13	CARBON-DIOXIDE	ISOPROPYL-ETHANOATE	124389	108214	0.0019	-0.0158	0
14	PROPYLENE	BENZENE	115071	71432	0.0087	0.0166	0
15	DIETHYL-ETHER	CARBON-DISULFIDE	60297	75150	0.0057	0.0206	1
16	CARBON-DIOXIDE	TOLUENE	124389	108883	0.0246	0.0461	0
17	NITROGEN	TOLUENE	7727379	108883	0.1223	0.1304	0
18	CARBON-DIOXIDE	CHLOROBENZENE	124389	108907	0.0351	0.03	1
19	PROPYLENE	ETHYLENE	115071	74851	-0.0004	0.0054	0
20	1-HEXENE	ETHYLENE	592416	74851	0.0443	0.013	0
21	1-HEXENE	BENZENE	592416	71432	0.0011	0.0076	0
			Test set				
22	BENZENE	CHLOROBENZENE	71432	108907	-0.002	0.0016	0
23	CARBON-DIOXIDE	ETHYLBENZENE	124389	100414	0.0345	0.0451	0
24	BENZENE	TOLUENE	71432	108883	-0.0062	0.0015	0

Table E.7: Regression results for mixtures of two non-associating, quadrupolar components (group 5). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

			Group 5 (nAqP - nAqP)				
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
			Training set				
1	CARBON-DIOXIDE	BENZENE	124389	71432	0.0502	0.0479	0
2	CARBON-DISULFIDE	BENZENE	75150	71432	0.0134	0.0032	0
3	NITROGEN	BENZENE	7727379	71432	0.1374	0.1365	0
			Test set				
4	NITROGEN	ETHYLENE	7727379	74851	0.0616	0.0326	0

Table E.8: Regression results for mixtures of two non-associating, dipolar components (group 6). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

Group 6 (nAdP - nAdP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
Training set							
1	N-N-DIMETHYL-FORMAMIDE	1-METHYLNAPHTHALENE	68122	90120	-0.0059	0	0
2	m-XYLENE	o-XYLENE	108383	95476	0	0.002	0
3	DICHLOROMETHANE	o-XYLENE	75092	95476	-0.0002	0.0199	0
4	ISOBUTENE	N-METHYL-2 PYRROLIDONE	115117	872504	-0.014	-0.0246	0
5	DIETHYL-ETHER	ACETONE	60297	67641	0.0028	-0.001	0
6	ACETONE	TOLUENE	67641	108883	0.0007	0.0138	0
7	PROPYLENE	DIMETHYL-ETHER	115071	115106	0	0.0078	0
8	ACETONE	3-PENTANONE	67641	96220	0.0031	0.0085	0
9	2-METHYLPROPANAL	N-N-DIMETHYL-FORMAMIDE	78842	68122	-0.0013	-0.001	0
10	ACETALDEHYDE	ACETONITRILE	75070	75058	-0.0017	0.0014	0
Test set							
11	ACETONITRILE	PROPIONITRILE	75058	107120	0	0.0078	0
12	ACETONE	o-XYLENE	67641	95476	0.0095	0.013	0

Table E.9: Regression results for mixtures of non-associating, non-polar with associating, dipolar components (group 7). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

Group 7 (nAnP - AdP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
Training set							
1	ETHANE	1-OCTANOL	74840	111875	0.0011	0.0011	0
2	CARBON-MONOXIDE	PROPIONIC-ACID	630080	79094	0.0708	0.1111	0
3	DIETHYLAMINE	HEPTANE	109897	142825	-0.0333	-0.004	0
4	DODECANE	epsilon-CAPROLACTAM	112403	105602	-0.0136	0.0033	1
5	CYCLOHEXANE	n-HEXANOIC-ACID	110827	142621	-0.0236	-0.0119	0
6	HEXANE	ISOPROPANOL	110543	67630	0.0402	0.0338	1
7	METHYLAMINE	HEXANE	74895	110543	0.01	0.0073	0
8	METHANOL	HEPTANE	67561	142825	-0.0093	0.0097	1
9	TRIDECANE	1-DODECANOL	629505	112538	0.012	-0.0075	0
10	1-OCTANOL	n-HEPTYLBENZENE	111875	1078713	0.0274	0.0217	0
11	ETHANOL	NONANE	64175	111842	0.0189	0.02	1
12	HEXANE	1-HEXADECANOL	110543	36653824	0.0008	-0.013	0
13	HEPTANE	NEOPENTANOIC-ACID	142825	75989	0.01	0.025	0
14	HEPTANE	2-METHYL-1-BUTANOL	142825	137326	0.0149	0.0094	1

15	CYCLOHEXANE	NEOPENTANOIC-ACID	110827	75989	0.0272	0.0406	0
16	PENTANE	1-PENTANOL	109660	71410	0.01	0.0081	0
17	1-PROPANOL	HEPTANE	71238	142825	0.026	0.0193	1
18	METHANOL	CYCLOHEXANE	67561	110827	0.0122	0.0205	1
19	CYCLOHEXANE	1-HEXANOL	110827	111273	0.0217	0.0114	0
20	PROPANE	1-DECANOL	74986	112301	0.0042	-0.0018	0
21	CYCLOHEXANE	PYRROLIDINE	110827	123751	-0.0158	-0.0049	1
22	BUTANE	ETHANOL	106978	64175	0.0105	0.0108	1
23	CYCLOHEXANE	1-BUTANOL	110827	71363	0.0426	0.0335	1
24	2-2-DIMETHYLBUTANE	1-PENTANOL	75832	71410	0.0141	0.0158	0
25	HEXANE	1-HEXANOL	110543	111273	0.011	0.0027	0
26	METHYLCYCLOHEXANE	1-BUTANOL	108872	71363	0.0467	0.041	1
27	DECANE	1-OCTANOL	124185	111875	0.0201	0.0073	1
28	1-BUTANOL	DODECANE	71363	112403	0.0246	0.0256	0
29	PROPANE	1-PROPANOL	74986	71238	-0.0044	0.0078	0
30	ETHANOL	HEXADECANE	64175	544763	0.0167	0.0266	0
31	CYCLOHEXANE	1-PROPANOL	110827	71238	0.0321	0.0329	1
32	CYCLOPENTANE	CYCLOHEXANOL	287923	108930	0.019	0.0132	0
33	CYCLOHEXANE	CYCLOHEXANOL	110827	108930	0.0245	0.0195	0
34	1-PROPANOL	UNDECANE	71238	1120214	0.0289	0.0247	0
35	ETHYLAMINE	HEXANE	75047	110543	-0.0006	0.0009	0
36	HEXANE	1-DECANOL	110543	112301	0.0054	-0.0059	0
37	BUTANE	ETHYLAMINE	106978	75047	0.003	0.0049	1
38	3-METHYLPENTANE	1-PENTANOL	96140	71410	0.0157	0.0118	0
39	HEXANE	ETHANOL	110543	64175	0.0102	0.0165	1
40	1-BUTANOL	DECANE	71363	124185	0.0351	0.0223	1
41	PENTANE	1-PROPANOL	109660	71238	0.022	0.0165	0
42	1-PROPANOL	NONANE	71238	111842	0.0296	0.0221	1
43	METHYLAMINE	NONANE	74895	111842	0.0106	0.0094	0
44	HEPTANE	1-OCTANOL	142825	111875	0.01	0.0056	0
			Test set				
45	OCTANE	1-OCTANOL	111659	111875	0.0223	0.0071	0
46	2-3-DIMETHYLBUTANE	1-PENTANOL	79298	71410	0.016	0.0153	0
47	2-METHYLPENTANE	1-PENTANOL	107835	71410	0.0134	0.0106	0
48	HEPTANE	1-PENTANOL	142825	71410	0.0255	0.0102	1
49	HEXANE	1-PENTANOL	110543	71410	0.0179	0.0097	0
50	HEXANE	1-PROPANOL	110543	71238	0.0238	0.0186	1

