

Thermophysical Properties of CO₂ with Impurities: Applications to Transportation and Injection

Thermophysical Properties of CO₂ with Impurities: Applications to Transportation and Injection

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus, prof. dr. ir. H. Bijl,
chair of the Board for Doctorates to be defended publicly on
Friday 19 February 2027 at 12:30 o'clock

by

Darshan RAJU

Master of Science in Mechanical Engineering,
Delft University of Technology, The Netherlands,
born in Coimbatore, India.

This dissertation has been approved by the (co)promoters.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. T.J.H. Vlugt,	Delft University of Technology, <i>promotor</i>
Dr. ir. M. Ramdin,	Delft University of Technology, <i>copromotor</i>

Independent members:

Prof. dr. ir. A.B. de Haan,	Delft University of Technology
Prof. dr. ir. R.A.W.M. Henkes,	Delft University of Technology
Dr. ir. T.M.J. Nijssen,	Delft University of Technology
Prof. dr. R. Pecnik,	Delft University of Technology
Prof. dr. J.P.M. Trusler,	Imperial College London, UK
Dr. O. Moulton,	Delft University of Technology, <i>reserve member</i>

This work was sponsored by NWO Domain Science for the use of super-computer facilities. The work presented herein is part of the European Network of Research Infrastructures for CO₂ Transport and Injection from the European Union's Horizon Europe Research and Innovation program under grant Number 101094664. The author acknowledges the use of computational resources of the DelftBlue supercomputer, provided by the Delft High Performance Computing Center (<https://www.tudelft.nl/dhpc>).



Keywords: Monte Carlo, Phase Equilibria, Thermophysical properties, Molecular Dynamics, and Machine Learning

Printed by: Ridderprint | <https://www.ridderprint.nl>

Cover by: Darshan Raju

Copyright © 2026 by D. Raju

ISBN 978-94-6518-464-7

An electronic version of this dissertation is available at
<https://repository.tudelft.nl/>

Contents

1. Introduction	1
1.1. CO ₂ Capture, Transport, and Storage	1
1.1.1. CO ₂ Transport and Storage Infrastructure	2
1.2. Flow Properties of CO ₂	4
1.2.1. Effect of Impurities	7
1.3. Modeling Thermophysical Properties of CO ₂	8
1.3.1. Thermophysical Properties from Equations of State . .	10
1.3.2. Thermodynamic Properties from Molecular Simulation	12
1.3.3. Transport Properties from Molecular Simulation . . .	13
1.4. Outline of this Thesis	14
2. Thermophysical Properties and Phase Behavior of CO₂ with Impurities	17
2.1. Introduction	18
2.2. Theoretical Background	22
2.3. Simulation Details	25
2.3.1. Thermodynamic Properties	27
2.3.2. Transport Properties	28
2.4. Results and Discussion	29
2.4.1. Thermodynamic Properties	30
2.4.2. Transport Properties	47
2.5. Conclusions	49
3. Vapor-Liquid Interfacial Properties of CO₂ Binary Mixtures	51
3.1. Introduction	52
3.2. Methodology	55
3.2.1. MD Simulations	55
3.2.2. Classical Density Functional Theory + PC-SAFT . . .	59
3.2.3. Parachor Model	60
3.2.4. Winterfeld–Scriven–Davis Model	61
3.2.5. Classical Nucleation Theory	62
3.3. Results and discussion	63
3.3.1. Pure CO ₂	65
3.3.2. Binary CO ₂ mixtures	70
3.3.3. Effect of Impurities on IFT	80
3.3.4. Computation of IFT from Parachor and WSD models .	84
3.3.5. Metastable Phase Boundaries of CO ₂ and CO ₂ -Binary Mixtures	87

3.4.	Conclusions	93
4.	ML Predictions of CO₂ Phase Equilibria & Interfacial Properties	97
4.1.	Introduction	98
4.2.	Methodology	103
4.2.1.	Classical Density Functional Theory + PCP-SAFT	103
4.2.2.	Winterfeld–Scriven–Davis Model	105
4.2.3.	Machine Learning Models	107
4.2.4.	Composition Space Sampling	110
4.2.5.	Model Evaluation	114
4.3.	Results and Discussion	115
4.3.1.	Composition Pool	115
4.3.2.	Effect of Impurities	118
4.3.3.	Correlation Analysis of Features	121
4.3.4.	Machine Learning Model Performance and Outlier Analysis	123
4.3.5.	Composition Space Sampling	133
4.3.6.	Hybrid Machine Learning Model for Interfacial Tension	136
4.4.	Conclusions	145
5.	Finite-size effects of the excess entropy	149
5.1.	Introduction	150
5.2.	Truncation of the Integral for S^{ex}	151
5.3.	Simulation Details	158
5.4.	Results and Discussion	162
5.5.	Conclusions	173
6.	Conclusions	175
A.	Appendix	179
A.1.	Supporting Information	179
A.2.	Data Availability	181
	Summary	217
	Samenvatting	219
	Curriculum Vitæ	223
	List of Publications	225
	Acknowledgements	227

1

Introduction

1.1. CO₂ Capture, Transport, and Storage

The scientific evidence for climate change is unequivocal. Human activities, primarily through greenhouse gas emissions, increased the global surface temperature by ca. 1.1 °C during 2011–2020 relative to 1850–1900 [1]. The main greenhouse gases are CO₂, CH₄, and N₂O. CO₂ is the dominant contributor to anthropogenic warming because of its large emission volume, persistence in the atmosphere, and cumulative warming effect, making carbon budgets central to climate policy [1, 2]. Several economies, including the United Kingdom (UK) and the European Union (EU), have substantially reduced emissions relative to 1990 while maintaining economic growth [3, 4]. Despite this progress, deep, rapid, and sustained emission reductions are needed to limit global warming to 1.5–2 °C relative to pre-industrial levels [1, 5]. Meeting these targets depends on a combination of mitigation strategies, particularly in sectors where emissions cannot be eliminated through conventional measures alone. In this context, Carbon Capture, Utilization, and Storage (CCUS) provides an important mitigation pathway for hard-to-abate sectors, where direct large-scale electrification or major process changes remain difficult [2, 6].

Conventional point-source capture separates CO₂ from concentrated exhaust streams produced by industrial facilities and power plants before it is released into the atmosphere [7–9]. Complementing this approach, direct air capture (DAC) technologies remove CO₂ that is already dispersed in the atmosphere [10, 11]. Figure 1.1 illustrates both conventional point-source capture and DAC as the two main CO₂ capture routes. The large-scale deployment of these technologies depends on safe and efficient infrastructure for CO₂ capture, conditioning, transport, and storage [6, 12].