Table E.10: Regression results for mixtures of non-associating, dipolar with associating, dipolar components (group 8). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

Group 8 (nAdP - AdP)							
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
			Training set				
1	ETHANOL	ACETONITRILE	64175	75058	-0.0313	-0.0571	1
2	ACETONE	ANILINE	67641	62533	-0.0286	0.0079	0

3	DIETHYLAMINE	CHLOROBENZENE	109897	108907	-0.0446	-0.0313	0
4	WATER	SULFOLANE	7732185	126330	-0.0244	-0.0405	0
5	PHENOL	NITROBENZENE	108952	98953	-0.0151	-0.0406	0
6	ACETONE	METHANOL	67641	67561	-0.033	-0.0358	1
7	ETHANOL	TOLUENE	64175	108883	0.0332	0.0205	1
8	DIETHYLAMINE	ETHYL-ETHANOATE	109897	141786	-0.0036	0.0192	0
9	NITROMETHANE	1-HEXANOL	75525	111273	0.0078	-0.0264	0
10	METHYLAMINE	TRIMETHYLAMINE	74895	75503	-0.024	-0.0019	1
11	METHANOL	gamma-BUTYROLACTONE	67561	96480	-0.0212	-0.0334	0
12	CHLOROBENZENE	PROPIONIC-ACID	108907	79094	-0.0335	-0.0541	0
13	1-PROPANOL	N-METHYL-2-PYRROLIDONE	71238	872504	-0.0594	-0.0183	0
14	1-BROMOBENZENE	CYCLOHEXANOL	108861	108930	0.0337	0.0241	1
15	METHANOL	DIMETHYL-DISULFIDE	67561	624920	0.0275	-0.0005	1
16	ACRYLONITRILE	WATER	107131	7732185	-0.05	-0.0416	1
17	1-HEXENE	PROPIONIC-ACID	592416	79094	-0.0154	-0.0508	0
18	CHLOROBENZENE	ANILINE	108907	62533	-0.0051	-0.0508	0
19	METHYL-PROPANOATE	1-BUTANOL	554121	71363	0.006	-0.0015	0
20	METHANOL	METHYL-n-BUTYL-ETHER	67561	628284	-0.0355	-0.0089	1
20							
21	ACETONITRILE	1-PROPANOL	75058	71238	-0.0381	-0.0427	1
22	DIETHYLAMINE	TRIETHYLAMINE	109897	121448	-0.0381	-0.0059	0
23	METHYL-ACETATE	ETHANOL	79209	64175	-0.0327	-0.0157	0
24	DIISOPROPYL-ETHER	PROPYLENE-GLYCOL-MONOMETHYL-ETHER	108203	107982	-0.0035	-0.0067	0
25	1-PROPANOL	N-N-DIMETHYL-FORMAMIDE	71238	68122	-0.0593	-0.0302	0
26	ACETONITRILE	WATER	75058	7732185	-0.0485	-0.0349	1
27	ETHANOL	N-METHYL-2-PYRROLIDONE	64175	872504	-0.078	-0.0299	0
28	TOLUENE	ANILINE	108883	62533	-0.005	-0.0388	0
29	ACETONE	CYCLOHEXANOL	67641	108930	0.0031	-0.013	0
30	1-2-DICHLOROETHANE	1-PROPANOL	107062	71238	0.0361	0.0095	1
31	ETHANOL	N-N-DIMETHYL-ACETAMIDE	64175	127195	-0.078	-0.0302	0
32	TOLUENE	1-BUTANOL	108883	71363	0.0433	0.0268	1
33	METHANOL	ACETONITRILE	67561	75058	-0.0284	-0.0613	1
34	DIISOPROPYL-ETHER	1-BUTANOL	108203	71363	0.0132	0.0044	0
35	CYCLOHEXANONE	CYCLOHEXANOL	108941	108930	0.0049	0.0086	0
36	METHANOL	4-HEPTANONE	67561	123193	-0.0266	-0.0106	0
37	ACETONE	ETHANOL	67641	64175	-0.0376	-0.0256	1
38	ETHANOL	DI-n-BUTYL-ETHER	64175	142961	0.0022	0.0026	0
39	METHANOL	1-2-DICHLOROETHANE	67561	107062	0.0039	-0.0086	1
40	ETHYL-PROPANOATE	PROPIONIC-ACID	105373	79094	-0.0745	-0.0542	0
41	TOLUENE	n-BUTYRIC-ACID	108883	107926	-0.0235	-0.0413	0
42	METHANOL	VINYL-ACETATE	67561	108054	-0.0255	-0.0214	1
43	DIETHYL-ETHER	ETHANOL	60297	64175	-0.0173	-0.0078	0
44	PROPYLENE	ETHANOL	115071	64175	0.0223	-0.008	0
45	ETHANOL	1-2-DICHLOROETHANE	64175	107062	0.0269	0.0017	1
46	DIETHYLAMINE	ETHYLBENZENE	109897	100414	-0.0539	-0.0236	0
47	PROPYLBENZENE	PHENOL	103651	108952	0.0234	0.0055	1

48	NITROMETHANE	1-BUTANOL	75525	71363	0.0137	-0.0214	1
49	METHANOL	BUTYL-ETHANOATE	67561	123864	-0.0514	-0.0117	0
50	DIETHYL-SULFIDE	1-PROPANOL	352932	71238	0.03	0.0168	1
51	CHLOROFORM	ETHANOL	67663	64175	-0.0032	-0.0014	1
52	DI-n-PROPYL-ETHER	1-OCTANOL	111433	111875	0.0043	-0.0028	0
53	THIOPHENE	PHENOL	110021	108952	0.0221	0.0006	0
54	1-2-DICHLOROETHANE	1-BUTANOL	107062	71363	0.0356	0.0121	0
55	DIISOPROPYL-ETHER	ETHANOL	108203	64175	-0.0009	-0.0032	1
56	DI-n-PROPYL-ETHER	1-BUTANOL	111433	71363	0.0122	0.0091	0
			Test set				
57	METHYL-BUTANOATE	1-BUTANOL	623427	71363	0.0065	0.002	0
58	ETHYLBENZENE	ANILINE	100414	62533	-0.008	-0.0408	0
59	DICHLOROMETHANE	ETHANOL	75092	64175	0.0101	-0.0153	1
60	ETHANOL	METHYL-PROPANOATE	64175	554121	-0.0033	-0.0122	1
61	TOLUENE	PHENOL	108883	108952	0.0181	0.0034	0
62	DIETHYLAMINE	TOLUENE	109897	108883	-0.0396	-0.0222	0
63	TOLUENE	p-CRESOL	108883	106445	0.0151	0.0029	0

Table E.11: Regression results for mixtures of non-associating, quadrupolar with associating, dipolar components (group 9). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