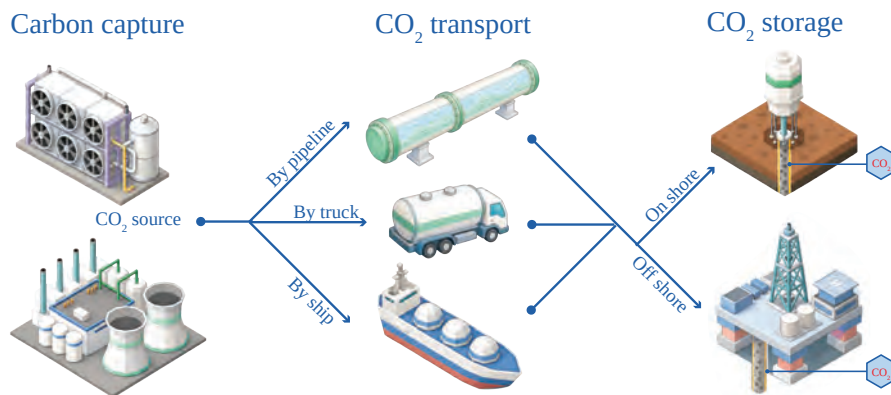


Figure 1.1. Schematic representation of the CCUS chain. CO₂ is captured either from concentrated emission sources, such as industrial facilities and power plants, or directly from the atmosphere using DAC. Following capture, the CO₂ may be transported by pipeline, truck, train, or ship to suitable onshore or offshore geological storage sites for permanent containment. Train transport is not shown in the figure for visual clarity.

1.1.1. CO₂ Transport and Storage Infrastructure

North America currently leads global CCUS deployment, with ca. 30 Mt per year of operating capture capacity [13]. This deployment is supported by policy incentives such as the 45Q tax credit [14]. This credit grants the operator a fixed amount for every tonne of CO₂ that is captured and permanently sequestered in a geological formation. The amount is up to 85 USD per tonne for CO₂ captured at an industrial facility or power plant, and up to 180 USD per tonne for CO₂ captured directly from the atmosphere. Europe is expanding rapidly as well, with several major projects concentrated around the North Sea [12, 15]. The Northern Lights project in Norway is the first open-access CO₂ transport and storage network in Europe [16, 17]. Injection began in 2025 with an initial capacity of ca. 1.5 Mt per year, and a second phase is expected to increase this capacity to at least 5 Mt per year by 2028 [16, 17]. In the Netherlands, the Porthos project will transport and store ca. 2.5 Mt of CO₂ per year for 15 years in the depleted gas fields beneath the North Sea, corresponding to ca. 37 Mt in total [18, 19]. Planned open-access infrastructure such as Aramis aims to further increase the Dutch CO₂ transport capacity to ca. 22 Mt per year [20]. Together with projects such as Greensand in Denmark [21] and the industrial clusters in the UK [22], these developments are expanding CO₂ storage capacity throughout Europe, supporting the EU target of at least 50 Mt per year of injection capacity by 2030 [15, 23]. Figure 1.2 illustrates this rapid expansion. The announced CO₂ injection capacity

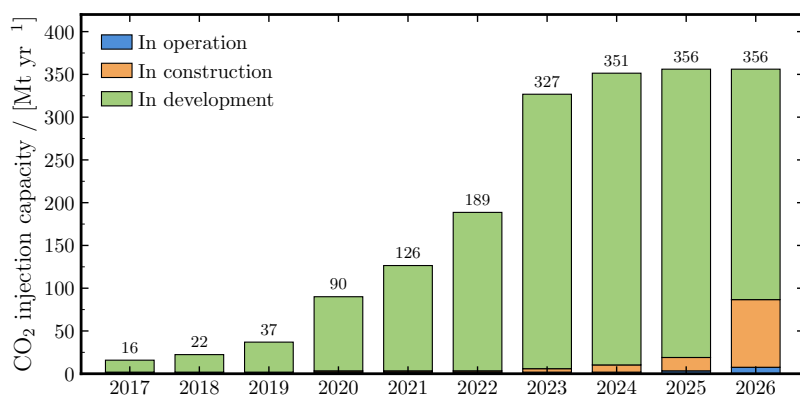


Figure 1.2. Evolution of the annual CO₂ injection capacity of European storage, transport and storage, and full-chain projects from 2017 to 2026. The projects are classified according to the development status. In 2026, ca. 7.6 Mt per year is in operation, ca. 79 Mt per year is in construction, and ca. 269 Mt per year is in development, out of an announced total of ca. 356 Mt per year. Canceled and suspended projects are excluded. Data were obtained from the 2026 edition of the IEA CCUS Projects Database [12].

of the European storage project pipeline has grown more than twentyfold within a decade, from ca. 16 Mt per year in 2017 to ca. 356 Mt per year in 2026 [12]. However, only ca. 7.6 Mt per year of this capacity is in operation in 2026, and a further ca. 79 Mt per year is in construction, underlining the gap between announced ambitions and realized infrastructure. Although primarily intended for CO₂ captured from industrial point sources, these networks could also provide transport and storage infrastructure for CO₂ removed from the atmosphere via DAC. Several DAC projects are being developed in European countries, including the Dreamcatcher [24] and Sizewell C [25] projects in the UK, as well as a planned facility near Kollsnes in Norway [26]. Dreamcatcher was proposed with a capture capacity of 0.5 Mt to 1 Mt of CO₂ per year, while the Kollsnes facility is planned to capture up to 0.5 Mt per year [24, 26]. These developments highlight the need for safe, reliable, and cost-effective CO₂ transport infrastructure.