			Group 9 (nAqP - AdP)				
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
			Training set				
1	ETHYLENE	1-DECANOL	74851	112301	-0.0054	0.0067	0
2	CARBON-DIOXIDE	ACETIC-ACID	124389	64197	-0.0341	-0.0174	0
3	CARBON-DIOXIDE	METHANOL	124389	67561	-0.0081	-0.0017	0
4	BENZENE	ACETIC-ACID	71432	64197	-0.0565	-0.072	0
5	NITROGEN	1-DECANOL	7727379	112301	0.153	0.0955	0
6	AMMONIA	BENZENE	7664417	71432	0.0416	0.0237	0
7	CARBON-DIOXIDE	n-DECANOIC-ACID	124389	334485	-0.0176	-0.0182	0
8	CARBON-DIOXIDE	1-BUTANOL	124389	71363	-0.0094	-0.0068	0
9	CARBON-DIOXIDE	METHYL-LACTATE	124389	547648	-0.0205	-0.049	0
10	NITROGEN	CYCLOHEXANOL	7727379	108930	0.1133	0.0621	0
11	BENZENE	PROPIONIC-ACID	71432	79094	-0.0638	-0.0375	0
12	DIETHYLAMINE	BENZENE	109897	71432	-0.0601	-0.0345	0
13	BENZENE	1-PROPANOL	71432	71238	0.0256	0.0247	1
14	NITROGEN	ETHANOL	7727379	64175	0.0355	0.0382	0
15	CARBON-DIOXIDE	PHENOL	124389	108952	-0.0041	0.0205	0
16	BENZENE	ANILINE	71432	62533	-0.0328	-0.0427	0
17	CARBON-DIOXIDE	ETHANOL	124389	64175	-0.0112	-0.0029	0
18	NITROGEN	1-OCTANOL	7727379	111875	0.1203	0.0665	0
19	CARBON-DIOXIDE	ETHYLENE-GLYCOL-MONOPROPYL-ETHER	124389	2807309	-0.0179	0.0119	0
20	CARBON-DIOXIDE	1-DECANOL	124389	112301	0.0146	-0.0046	0
21	NITROGEN	1-PROPANOL	7727379	71238	-0.0154	0.0361	0

22	CARBON-DISULFIDE	1-PENTANOL	75150	71410	0.0114	0.027	0
23	ETHANOL	BENZENE	64175	71432	0.0165	0.0248	1
24	ISOPROPANOL	BENZENE	67630	71432	0.0444	0.0432	1
25	BENZENE	PHENOL	71432	108952	0.0034	0.0004	0
26	CARBON-DIOXIDE	1-HEPTANOL	124389	111706	0.0092	-0.0019	0
27	CARBON-DISULFIDE	1-BUTANOL	75150	71363	0.0265	0.0418	0
28	NITROGEN	2-HEXANOL	7727379	626937	0.0818	0.053	0
29	CARBON-DIOXIDE	1-PROPANOL	124389	71238	-0.0194	-0.0058	0
30	CARBON-DIOXIDE	1-PENTANOL	124389	71410	-0.0012	0.0007	0
31	CARBON-DIOXIDE	1-OCTANOL	124389	111875	0.0065	-0.0029	0
			Test set				
32	CARBON-DIOXIDE	1-HEXANOL	124389	111273	0.001	-0.001	0
33	CARBON-DISULFIDE	1-PROPANOL	75150	71238	0.0293	0.0412	0
34	CARBON-DIOXIDE	2-METHYL-1-BUTANOL	124389	137326	0.0143	-0.004	0
35	CARBON-DIOXIDE	n-DODECANOIC-ACID	124389	143077	-0.0069	-0.0151	0
36	CARBON-DIOXIDE	p-CRESOL	124389	106445	-0.0032	0.0164	0

Table E.12: Regression results for mixtures of two associating, dipolar components (group 10). For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column. The mixtures are listed separately for the training and the test set.

			Group 10 (AdP - AdP)				
#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
			Training set				
1	ACETIC-ACID	ACETAMIDE	64197	60355	-0.0993	-0.0734	0
2	CYCLOHEXYLAMINE	CYCLOHEXANOL	108918	108930	-0.1334	-0.1459	0
3	PROPIONIC-ACID	n-BUTYRIC-ACID	79094	107926	-0.0164	-0.0744	0
4	METHANOL	ANILINE	67561	62533	0.0558	0.0467	0
5	WATER	1-2-BENZENEDIOL	7732185	120809	-0.1136	-0.1507	0
6	PROPYLENE-GLYCOL-MONOMETHYL-ETHER	1-2-PROPYLENE-GLYCOL	107982	57556	-0.003	-0.0379	0
7	ISOPROPANOL	PROPYLENE-GLYCOL-MONOMETHYL-ETHER	67630	107982	0.0328	0.0265	0
8	p-CRESOL	1-2-BENZENEDIOL	106445	120809	-0.0486	-0.0633	0
9	CYCLOHEXYLAMINE	ANILINE	108918	62533	-0.1601	-0.1225	0
10	METHANOL	1-DECANOL	67561	112301	-0.0193	-0.0186	0
11	ANILINE	ETHYLENE-GLYCOL	62533	107211	0.0608	-0.007	1
12	3-METHYL-2-BUTANOL	1-PENTANOL	598754	71410	0.0035	-0.0089	0
13	WATER	PROPIONIC-ACID	7732185	79094	-0.1378	-0.1047	1
14	AMMONIA	DIETHYLAMINE	7664417	109897	0.0827	0.0438	0
15	ACETIC-ACID	n-BUTYRIC-ACID	64197	107926	-0.0513	-0.021	0
16	ACETIC-ACID	1-DODECANOL	64197	112538	-0.0451	0.0048	0
17	WATER	GLYCEROL	7732185	56815	-0.0679	-0.0845	0
18	PROPIONIC-ACID	PHENOL	79094	108952	-0.0625	-0.0471	0
19	n-BUTYLAMINE	1-BUTANOL	109739	71363	0.0025	0.0422	0
20	AMMONIA	HYDRAZINE	7664417	302012	0.0158	-0.0031	0
21	WATER	PROPYLENE-GLYCOL-MONOMETHYL-ETHER	7732185	107982	-0.094	-0.0334	1

22	ETHANOL	WATER	64175	7732185	-0.0997	-0.0665	1
23	ETHANOL	1-PROPANOL	64175	71238	0.0001	-0.0443	0
24	n-BUTYLAMINE	PIPERIDINE	109739	110894	0.0058	-0.0012	0
25	METHANOL	2-BUTANOL	67561	78922	-0.0443	-0.0339	0
26	WATER	ETHYLENE-GLYCOL	7732185	107211	-0.0546	-0.0951	0
27	WATER	2-ETHOXYETHANOL	7732185	110805	-0.0987	-0.0264	1
28	1-PROPANOL	WATER	71238	7732185	-0.0736	-0.0413	1
29	1-BUTANOL	PYRROLE	71363	109977	0.0081	0.0033	1
30	ISOPROPANOL	WATER	67630	7732185	-0.0958	-0.0264	1
31	METHANOL	WATER	67561	7732185	-0.0994	-0.0871	0
32	ETHANOL	2-METHYL-1-PROPANOL	64175	78831	0.0011	-0.0344	0
33	ISOBUTYRIC-ACID	n-PENTANOIC-ACID	79312	109524	-0.0091	-0.011	0
34	ISOPROPANOL	1-PROPANOL	67630	71238	0.0002	-0.0233	0
35	WATER	1-BUTANOL	7732185	71363	-0.0762	-0.0304	1
36	2-METHYL-1-BUTANOL	1-PENTANOL	137326	71410	-0.0015	-0.0215	0
			Test set				
37	2-BUTANOL	1-BUTANOL	78922	71363	0.0006	-0.0172	0
38	ETHANOL	1-BUTANOL	64175	71363	-0.0083	-0.0359	0
39	METHANOL	ETHANOL	67561	64175	-0.0156	-0.0717	0
40	ACETIC-ACID	PROPIONIC-ACID	64197	79094	-0.0297	-0.0316	0
41	1-PROPANOL	PYRROLE	71238	109977	0.0068	-0.0009	0

Table E.13: List of the binary mixtures identified as outliers. These mixtures are not included neither in the training nor in the test set. For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the group of mixtures they are categorized to (*group*). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column.

excluded mixtures (outliers)							
#	component i	component j	CAS-i	CAS-j	group	k_{ij}^{fit}	azeo
1	'CARBON-MONOXIDE'	'p-XYLENE'	630080	106423	1	0.0902	0
2	'ETHYL-METHANOATE'	'CYCLOHEXANE'	109944	110827	2	0.0686	1
3	'ETHYLENE'	'DODECANE'	74851	112403	3	-0.0412	0
4	'NITROGEN'	'ETHYLCYCLOHEXANE'	7727379	1678917	3	0.1241	0
5	'NITROGEN'	'HEXADECANE'	7727379	544763	3	0.1693	0
6	'CARBON-DIOXIDE'	'ACETONE'	124389	67641	4	-0.0045	0
7	'DICHLOROMETHANE'	'CARBON-DIOXIDE'	75092	124389	4	0.0089	0
8	'NITROGEN'	'CARBON-DIOXIDE'	7727379	124389	4	-0.0910	1
9	'CARBON-DIOXIDE'	'ETHYL-ETHANOATE'	124389	141786	4	-0.1641	0
10	'CARBON-DIOXIDE'	'ETHYL-METHANOATE'	124389	109944	4	-0.1446	0
11	'CARBON-DIOXIDE'	'ETHYL-PROPANOATE'	124389	105373	4	-0.1016	0
12	'CARBON-DIOXIDE'	'PROPYL-ETHANOATE'	124389	109604	4	-0.1413	0
13	'CARBON-DIOXIDE'	'PROPYL-PROPIONATE'	124389	106365	4	-0.0652	0
14	'CARBON-DIOXIDE'	'CARBON-DISULFIDE'	124389	75150	5	0.1145	0
15	'ACETONE'	'ACETIC-ACID'	67641	64197	6	-0.0898	0
16	'DIETHYL-ETHER'	'METHANOL'	60297	67561	6	-0.0602	1
17	'CHLOROFORM'	'PIPERIDINE'	67663	110894	6	-0.0845	0
18	'WATER'	'PROPYLENE-CARBONATE'	7732185	108327	6	-0.0097	0
19	'1-METHYLNAPHTHALENE'	'TRIETHYLENE-GLYCOL'	90120	112276	6	0.0416	0
20	'HYDROGEN-CHLORIDE'	'TRIFLUOROACETIC-ACID'	7647010	76051	6	0.0928	0

21	'ACETONE'	'WATER'	67641	7732185	6	-0.1167	0
22	'ETHYLENE-OXIDE'	'WATER'	75218	7732185	6	-0.0702	0
23	'TETRAHYDROFURAN'	'WATER'	109999	7732185	6	-0.0910	1
24	'HYDROGEN'	'1-OCTANOL'	1333740	111875	8	0.0514	0
25	'METHANE'	'1-PROPANOL'	74828	71238	8	-0.0373	0
26	'CARBON-MONOXIDE'	'ACETIC-ACID'	630080	64197	8	0.0662	0
27	'HYDROGEN'	'ACETIC-ACID'	1333740	64197	8	0.0850	0
28	'CYCLOHEXANE'	'ANILINE'	110827	62533	8	0.0255	0
29	'METHYLCYCLOHEXANE'	'ANILINE'	108872	62533	8	0.0078	0
30	'PHENOL'	'DODECANE'	108952	112403	8	0.0386	1
31	'CARBON-MONOXIDE'	'ETHANOL'	630080	64175	8	-0.0253	0
32	'HYDROGEN'	'METHANOL'	1333740	67561	8	-0.0259	0
33	'OCTANE'	'epsilon-CAPROLACTAM'	111659	105602	8	-0.0217	0
34	'WATER'	'3-HEPTANONE'	7732185	106354	10	-0.0644	0
35	'ETHYLAMINE'	'ETHANOL'	75047	64175	10	-0.1002	0
36	'AMMONIA'	'METHANOL'	7664417	67561	10	-0.1388	0
37	'FORMIC-ACID'	'PROPIONIC-ACID'	64186	79094	10	-0.1562	0
38	'AMMONIA'	'WATER'	7664417	7732185	10	-0.1768	0

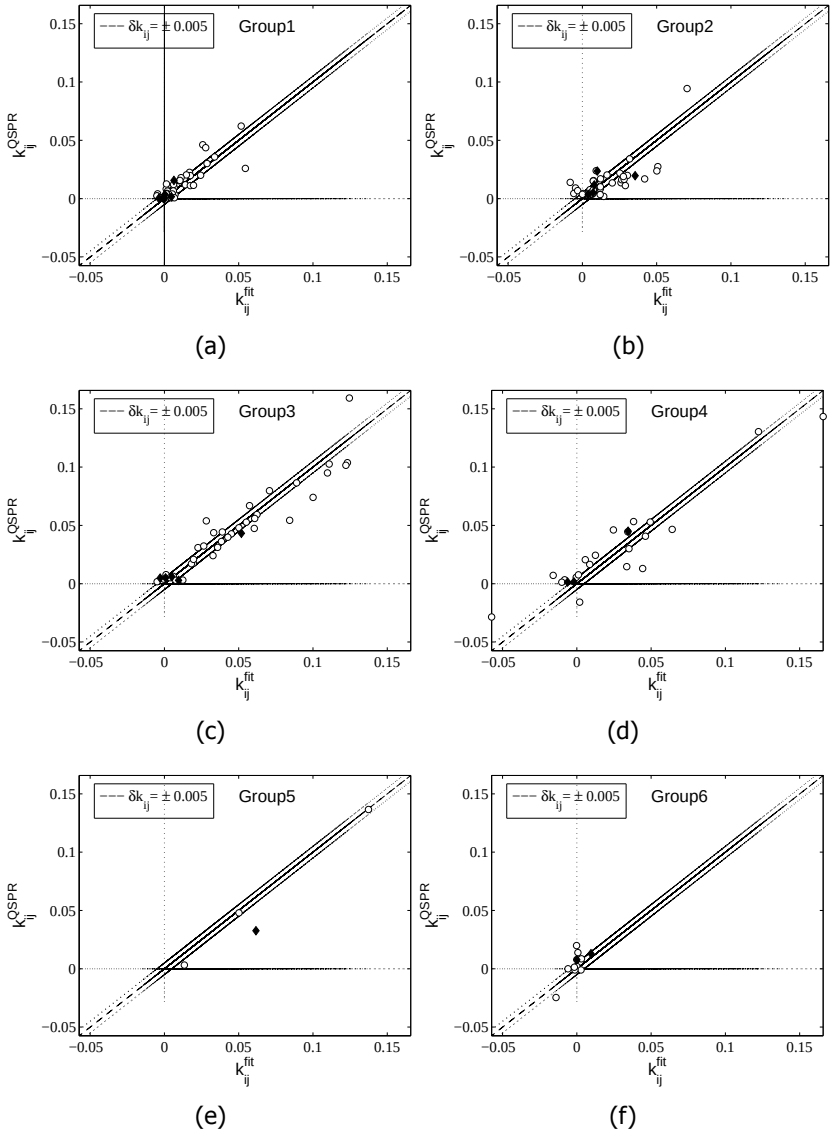


Figure E.1: Parity plots for the prediction in mixtures of non-associating components (groups 1 to 6). The fitted k_{ij} values (k_{ij}^{fit}) are plotted against the predicted k_{ij} values (k_{ij}^{QSPR}). The white circles correspond to the mixtures of the training set and the black diamonds correspond to the mixtures of the test set. The dashed lines designate the ± 0.005 absolute deviation on the value of (k_{ij}^{fit}).