Figure 1.1 shows a schematic of the CCUS chain, connecting the capture of CO₂ from industrial point sources and the atmosphere, via transport, to utilization or geological storage. Between capture and transport, the captured stream is conditioned by dehydration, purification, and compression or liquefaction, which is a step that is omitted in the schematic. Pipelines are the most established option for transporting large, continuous CO₂ flows, whereas ships, trains, and trucks offer greater flexibility in connecting geographically dispersed emitters to offshore storage sites. Currently, ship and pipeline transport are often combined. Northern Lights transports

liquefied CO₂ by ship to an onshore receiving terminal, followed by pipeline transport to the offshore storage site [17]. Similarly, the planned Aramis network will receive CO₂ by ship or onshore pipeline before transporting it through an offshore trunk pipeline to storage sites beneath the North Sea [20]. For economic pipeline transport, CO₂ is preferably maintained in a dense phase, either as a liquid or as a supercritical fluid, because its high density reduces the volumetric flow rate and improves transport efficiency. Dense-phase transport is commonly achieved by compressing the stream to ca. 85 bar to 150 bar, combined with appropriate temperature control to avoid two-phase conditions and the associated flow instabilities [27]. This approach has been used for several decades in North American CO₂ pipeline networks, many of which were developed to supply CO₂ for enhanced oil recovery [13, 27]. In sharp contrast, the emerging European networks are primarily designed to transport captured CO₂ to offshore geological storage sites beneath the North Sea [15]. In the Porthos project, CO₂ captured from industrial sources in and around Rotterdam is first collected in a common low-pressure onshore pipeline and subsequently compressed at the Maasvlakte before being transported in the dense phase through an offshore pipeline for injection into depleted gas fields beneath the seabed [19]. The Aramis project follows a similar shared-trunk-pipeline concept but additionally includes a terminal in Rotterdam for receiving liquefied CO₂ delivered by ship prior to compression and offshore pipeline transport [20].

1.2. Flow Properties of CO₂

Designing CO₂ transport and storage infrastructure that is safe, cost-effective, and efficient requires accurate knowledge of the thermophysical and flow behavior of the transported stream. The starting point for this description is the phase diagram of pure CO₂, shown in Figure 1.3(a). The sublimation, saturation, and melting curves separate the solid, liquid, and gaseous regions and meet at the triple point at 216.59 K and 5.18 bar. The saturation curve, which separates the liquid and gaseous regions, terminates at the critical point at 304.13 K and 73.77 bar, beyond which the liquid and the vapor are no longer distinguishable and the fluid is supercritical. The phase boundaries and thermodynamic properties shown in Figure 1.3 are computed with the multiparameter Equation of State (EoS) model of Span and Wagner [28], which reproduces the available experimental data of pure CO₂ to within the experimental uncertainty over the range of conditions relevant to CO₂ transport and storage. Within the supercritical region, liquid-like and gas-like states are commonly distinguished by the Widom line, a pseudo-boundary that extends the saturation curve beyond the critical point [29]. The Widom line is defined by the loci of extrema of response functions, such as the maximum of the isobaric heat capacity, the maximum of the isothermal compressibility, and the minimum of the

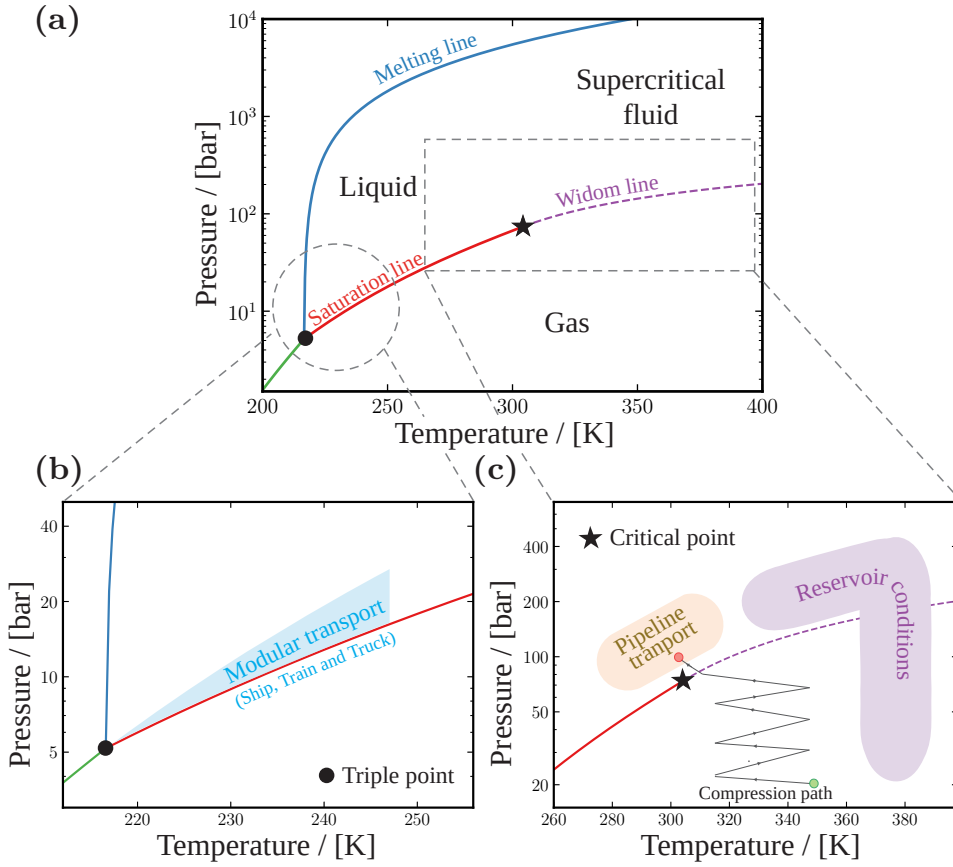


Figure 1.3. Pressure–temperature (PT) phase diagram of pure CO₂, computed using the Span–Wagner Equation of State (EoS) [28], showing typical operating windows for CO₂ transport and geological storage. (a) Complete phase diagram including the sublimation, saturation, and melting curves, the Widom line, the triple point at 216.59 K and 5.18 bar, and the critical point at 304.13 K and 73.77 bar. The Widom line is computed from the locus of maxima of the isothermal compressibility. (b) Enlarged view of the modular transport window for ships, trains, and trucks, where CO₂ is transported in the liquid phase. (c) High-pressure section of the transport and storage chain, including multistage compression with inter-cooling, dense-phase pipeline transport, and representative reservoir conditions for geological storage.

speed of sound [30, 31]. These loci converge at the critical point but diverge from one another as pressure increases, so that a Widom region rather than a single line is sometimes identified in the literature [32, 33]. In Figure 1.3(a), the Widom line is computed from the locus of maxima of the

isothermal compressibility. Crossing this boundary is associated with a rapid change in thermophysical properties, marking the transition between liquid-like and gas-like supercritical behavior [31, 34]. The transport and storage chain occupies distinct parts of this diagram, and each transport mode operates in a different region. Figure 1.3(b) shows the modular transport window used for ships, trains, and trucks in which CO₂ is carried as a liquid close to the saturation curve at temperatures of ca. 220 K to 250 K, which correspond to pressures of ca. 6 bar to 20 bar. Figure 1.3(c) shows the high-pressure part of the chain, in which the captured stream is compressed in stages with inter-cooling, transported in the dense phase at ca. 85 bar to 150 bar, and finally injected at reservoir conditions. During transport, injection, and storage, the CO₂ therefore passes through the gaseous, liquid, and supercritical regions of the phase diagram.