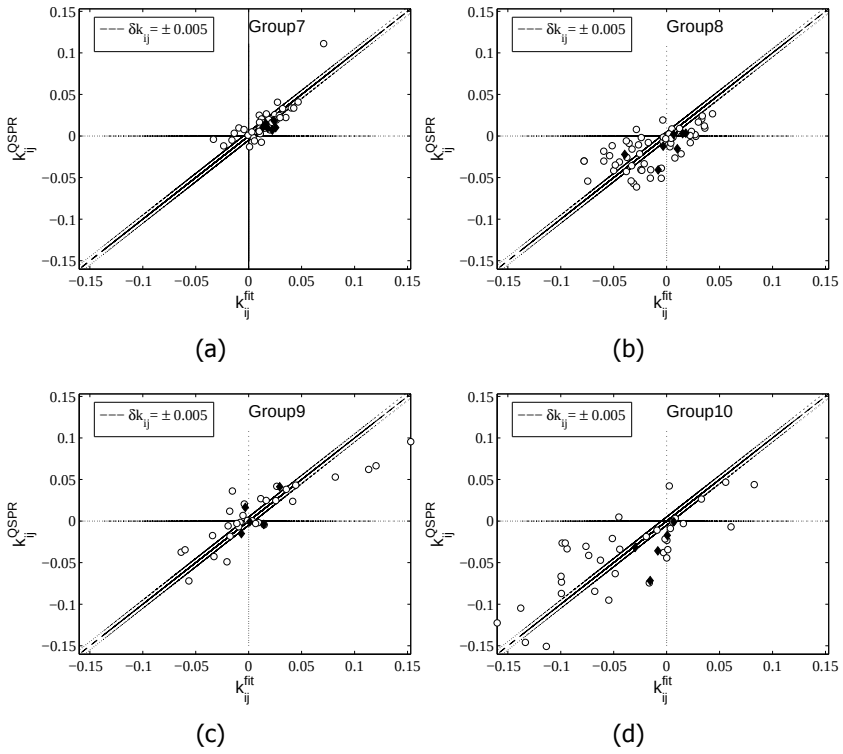


Figure E.2: Parity plots for the prediction in mixtures including associating components (groups 7 to 10). The fitted k_{ij} values (k_{ij}^{fit}) are plotted against the predicted k_{ij} values (k_{ij}^{QSPR}). The white circles correspond to the mixtures of the training set and the black diamonds correspond to the mixtures of the test set. The dashed lines designate the ± 0.005 absolute deviation on the value of (k_{ij}^{fit}).

E.3. Evaluation of phase equilibrium calculations with the estimated k_{ij} values

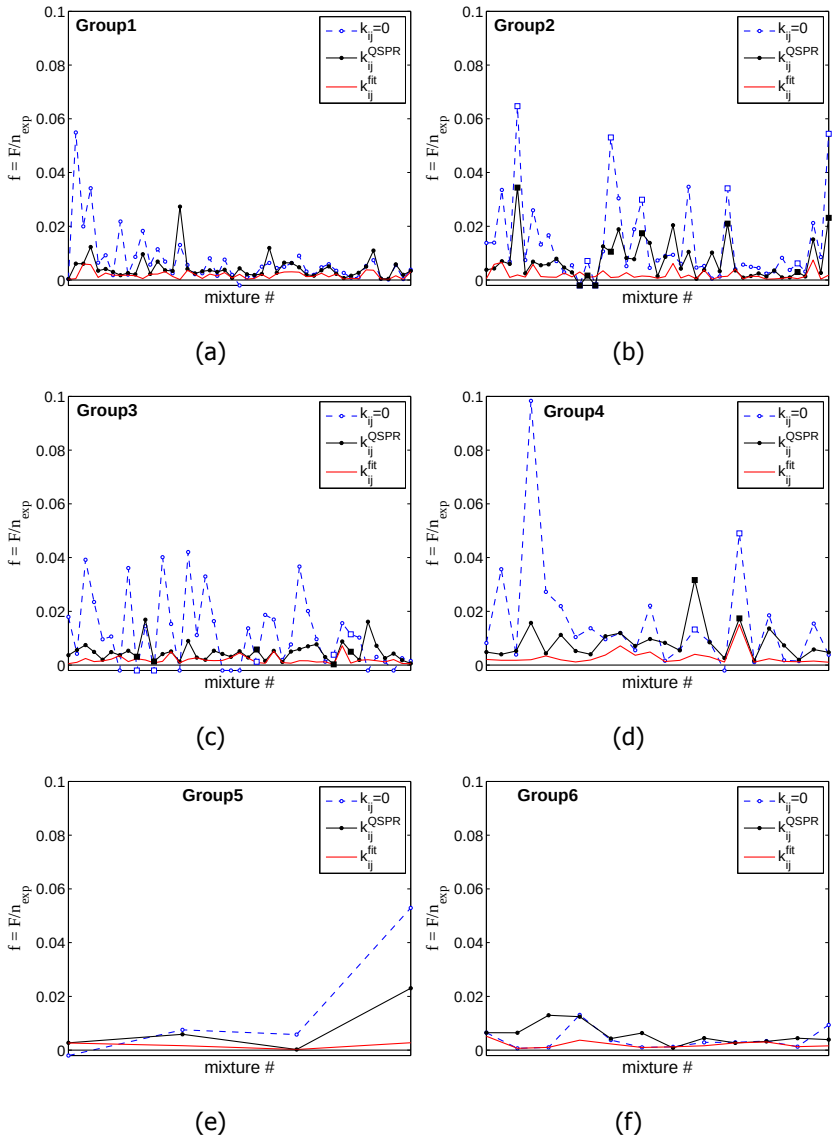


Figure E.3: Deviations of PC-SAFT from experimental binary phase equilibrium data for mixtures of non-associating components (groups 1 to 6). The deviations are defined as $f = F/n_{\text{exp}}$ according to Eq.(3.23). The red line serves as a reference and is obtained for individually optimized k_{ij}^{fit} values. The symbols (connected by black line) are obtained using estimated values k_{ij}^{QSPR}. The blue dashed line represent phase equilibrium calculations, when the PC-SAFT EoS is not corrected ($k_{ij} = 0$). The negative values indicate cases, when either the isobaric-isothermal flash calculations or the bubble point calculations did not converge for certain equilibrium data points. Azeotropic mixtures are marked with a square instead of a circle.