Operating in all three regions has direct consequences for the design of the transport chain. Supercritical CO₂ combines a liquid-like density with appreciable compressibility [28, 35]. Near the critical point, and particularly within the Widom region, small changes in temperature or pressure produce large variations in density, compressibility, and heat capacity [32, 33]. Such variations are important for valve sizing and selection. The fluid state at the valve outlet may differ considerably from that at the inlet, and pressure reduction through a valve can cause pronounced Joule–Thomson cooling and trigger a transition from single-phase to two-phase flow. Valve computations should therefore account for the fluid properties both upstream and downstream [35, 36]. The density, the Joule–Thomson coefficient, the speed of sound, and the Interfacial Tension (IFT) also directly influence the broader design and operation of CO₂ pipelines and injection wells [27, 35]. Density determines the volumetric flow rate for a given mass flow rate and, consequently, affects pipeline capacity, pressure drop, compression requirements, and transportation costs [35, 37]. Accurate density data and flow measurements are also essential for fiscal and custody-transfer applications because small measurement errors can result in considerable financial consequences [38, 39]. The Joule–Thomson coefficient determines the temperature change during isenthalpic expansion and is therefore important for evaluating the risks of low-temperature embrittlement and solid CO₂ formation during depressurization. The speed of sound governs the propagation of pressure and decompression waves and is an important input for transient-flow simulations and assessments of running ductile fracture [35, 37]. Finally, two-phase conditions may arise when rapid changes in pressure and temperature during startup, shutdown, depressurization, valve throttling, or accidental release cause the CO₂ stream to cross a phase boundary [37, 40]. Similar conditions may occur near the wellbore during injection into low-pressure reservoirs [41, 42]. At these conditions, the IFT and related interfacial properties govern droplet and bubble formation, which in turn affects flow regimes and the transfer of heat and mass between the phases [40, 41].

1.2.1. Effect of Impurities

In practice, captured CO₂ streams are never pure, and the type and amount of impurities depend on the point source and on the efficiency of the capture process [8, 43, 44]. Three main capture technologies are distinguished [8]. In post-combustion capture, CO₂ is separated from the flue gas after combustion in air, typically by chemical absorption in aqueous amine solvents [45], and the product stream mainly contains N₂, O₂, and H₂O [8, 43]. In pre-combustion capture, the fuel is first converted into synthesis gas from which CO₂ is removed before combustion, leading to impurities such as H₂, N₂, CO, CH₄, and H₂S [8, 43]. In oxy-fuel combustion, the fuel is burned in nearly pure oxygen, producing a CO₂-rich flue gas whose main impurities are O₂, N₂, Ar, H₂O, and SO_x/NO_x [43, 46]. These impurities are commonly divided into noncondensable gases, such as N₂, O₂, Ar, H₂, CH₄, and CO, which strongly affect the phase behavior and compression requirements, and condensable and acid-forming impurities, such as H₂O, H₂S, SO_x, and NO_x, which can cause corrosion, hydrate formation, and the formation of corrosive aqueous phases [43, 47]. Corrosion is driven by the appearance of a separate water phase [48, 49]. Dense CO₂ dissolves only a limited amount of water, and the solubility decreases further when noncondensable impurities are present, so that a free aqueous phase drops out once the water content exceeds the solubility limit [48]. This aqueous phase dissolves CO₂, SO₂, and NO₂ to form carbonic, sulfuric, and nitric acid, which attack the carbon steel of the pipeline wall [49, 50]. The combination of SO₂ and NO₂ is particularly damaging because the attack shifts from uniform corrosion to localized pitting [50]. The same separate water phase also allows gas hydrates to nucleate, which can block valves, small-bore instrument lines, and injection wells at temperatures well above 273.15 K, the melting point of ice at atmospheric pressure [51, 52]. Because of these integrity and safety risks, acid impurities are restricted to the ppm level in CO₂ transport specifications, as exemplified by the Porthos and Dynamis quality recommendations [19, 43, 46]. Even at very low concentrations, these impurities significantly alter the thermophysical properties of the CO₂ stream [44, 53]. Whereas pure CO₂ exhibits a single vapor-liquid saturation line in the pressure-temperature phase diagram, an impure stream shows a two-phase envelope bounded by a bubble-point curve, above which the mixture is entirely liquid, and a dew-point curve, below which it is entirely gaseous [35, 44]. Between these curves, vapor and liquid coexist, and the extent to which this two-phase region broadens depends on the type and concentration of the impurities [44, 53]. Noncondensable gases such as N₂ and H₂ widen the envelope considerably even at low concentrations [35, 44, 53]. In addition, impurities shift the critical point, reduce the density, and affect the Joule-Thomson coefficient, heat capacity, viscosity, and speed of sound [35, 44]. As a result, impure streams require higher operating pressures or larger pipelines for the same capacity, and the decompression behavior

of these streams differs from that of pure CO₂, which directly affects fracture control and safety assessments [35, 53]. The modeling of pure CO₂ transport in pipelines and wells at single-phase conditions is relatively well established, with reliable models available for describing thermodynamic properties and flow behavior [35, 54]. However, the uncertainty increases for CO₂ streams containing impurities, particularly when two-phase flow occurs [55]. Correlations for predicting flow regimes, pressure drop, and liquid holdup have not yet been sufficiently established or validated for CO₂-rich mixtures at conditions relevant to CCUS [55]. Rapid transient events, including startup, shutdown, depressurization, and accidental release, introduce further challenges because pressure and temperature may change faster than phase equilibrium can be established [56]. The resulting metastable states, delayed phase transitions, vapor collapse, and nonequilibrium heat and mass transfer can strongly affect the pressure response, temperature evolution, and phase distribution [56, 57]. During these events, interfacial properties such as the IFT, together with thermodynamic and transport properties, influence bubble and droplet nucleation, growth, coalescence, breakup, and collapse [41, 57]. Accurate representation of the thermodynamic, transport, and interfacial behavior of the mixture is therefore essential [35, 57]. Thermodynamic properties are commonly computed using multiparameter mixture models such as the Groupe Européen de Recherches Gazières (GERG-2008) EoS [58], which was originally developed for natural gas mixtures, and its CCUS-oriented extension, EOS-CG [59]. The performance of these models depends on the property of interest, temperature, pressure, composition, phase region, and the quality and availability of experimental data for the binary subsystems present in the mixture [58, 59]. Transport and interfacial properties are not obtained directly from standalone EoS models and require additional correlations or dedicated models [35, 60]. The accuracy of these approaches is ultimately limited by the available experimental data, which remain scarce for many CO₂-rich mixtures at relevant transport and injection conditions [35, 44]. Multiparameter reference EoSs can reproduce the thermodynamic properties of well-characterized systems within experimental uncertainty, whereas semi-empirical transport and interfacial models are generally less accurate and less reliable beyond the conditions used in the model development [58–60].

1.3. Modeling Thermophysical Properties of CO₂

Experimental determination of the thermophysical properties of CO₂-rich mixtures for a broad range of compositions, temperatures, and pressures relevant to CCUS operations is technically demanding, time-consuming, and costly. Comprehensive measurement campaigns may span several years, making predictive models essential to complement the available

experimental data. Several classes of models have been developed, differing in theoretical foundation, accuracy, range of applicability, and computational cost [61]. Empirical correlations can represent individual properties of a specific fluid, such as the saturation pressure, viscosity, and IFT of pure CO₂, with high accuracy [62, 63]. The applicability of these correlations is generally limited to fluids and thermodynamic conditions for which sufficient experimental data are available [62–64]. EoSs provide a thermodynamically consistent framework for predicting a broad range of thermodynamic properties [61]. These models include cubic EoSs, such as the Peng–Robinson (PR) and Soave–Redlich–Kwong (SRK) models [65–67], molecular-based models within the Statistical Associating Fluid Theory (SAFT) family [68, 69], and multiparameter reference EoSs fitted to extensive experimental datasets [28, 58, 59]. Thermodynamic quantities obtained from an EoS can also serve as inputs to models for transport and interfacial properties, including classical Density Functional Theory (cDFT) [70], Density Gradient Theory (DGT) [71, 72], friction theory [73], entropy scaling [74, 75], and the Extended Corresponding States (ECS) model [76–78]. Molecular simulations based on Monte Carlo (MC) and Molecular Dynamics (MD) techniques instead predict thermodynamic, transport, and interfacial properties from a molecular-level description of the fluid [79, 80]. In classical molecular simulations, intermolecular interactions are represented by force fields, which describe the potential energy of the system as a function of the molecular positions and configurations [79, 80]. The force field parameters are tuned to reproduce experimental measurements, fitted to first-principles computations, or a combination of both [81, 82]. First-principles methods, such as Density Functional Theory (DFT), can also be applied directly without relying on predefined force fields. However, the high computational cost of DFT restricts accessible system sizes and simulation time scales, which limits its use for transport and interfacial-property computation that require large systems and long trajectories [83]. Machine-Learned Interatomic Potentials (MLIPs) address this limitation by learning the potential energy surface from electronic-structure data while retaining an accuracy close to that of the underlying first-principles method [84, 85]. However, the accuracy of an MLIP is fundamentally constrained by the reference method used to generate the training data [86]. Because experimental properties are generally not included directly in the fitting procedure, systematic errors in the underlying DFT computations are inherited by the resulting MLIP [87]. Nevertheless, DFT-based reference data are particularly valuable when chemical reactions, bond formation and breaking, or intramolecular charge redistribution must be described [88]. For the nonreactive CO₂-rich mixtures considered in this thesis, molecular simulations using established force fields are preferred because these methods enable longer simulations of larger systems and can be parameterized and validated with available

thermophysical property data. The following subsections briefly describe the modeling approaches used in this thesis.

1.3.1. Thermophysical Properties from Equations of State

Cubic EoSs such as SRK [66, 67] and PR [65], both descendants of the equation of van der Waals [89], remain the default models in commercial process simulators, since only the critical temperature, the critical pressure, and the acentric factor of each component are required and the volume roots follow analytically [36, 90]. The accuracy of these models depends directly on experimental data, because the attractive term is fitted to pure-component vapor pressures and every component pair requires a binary interaction parameter regressed from experimental Vapor–Liquid Equilibrium (VLE) data, which are scarce for CO₂-rich mixtures at transport conditions [44, 91]. Liquid densities are reproduced only within several percent unless an empirical volume translation is applied [92], and second-derivative properties such as the heat capacity, the speed of sound, and the Joule–Thomson coefficient are described considerably less accurately than the phase behavior [93, 94]. Deviations are largest near the critical point, which is precisely where dense-phase CO₂ pipelines operate, so cubic EoSs describe the phase envelopes of CO₂ mixtures with N₂, O₂, Ar, CH₄, and H₂ less accurately than SAFT-type and multiparameter models [44, 95]. Higher accuracy over wide ranges of temperature and pressure is obtained from modern multiparameter EoSs, which are formulated explicitly in the reduced Helmholtz energy [28, 61]

$$\frac{A(T, \rho)}{nRT} = \alpha(\tau, \delta) = \alpha^{\text{ig}}(\tau, \delta) + \alpha^{\text{r}}(\tau, \delta), \quad (1.1)$$

in which α^{ig} is the ideal gas contribution, α^{r} is the residual contribution accounting for intermolecular interactions, $\tau = T_c/T$ is the inverse reduced temperature, and $\delta = \rho/\rho_c$ is the reduced density. All thermodynamic properties, including the pressure, heat capacities, speed of sound, and Joule–Thomson coefficient, follow from analytic derivatives of α with respect to τ and δ . The ideal gas contribution is obtained by integrating the ideal gas heat capacity twice with respect to τ , which introduces two constants of integration [28, 61]. These two constants are arbitrary and do not affect any measurable property, because every thermodynamic property follows either from a derivative of α or from a difference between two states, in which the constants cancel. The constants only fix the zero points of the enthalpy and the entropy, and are assigned by convention through a selected reference state [28, 61]. The residual contribution is an empirical sum of polynomial, exponential, and Gaussian bell-shaped terms. The coefficients and exponents of these terms carry no physical meaning individually and are regressed simultaneously to large sets of experimental data of several properties, including $P\rho T$ data, vapor pressures, saturated liquid and vapor densities, speeds of sound, and heat capacities [28,