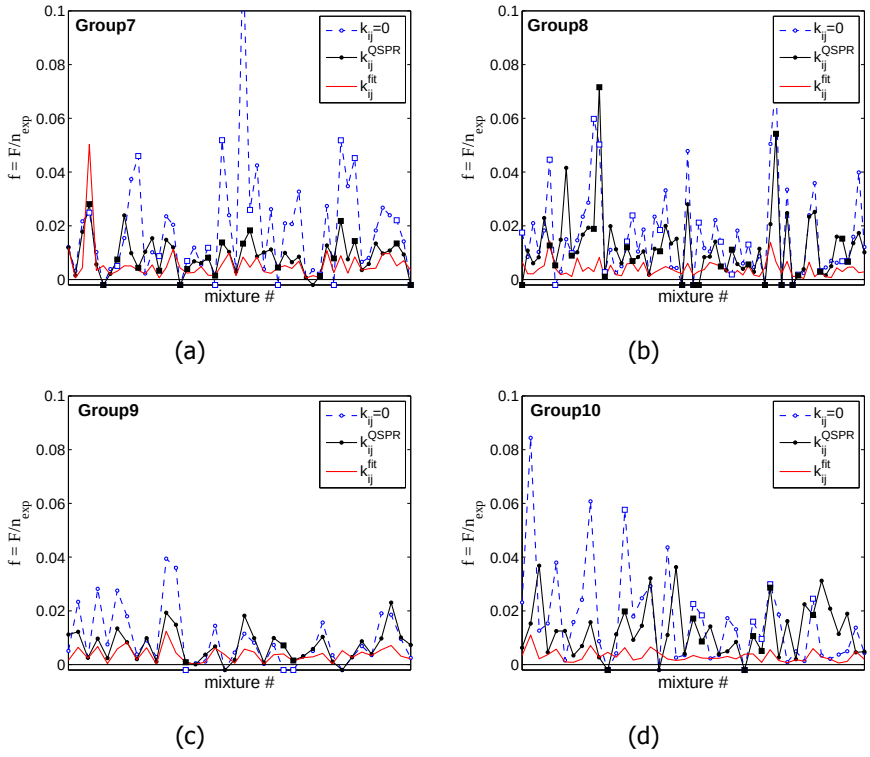


Figure E.4: Deviations of PC-SAFT from experimental binary phase equilibrium data for mixtures with at least one associating component (groups 7 to 10). The deviations are defined as $f = F/n_{\text{exp}}$ according to Eq.(3.23). The red line serves as a reference and is obtained for individually optimized k_{ij}^{fit} values. The symbols (connected by black line) are obtained using estimated values k_{ij}^{QSPR}. The blue dashed line represent phase equilibrium calculations, when the PC-SAFT EoS is not corrected ($k_{ij} = 0$). The negative values indicate cases, when either the isobaric-isothermal flash calculations or the bubble point calculations did not converge for certain equilibrium data points. Azeotropic mixtures are marked with a square instead of a circle.

E.4. Estimation of k_{ij} for binary mixtures with H_2

Table E.14: Results of the regressed model coefficients for mixtures of non-associating components with H_2. The values of the coefficients c_L and their 95% confidence interval $\Delta c_{L,95}$ for the QSPR model defined in Eq. (3.19) and (3.21), with $c^{\text{assoc,s}} = 0$, $c^{\text{assoc,c}} = 0$ and $c^{\text{assoc,sc}} = 0$.

mixtures of non-associating components with H_2		
	$c_L \cdot 10^4$	$\Delta c_{L,95} \cdot 10^4$
c^{LJ}	-21.671	± 1.031
$c^{\text{dd,a}}$	-1315.516	± 64.438
$c^{\text{dd,b}}$	0.000	± 0.000
c^{qq}	-0.766	± 0.355
c^{dq}	0.000	± 0

Table E.15: Regression results for mixtures of non-associating components with H_2. For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column.

#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
			Training set				
1	HYDROGEN	BICYCLOHEXYL	1333740	92513	0.3774	0.4919	0
2	HYDROGEN	PROPANE	1333740	74986	0.2133	0.1309	0
3	HYDROGEN	CYCLOHEXANE	1333740	110827	0.2901	0.2842	0
4	HYDROGEN	DECANE	1333740	124185	0.2610	0.2160	0
5	HYDROGEN	PENTANE	1333740	109660	0.2398	0.1841	0
6	HYDROGEN	TETRADECANE	1333740	629594	0.2848	0.2542	0
7	HYDROGEN	BUTANE	1333740	106978	0.2069	0.1623	0
8	HYDROGEN	HEXANE	1333740	110543	0.1511	0.1971	0
9	HYDROGEN	1-METHYLNAPHTHALENE	1333740	90120	0.2633	0.3703	0
10	HYDROGEN	PROPYLENE-CARBONATE	1333740	108327	-0.7299	-0.5732	0
11	HYDROGEN	1-HEXENE	1333740	592416	0.1498	0.1030	0
12	HYDROGEN	CYCLOHEXANONE	1333740	108941	0.1460	-0.1330	0
13	HYDROGEN	BENZENE	1333740	71432	0.2007	0.2007	0
			Test set				
14	HYDROGEN	2-2-4-TRIMETHYLPENTANE	1333740	540841	0.3907	0.2756	0
15	HYDROGEN	HEXADECANE	1333740	544763	0.3022	0.2582	0
16	HYDROGEN	1-OCTENE	1333740	111660	0.1424	0.1250	0
17	HYDROGEN	CHLOROBENZENE	1333740	108907	0.2080	0.0772	0

E.5. Estimation of k_{ij} for binary mixtures with CO_2

Table E.16: Results of the regressed model coefficients for mixtures of non-associating dipolar with non-associating quadrupolar components. The values of the coefficients c_L are adjusted only for mixtures of group 4 (Table E.17). The 95% confidence interval $\Delta c_{L,95}$ for the QSPR model defined in Eq. (3.19) and (3.21), with $c^{\text{assoc,s}} = 0$, $c^{\text{assoc,c}} = 0$ and $c^{\text{assoc,sc}} = 0$.

mixtures of non-associating dipolar with non-associating quadrupolar components		
	$c_L \cdot 10^4$	$\Delta c_{L,95} \cdot 10^4$
c^{LJ}	-118.756	± 0.261
$c^{\text{dd,a}}$	-33.984	± 1.373
$c^{\text{dd,b}}$	0.000	± 0.000
c^{qq}	-0.300	± 0.003
c^{dq}	-0.132	± 0.007

Table E.17: Regression results for mixtures of a non-associating dipolar with a non-associating quadrupolar component (group 4) for the QSPR model coefficients adjusted only to mixtures of this group. For each mixture the component names are given (*component i, component j*), their CAS-Numbers (*CAS-i, CAS-j*), the values of k_{ij} fitted to experimental VLE data (k_{ij}^{fit}) and the values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}). The mixtures that demonstrate azeotropic behavior are noted with "ones" in the last column.