61]. For pure CO₂, the reference EoS of Span and Wagner [28] is valid from the triple-point temperature to 1100 K at pressures up to 8000 bar, with uncertainties in density as low as 0.03% in the region relevant for CO₂ pipeline transport. For mixtures, the multifluid Helmholtz approach combines the pure-component reference EoSs with binary-specific reducing and departure functions [58, 96]. The reducing functions define the composition-dependent reducing temperature and reducing density [58, 96]. The departure function accounts for nonideal interactions between unlike components that are not fully represented by the pure-component EoSs and the reducing functions [58, 96]. The GERG-2008 EoS [58], developed for natural gases, covers 21 components, while EOS-CG [59] reparameterizes the GERG-2008 framework for combustion gases and mixtures relevant to CCUS. A theoretically based alternative to the empirical multiparameter EoS is the SAFT family of EoSs, derived from Wertheim perturbation theory, in which molecules are modeled as chains of interacting segments [97–101]. Widely used variants include Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) [68, 102], its polar extension Perturbed-Chain Polar Statistical Associating Fluid Theory (PCP-SAFT) that explicitly accounts for the strong quadrupole moment of CO₂ [103, 104], and SAFT-Variable Range (VR) Mie [69, 105]. In the PCP-SAFT framework, the residual Helmholtz energy is constructed as a sum of physically motivated contributions, as shown schematically in Figure 1.4 and expressed as follows [68, 103]:

$$\frac{A}{nRT} = \frac{A^{\text{ig}}}{nRT} + \frac{A^{\text{hs}}}{nRT} + \frac{A^{\text{chain}}}{nRT} + \frac{A^{\text{disp}}}{nRT} + \frac{A^{\text{assoc}}}{nRT} + \frac{A^{\text{polar}}}{nRT}, \quad (1.2)$$

in which the ideal gas term (ig) follows directly from statistical mechanics and depends only on temperature, density, and the internal degrees of freedom of the molecules, while the remaining terms constitute the residual contribution arising from intermolecular interactions. The hard-sphere (hs) and chain (chain) terms together form the hard-chain reference contribution, describing the repulsion of molecules modeled as chains of tangentially bonded hard spheres [68, 106]. The dispersion term (disp) accounts for the attractive van der Waals interactions between these chains [68], the association term (assoc) captures strong, short-ranged directional interactions such as hydrogen bonding [102, 106], and the polar term (polar) accounts for dipolar and quadrupolar interactions [103, 104]. For CO₂, the explicit quadrupolar contribution of PCP-SAFT improves the description of the phase behavior of CO₂-rich mixtures without introducing additional adjustable parameters [103]. SAFT-type models require only a few molecular-based parameters per component, unlike multiparameter EoSs. This physically based formulation generally provides more reliable extrapolation to conditions and mixtures outside the fitted range, although the absolute accuracy is often lower than that of multiparameter reference models [60]. A further advantage of Helmholtz energy-based models is the extension to inhomogeneous systems using cDFT, DGT, or Square Gradient

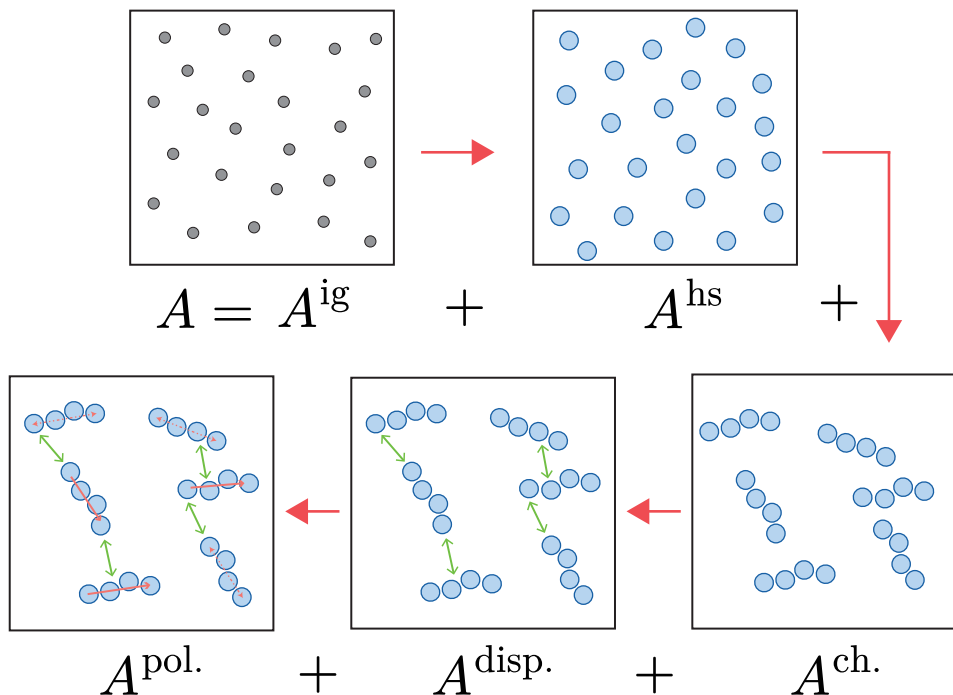


Figure 1.4. Perturbation scheme of PCP-SAFT, showing the contributions to the Helmholtz energy of Equation 1.2. Starting from the ideal gas (ig), the hard-sphere (hs) and chain (ch.) terms describe the repulsion of molecules modeled as chains of tangentially bonded hard spheres, the dispersion term (disp.) accounts for attractive van der Waals interactions between these chains, and the polar term (pol.) accounts for dipolar and quadrupolar interactions [68, 103, 106]. The association contribution is not depicted. Figure adapted from Refs. [68, 70, 107].

Theory (SGT) [70, 72, 108–111]. In cDFT, the grand potential is expressed as a functional of the spatially varying density profiles. Minimization of this functional yields the equilibrium density profiles that connect the coexisting bulk phases, from which interfacial properties such as the IFT, adsorption, and component enrichment can be determined [70, 71, 109]. In this way, cDFT predicts interfacial properties that are thermodynamically consistent with the underlying bulk EoS.

1.3.2. Thermodynamic Properties from Molecular Simulation

Molecular simulations predict macroscopic thermodynamic properties from a microscopic description of the fluid, in which intermolecular and intramolecular interactions are represented by a force field [79, 80]. The

predictive accuracy of a molecular simulation therefore depends directly on the ability of the selected force field to represent the molecular interactions of the fluid [112, 113]. Force fields that reproduce the experimental vapor–liquid equilibria and densities are available for CO₂ and for its most common impurities, such as N₂, Ar, H₂, and CH₄ [82, 113–116]. In MC simulations, molecular configurations are sampled from a specified statistical-mechanical ensemble using trial moves that are accepted or rejected according to the Metropolis criterion [117]. Thermodynamic properties are computed from ensemble averages or fluctuations of the sampled quantities. Depending on the selected ensemble, MC simulations are used to compute the density, internal energy, enthalpy, pressure, chemical potential, compressibility, thermal expansion coefficient, heat capacities, Joule–Thomson coefficient, and speed of sound [112, 118, 119]. Phase equilibria can be computed directly using specialized methods such as the Gibbs ensemble, in which molecular transfers and volume exchanges between two simulation boxes establish phase coexistence without requiring an explicit interface [120]. Advanced techniques such as Continuous Fractional Component Monte Carlo (CFCMC) improve the sampling of molecular insertions and deletions in dense fluids and multicomponent mixtures [121–125].

Thermodynamic and interfacial properties can also be computed using MD simulations, in which Newton's equations of motion are integrated numerically to generate molecular trajectories [79, 80]. Bulk properties such as the pressure, internal energy, enthalpy, and heat capacities can be obtained from time averages and fluctuations in the appropriate ensemble [79, 80, 118]. For vapor–liquid systems, the two coexisting phases can be simulated in a single elongated simulation box containing an explicit interface [126, 127]. The coexisting densities and phase compositions are obtained from the spatial density profiles, while the IFT is calculated from the difference between the normal and tangential components of the pressure tensor [126–129]. The interfacial thickness, adsorption, and preferential enrichment of impurities at the interface can also be determined from the density and composition profiles [126, 127]. Molecular simulations can therefore provide bulk, phase-equilibrium, and interfacial properties of impure CO₂ streams at conditions and compositions for which experimental data are limited. Standard textbooks provide detailed descriptions of molecular simulation methods [79, 80, 130–132].

1.3.3. Transport Properties from Molecular Simulation

In sharp contrast to thermodynamic properties, transport properties describe the dynamic response of a fluid to gradients in momentum, temperature, or chemical potential and therefore require the physical evolution of the molecular system over time. In MD simulations, intermolecular forces are evaluated at each time step and used to update the molecular positions and velocities, commonly using algorithms from the Verlet family, such

as the velocity-Verlet algorithm [79, 80, 133]. These algorithms are time-reversible and symplectic, so in the microcanonical ensemble (NVE) the total energy fluctuates around a constant value without systematic drift, provided that the integration time step is sufficiently small. Thermostats and barostats can be applied in other ensembles to maintain prescribed temperatures and pressures. Transport properties, including the self-diffusion coefficient, viscosity, and thermal conductivity, can be computed using either equilibrium or nonequilibrium MD. In equilibrium MD, the self-diffusion coefficient, viscosity, and thermal conductivity can each be obtained either from Green–Kubo integrals of the corresponding autocorrelation functions or from the equivalent Einstein–Helfand relations [79, 80, 134]. The relevant correlation functions are those of the particle velocities, the stress tensor, and the heat flux, respectively. In the Einstein form, these correspond to the mean-squared displacement of the positions and of the Helfand moments of momentum and energy. In nonequilibrium MD, an external perturbation is imposed, and the resulting steady-state response is measured directly [135–137]. Shear flow can be applied to estimate the viscosity, while a temperature gradient or an imposed heat flux can be used to compute the thermal conductivity [135–137]. Nonequilibrium methods often yield a stronger response signal than equilibrium methods [138, 139]. This advantage holds provided that the perturbation remains small enough to keep the system within the linear-response regime, and at the same time large enough to produce a measurable response. Entropy scaling offers an attractive complementary route to transport properties [75, 140]. The approach goes back to Rosenfeld [140] and rests on the finding that transport coefficients, once reduced appropriately, collapse onto nearly universal and monotonic functions of the excess entropy [74, 75, 141]. Since the excess entropy is accessible from an EoS, this framework links thermodynamic and transport behavior directly. Viscosity and thermal conductivity can thus be predicted by pairing entropy-scaling correlations with SAFT-type EoSs or multiparameter EoS [142, 143]. The same idea can be further exploited to reduce the computational cost of determining transport properties from MD simulations [144].

1.4. Outline of this Thesis

This thesis applies the modeling approaches introduced in Section 1.3 to CO₂-rich mixtures at transport conditions discussed in Section 1.2. The central aim is to provide accurate and computationally efficient descriptions of the thermodynamic, transport, and interfacial properties of impure CO₂ streams for the design and safe operation of CO₂ transport and storage systems. Experimental data for impure CO₂ remain scarce, and a systematic molecular-simulation dataset of thermodynamic and transport properties is therefore provided at conditions where measurements are lacking. The delayed phase change during rapid transient events cannot be described

with equilibrium models alone, and the metastability limits of CO₂-rich mixtures are therefore quantified from their computed interfacial properties. Physics-based interfacial models are too slow for repeated evaluation in flow simulations, and physics-informed Machine Learning (ML) surrogates are therefore developed that reduce the computational cost by several orders of magnitude, enabling their direct use in Computational Fluid Dynamics (CFD) and transient pipeline simulations. The two main error sources in computing excess entropy from radial distribution functions, the finite integration distance and the limited system size, are analyzed systematically, and the most reliable route to the thermodynamic limit is identified for small simulation systems. Together, the chapters of this thesis span the range from validated force fields and molecular-based EoSs to property surrogates that are fast enough for engineering-scale flow simulations. As a result, molecular-level property predictions can be combined with the engineering tools used in the design and safety assessment of CO₂ transport systems.

The thesis is organized in two parts. In the first part, molecular simulations are used to predict the thermodynamic, transport, and interfacial properties of impure CO₂ streams. In Chapter 2, the thermodynamic and transport properties of pure CO₂ and of CO₂-rich mixtures containing the noncondensable impurities N₂, Ar, H₂, and CH₄ are computed using force field-based MC and MD simulations. Both binary mixtures, with impurity concentrations between 1 and 10 mol%, and multicomponent mixtures representative of captured CO₂ streams are considered, covering the operating window of CO₂ pipeline systems (253.15 K to 313.15 K and pressures up to 200 bar). The computed densities, isothermal compressibilities, thermal expansion coefficients, heat capacities, Joule–Thomson coefficients, speeds of sound, and viscosities are compared with the GERG-2008 EoS. This comparison assesses both the predictive quality of the force fields used and the accuracy of the reference model outside the range of its experimental database. The influence of each impurity on the individual properties is quantified and shown to correlate with the critical temperature of the impurity. Chapter 3 shifts the focus from bulk properties to interfacial properties, which influence droplet and bubble formation, flow regimes, and interphase heat and mass transfer when transient pipeline operations cause the fluid to enter the two-phase region. The IFTs, interfacial thicknesses, and interfacial enrichment are computed for pure CO₂ and the binary mixtures CO₂–N₂, CO₂–Ar, CO₂–CH₄, and CO₂–H₂ using MD simulations and the PC-SAFT EoS combined with cDFT. The results are compared with predictions from the semi-empirical Parachor [145] and Winterfeld–Scriven–Davis (WSD) [146] models, which are widely used in engineering applications and process simulations. The importance of temperature-dependent Binary Interaction Parameters (BIPs) for accurately describing mixture interfaces is also demonstrated. The computed IFTs are then combined with Classical Nucleation Theory (CNT) to estimate the