#	component i	component j	CAS-i	CAS-j	k_{ij}^{fit}	k_{ij}^{QSPR}	azeo
			Training set				
1	NITROGEN	PROPYLBENZENE	7727379	103651	0.1658	0.1527	0
2	CARBON-DIOXIDE	PROPYLENE-CARBONATE	124389	108327	0.0314	0.0010	0
3	BUTYLBENZENE	BENZENE	104518	71432	-0.0101	-0.0026	0
4	CARBON-DIOXIDE	ACETYLACETONE	124389	123546	0.0502	0.0637	0
5	CARBON-DIOXIDE	BUTYLBENZENE	124389	104518	0.0473	0.0328	0
6	SULFUR-DIOXIDE	BENZENE	7446095	71432	0.0494	0.0373	0
7	CARBON-DIOXIDE	BENZALDEHYDE	124389	100527	0.0355	0.0230	0
8	PROPYLENE	NITROGEN	115071	7727379	0.0642	0.0484	0
9	CARBON-DIOXIDE	n-HEXYL-ACETATE	124389	142927	0.0004	0.0043	0
10	CARBON-DIOXIDE	1-BROMOBENZENE	124389	108861	0.0462	0.0447	0
11	VINYL-ACETATE	ETHYLENE	108054	74851	-0.0159	-0.0040	0
12	ACETONE	NITROGEN	67641	7727379	0.0336	0.0354	0
13	CARBON-DIOXIDE	gamma-VALEROLACTONE	124389	108292	0.0466	0.0243	0
14	CARBON-DIOXIDE	DIETHYL-OXALATE	124389	95921	-0.0314	-0.0023	0
15	CARBON-DIOXIDE	1-PENTENE	124389	109671	0.0276	0.0082	0
16	CARBON-DIOXIDE	1-METHYLNAPHTHALENE	124389	90120	0.0422	0.0604	0
17	CARBON-DIOXIDE	METHYL-BENZOATE	124389	93583	0.0096	0.0179	0
18	DIETHYL-ETHER	CARBON-DISULFIDE	60297	75150	0.0057	0.0155	1
19	CARBON-DIOXIDE	ETHYL-BENZOATE	124389	93890	0.0212	0.0255	0
20	CARBON-DIOXIDE	METHYL-DODECANOATE	124389	111820	0.0210	0.0105	0
21	CARBON-DIOXIDE	STYRENE	124389	100425	0.0260	0.0329	0
22	CARBON-DIOXIDE	gamma-BUTYROLACTONE	124389	96480	0.0184	0.0221	0
23	CARBON-DIOXIDE	PHENYL-ACETATE-	124389	122792	-0.0077	0.0092	0
24	PROPYLENE	BENZENE	115071	71432	0.0087	0.0136	0
25	CARBON-DIOXIDE	DIETHYL-PHTHALATE	124389	84662	0.0050	0.0071	0
26	CARBON-DIOXIDE	DECANAL	124389	112312	-0.0080	0.0058	0
27	BENZENE	1-BROMOBENZENE	71432	108861	-0.0082	-0.0037	0
28	CARBON-DIOXIDE	o-XYLENE	124389	95476	0.0343	0.0323	0
29	CARBON-DIOXIDE	CYCLOHEXANONE	124389	108941	0.0157	0.0273	0
30	NITROGEN	TOLUENE	7727379	108883	0.1223	0.1402	0
31	CARBON-DIOXIDE	DIISOPROPYL-ETHER	124389	108203	0.0178	-0.0003	0
32	CARBON-DIOXIDE	ACETOPHENONE	124389	98862	0.0241	0.0227	0
33	CARBON-DIOXIDE	1-HEPTENE	124389	592767	0.0271	0.0133	0
34	CARBON-DIOXIDE	METHYL-tert-BUTYL-ETHER	124389	1634044	-0.0301	0.0037	0
35	CARBON-DIOXIDE	ISOPROPYL-ETHANOATE	124389	108214	0.0019	-0.0026	0
36	CARBON-DIOXIDE	CYCLOPENTENE	124389	142290	0.0312	0.0224	0
37	CARBON-DIOXIDE	CHLOROBENZENE	124389	108907	0.0351	0.0352	1
38	CARBON-DIOXIDE	N-METHYL-2-PYRROLIDONE	124389	872504	0.0163	0.0131	0
39	CARBON-DIOXIDE	NONANAL	124389	124196	0.0033	0.0026	0
40	CARBON-DIOXIDE	TOLUENE	124389	108883	0.0246	0.0311	0
41	1-HEXENE	BENZENE	592416	71432	0.0011	0.0044	0
42	CARBON-DIOXIDE	DIETHYLENE-GLYCOL-DIMETHYL-ETHER	124389	111966	-0.0287	0.0078	0

43	CARBON-DIOXIDE	1-OCTENE	124389	111660	0.0287	0.0109	0
44	CARBON-DIOXIDE	n-OCTYL-ACETATE	124389	112141	0.0294	0.0068	0
45	1-HEXENE	ETHYLENE	592416	74851	0.0443	0.0099	0
46	CARBON-DIOXIDE	2-METHYL-1-PENTENE	124389	763291	0.0028	0.0108	0
47	CARBON-DIOXIDE	DIETHYL-SUCCINATE	124389	123251	-0.0229	0.0012	0
48	CARBON-DIOXIDE	CUMENE	124389	98828	0.0432	0.0306	0
49	CARBON-DIOXIDE	1-HEXENE	124389	592416	0.0125	0.0091	0
			Test set				
50	BENZENE	CHLOROBENZENE	71432	108907	-0.0020	-0.0057	0
51	CARBON-DIOXIDE	CYCLOHEXYL-ACETATE	124389	622457	0.0073	0.0072	0
52	CARBON-DIOXIDE	ETHYLBENZENE	124389	100414	0.0345	0.0313	0
53	PROPYLENE	ETHYLENE	115071	74851	-0.0004	0.0015	0
54	CARBON-DIOXIDE	PROPYLBENZENE	124389	103651	0.0382	0.0339	0
55	BENZENE	TOLUENE	71432	108883	-0.0062	-0.0022	0
56	CARBON-DIOXIDE	m-XYLENE	124389	108383	0.0287	0.0304	0

References

[1] M. Stavrou, A. Bardow, and J. Gross, *Estimation of the binary interaction parameter k_{ij} of the PC-SAFT Equation of State based on pure component parameters using a QSPR method,* Fluid Phase Equilib. **416**, 138 (2016), special Issue: SAFT 2015.

[2] G. H. Hudson and J. McCoubrey, *Intermolecular forces between unlike molecules,* Trans. Faraday Soc. **56**, 761 (1959).

Curriculum Vitæ

Marina-Eleni Stavrou

28-07-1981 Born in Athens, Greece.

Education

1993–1999 Pre-university education (Gymnasium-Lyceum), Athens, Greece.

1999–2005 Diploma in Chemical Engineering

National Technical University of Athens

Thesis: Contribution to anti-corrosive protection of steel using environmental friendly pigments.

Supervisor: Prof. Panagiota Vassiliou

2005–2007 Master of Science in Process, Energy and Environmental Systems Engineering

Technical University of Berlin

Thesis: Optimization of parameter estimation in membrane bioreactors.

Supervisors: Prof. Günter Wozny, Dr. Harvey Arellano-Garcia, Dr. Anja Drews

2011–2015 PhD student

Delft University of Technology

Thesis: Integrated process and solvent design for CO_2 capture using Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD).

Promotors: Prof. Dr.- Ing. Joachim Gross, Prof. Dr.- Ing. André Bardow

Professional

2003	Internship Bp Hellas A.E., Athens, Greece Department of retail lubricants - Trading, control and performance
2004	Internship Bp Hellas A.E., Piraeus, Greece Department of marine lubricants- Trading and logistics
2007	Internship BASF SE, Ludwigshafen, Germany Department of Technical - Automation and control *Project:* Design and optimization of process control strategies for multi-product distillation columns.
2008−2011	Chemical engineer AGET Heracles (Lafarge) - Cement Division, Evoia, Greece. Milaki cement plant, Department of process optimization.
2011−2015	Scientific associate University of Stuttgart, Germany Institute of Thermodynamics and Thermal Process Engineering
since 2016	Chemical engineer BASF SE, Ludwigshafen, Germany Department of Digitalization RC, Scientific Modeling, Group of Chemical Process Modeling (ROM/CP)

Distinctions

2006	Scholarship holder, Deutscher Akademischer Austausch Dienst (DAAD).
2013	Best Poster - CAST Directors Award 2013 from the Computing and Systems Technology Division of AIChE for the work presented at the 2013 AIChE Annual Meeting in San Francisco, USA, November 2013.
2014	Second place award in the competition for best lecture in the 24^{th} European Symposium on Computer Aided Process Engineering (ESCAPE 24), Budapest, Hungary, June 2014.

List of Journal Publications

6. **Stavrou, M.**; Bardow, A.; Gross, J., *Estimation of the binary interaction parameter k_{ij} of the PC-SAFT Equation of State based on pure component parameters using a QSPR method*, Fluid Phase Equilib. **416**, 138 (2016)

5. **Lampe, M.**; Stavrou, M.; Schilling, J.; Gross, J.; Bardow, A., *Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT*, Comput. Chem. Eng. **81**, 278 (2015)

4. **Stavrou, M.**; Lampe, M.; Bardow, A.; Gross, J., *Continuous Molecular Targeting-Computer-Aided Molecular Design (CoMT-CAMD) for Simultaneous Process and Solvent Design for CO_2 Capture*, Ind. Eng. Chem. Res. **53**(46), 18029 (2014)

3. **Sauer, E.**; Stavrou, M.; Gross, J., *Comparison between a Homo- and a Heterosegmented Group Contribution Approach Based on the Perturbed-Chain Polar Statistical Associating Fluid Theory Equation of State*, Ind. Eng. Chem. Res. **53**(38), 14854 (2014)

2. **Lampe, M.**; Stavrou, M.; Bücker, H. M..; Gross, J.; Bardow, A., *Simultaneous Optimization of Working Fluid and Process for Organic Rankine Cycles Using PC-SAFT*, Comput. Chem. Eng. **53**(21), 8821 (2015)

1. **Stavrou, M.**; Lampe, M.; Bardow, A.; Gross, J., *Physically-based Thermodynamic Models in Integrated Process and Molecular Design*, Comput. Aided Chem. Eng. **33**, 67 (2014)

Acknowledgment

"As you set out for Ithaca hope the voyage is a long one,
full of adventure, full of discovery.
Laistrygonians and Cyclops, angry Poseidon—don't be afraid of them:
you'll never find things like that on your way as long as you keep your thoughts raised high,
as long as a rare excitement stirs your spirit and your body.
Laistrygonians and Cyclops, wild Poseidon—you won't encounter them
unless you bring them along inside your soul, unless your soul sets them up in front of you.
Hope the voyage is a long one.
May there be many a summer morning when, with what pleasure, what joy,
you come into harbors seen for the first time;
may you stop at Phoenician trading stations to buy fine things,
mother of pearl and coral, amber and ebony, sensual perfume of every kind—
as many sensual perfumes as you can; and may you visit many Egyptian cities
to gather stores of knowledge from their scholars.
Keep Ithaca always in your mind. Arriving there is what you are destined for.
But do not hurry the journey at all. Better if it lasts for years,
so you are old by the time you reach the island, wealthy with all you have gained on the way,
not expecting Ithaca to make you rich.
Ithaca gave you the marvelous journey. Without her you would not have set out.
She has nothing left to give you now.
And if you find her poor, Ithaca won't have fooled you.
Wise as you will have become, so full of experience, you will have understood by then what these Ithacas mean."

(C.P. Cavafy, Collected Poems. Translated by Edmund Keeley and Philip Sherrard. Edited by George Savidis. Revised Edition. Princeton University Press, 1992)

This dissertation thesis is somehow like "Ithaca"; a destination which gave me a great and beautiful journey, full of new knowledge and discoveries. But also it has been a long journey, a tedious one, and definitely, I could not have succeeded all alone.

Joachim Gross and André Bardow, first I would like to thank you both for giving me the chance to work on a fascinating and challenging research subject. Moreover, for that you have been my mentors during this scientific endeavor. Your passion about thermodynamics and chemical process engineering, your pioneering ideas and your modern, humanitarian and visionary way of thinking have been an inspiration and a major drive for this work.

Joachim, I would particularly like to thank you for your major contribution to this work. Thank you first for giving me the chance to work with you in Stuttgart, for showing me the charm of molecular thermodynamics, for sharing with me your ideas,

your enthusiasm and your (so far above average) deep understanding of thermodynamics and PC-SAFT. Thank you for your engagement in our work and for taking the time to work together with me on my problems. In times I thought I was reaching a dead end, you could break the problem down and guide me to a solution. Thank you also for showing me the beauty of mutual respect and openness in scientific collaborations. Your free spirit, humor, ethic and compassion make you a great teacher; your sharp way of thinking and understanding of the physical world an exceptional scientist!

André, you have been the true catalyst to this work. In times I was feeling lost in the detail, you could bring the ”big picture” up again. After our meetings, I always left with a plan, a new perspective and with my spirits up. You showed me that scientific research should meet high standards and how one should analyze and get the essence of research results. I am especially thankful for your encouragement to be a bold scientist, to try new ideas. Thank you for all the discussions, for sharing with me your precise and so elegant way of thinking and for your support.

Joachim, André, it has been a great honor and a great pleasure working with both of you!

Kay Damen, thank you for welcoming me into the CO_2-Catch Up project and for your support all along. Your knowledge and experience on the pre-combustion CO_2 capture process have been of great essence for the formulation of an industrially relevant process problem. Thank you also for keeping an open mind to our ideas, while managing the CO_2-Catch Up project.

Matthias Lampe, it has been a great pleasure working with you! Our discussions have always given me courage and hope, our mutual successes great joy. I would especially like to thank you for the openness in our collaboration and wish you, Simone and your beautiful daughter from my heart all the best.

Spyro Tzamaria, you are also one of the persons I have to specially thank, for you have always been someone I could look up to, a passionate scientist and a true humanist. You were the first to show me, that science is a passion. Besides, it was your example and your clear words that encouraged me to take my final decision to pursue a PhD abroad.

My beloved sister Maria, I know it took a great deal of patience to cope with me, when you first came to Stuttgart. Thank you for being there, when I needed someone to talk about my ”strange” problems. Your presence has been a comfort in most lonely times.

Here, I would also like to thank all the members of the ITT group in Stuttgart. Florian Drunsel, Andrea Hemmen and Christoph Klink, your support, from the very beginning, made my everyday life in Stuttgart much easier and more pleasant. Thank you for your advice, patience and your immense willingness to help me. Thank you for trusting me as a colleague and as a friend and for sharing with me important

moments of your private lives. Elmar Sauer, a special thank to you for your great work during your master thesis, which helped us implement the group contribution PC-SAFT for the CoMT-CAMD method. Thank you also for the most interesting extracurricular discussions. Oliver, Gerni, Domi, Vassili, thank you for showing me the funny side of life in Stuttgart. Thank you all, older and younger members of the ITT-group, for the discussions and the funny times we enjoyed in and outside ITT. Last but not least, Dieter Hoehn and Mrs Agnes Schmidtgen, thank you for all your practical support and your constant, tireless efforts to organize our everyday life at ITT.

However, the completion of this thesis continued for some time after I left Stuttgart. Lawien Zubeir, thank you for your willingness and help by translating the summary of this thesis to dutch. Jan-Oliver Weidert, thank you for all your support and for all our post-ITT discussions on thermodynamics and process simulation, which helped me greatly to work out part of the propositions accompanying this thesis. Nicole and Benjamin Bruns, thank you for taking the time to go through this text and thank you for all the kindness you have showed me, since I moved to Mannheim.
My dear Frank, thank you for your support and for being a beautiful and inspiring part of my life since the last couple of months.

And many thanks should go to all my friends, those back in Greece and those who live abroad. Yasaman Mirfendereski, thank you for our friendship along the past ten years, which has given us great times long after Berlin, like during the visits in Amsterdam before or after my meetings in TU-Delft. You have always been a caring friend and an inspiration for me, in both my professional and personal life! Francisco Mendoza and Omar Perez, thank you both for dragging me out of my simulations from time to time. Thank you for your visits in Stuttgart, our serious talks, the hilarious moments and the great parties! Christo, Stathi, Taso, Kosta, Oresti, Nikoleta, Eleni, Vassili, thank you all for making me feel at home, each time we meet! Thanks for the long hours we can still spend together and the discussions that unfold from everyday life problems and silly jokes to philosophical and political questions.

Here, finally, I would like to thank my mother; Mama, a ”thank you” is not enough. But, thank you!, for your unconditional support, for teaching me not to give up, for showing me the significance of compassion, hard work and free will. Without your devotion and personal sacrifices the last twenty five years, this work would not have been possible.

Marina
Mannheim, June 2017