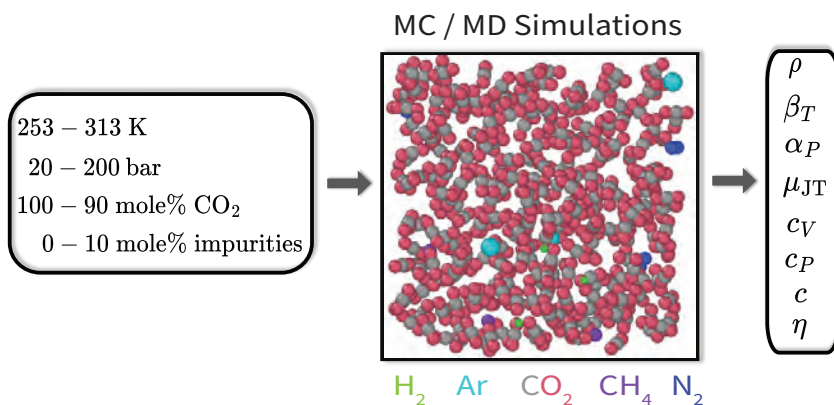
metastability limits of the mixtures, represented by the Superheat Limit (SHL) and the Supercool Limit (SCL). These limits indicate how far a fluid can be expanded beyond the phase boundary before phase change begins. The results help describe delayed phase change during fast transient events, such as startup, rapid depressurization, emergency shutdown, and sudden valve opening, and can serve as input for Computational Fluid Dynamics (CFD) simulations of pipeline decompression, relief-valve operation, blowdown processes, vessel rupture, two-phase flow, and safety-system performance.

The second part of this thesis focuses on making such property predictions computationally efficient and on the methodological foundations of the models used. In Chapter 4, the interfacial framework of Chapter 3 is extended to multicomponent mixtures and combined with Machine Learning (ML). The PCP-SAFT EoS combined with cDFT is used to generate a large dataset of bubble-point pressures, dew-point pressures, IFTs, and interfacial thicknesses of CO₂-rich mixtures containing between 95 and 99.99 mol% CO₂ and up to seven transportation-relevant impurities (N₂, Ar, CH₄, H₂, O₂, CO, and H₂S). Using this dataset, fast and uncertainty-aware ML surrogate models are developed for these four properties. Several fully data-driven classical ML models are first evaluated and compared. Active-learning strategies based on Gaussian Process Regression (GPR) are then applied to reduce the number of training points required from the PCP-SAFT EoS. In addition, hybrid physics-informed models are developed to learn the residuals of the WSD model, combining the accuracy of data-driven methods with the improved extrapolation behavior of a physics-based model. The resulting surrogates predict phase equilibria and interfacial properties at a small fraction of the computational cost of the underlying EoS-based computations. This substantial reduction in evaluation time enables the use of these surrogates in large-scale transport-process models and CFD simulations that require repeated property evaluations. Chapter 5 addresses an important methodological question related to the entropy-scaling approach discussed in Section 1.3. How accurately can excess entropy be computed from radial distribution functions obtained using molecular simulations? The chapter systematically examines the two main sources of error in the commonly used two-body expansion of excess entropy. These arise from truncating the integral at a finite distance and from finite-size effects in radial distribution functions computed using systems with a limited number of particles. Finite-size corrections for the radial distribution function are assessed together with extrapolation methods similar to those used for Kirkwood–Buff integrals. Based on this analysis, practical guidelines are proposed to accurately estimate excess entropy in small simulation systems. Chapter 6 summarizes the main conclusions of this thesis and presents directions for future research.

2

Thermophysical Properties and Phase Behavior of CO₂ with Impurities

This chapter is based on the following publication: D. Raju, M. Ramdin and T. J. H. Vlugt. 'Thermophysical Properties and Phase Behavior of CO₂ with Impurities: Insight from Molecular Simulations'. *Journal of Chemical & Engineering Data* 69 (2024), pp. 2735–2755.



2.1. Introduction

Climate change is being driven mostly by CO₂ emissions from the combustion of fossil fuels (oil, natural gas, and coal) for electricity production, transportation, heating, and other industrial applications [148–150]. The 2023 statistical review of world energy indicates that more than 80% of world energy consumption comes from fossil fuels [149, 151]. It is unfeasible to shift entirely to renewable energy resources on a very short time scale [152]. Carbon Capture and Storage (CCS) is the most popular and large-scale process used in industries to meet the anthropogenic CO₂ emission targets [7]. CCS is carried out in three different stages, namely, capture, transportation, and sequestration [8]. Carbon capture and sequestration depends upon safe and economical transportation of CO₂ [37]. In the past four decades, pipelines have been used to successfully inject pure CO₂ into depleted or nearly oil/gas fields for Enhanced Oil/gas Recovery (EOR) [150]. The injection of CO₂ captured from the flue gas stack is significantly different from pure CO₂ injection for EOR [35]. The difference is the presence of impurities, since no gas separation process is 100% efficient [153]. Impure CO₂ also differs in composition depending on the source and technology of capture [154]. It is possible to obtain pure CO₂ from impure CO₂, but this will result in additional costs and carbon footprint [44, 154]. The most efficient and preferred phase of transporting CO₂ in the pipeline is the dense supercritical or liquid phase [37]. Due to the presence of impurities, especially noncondensable gases (for example, Ar, N₂, H₂, O₂, and CH₄ which have a low boiling point compared to CO₂) reduces the density of impure CO₂ mixtures and is likely to introduce two-phase flow behavior [47]. Two-phase flows during transportation lead to numerous challenges, mainly pressure surge [155], which will sequentially lead to possible pipeline failure [156]. The recommended level of major impurities in CO₂ rich stream for safe and efficient pipeline transportation and sequestration from different standards (National Energy Technology Laboratory (NETL) [46], Dynamis [43], and International Standard Organization (ISO) [27]) and projects (Porthos [19] and Teeside [47]) are summarized in Tables 2.1 and 2.2. For quality standards of additional minor impurities in CO₂ transportation systems, the reader is referred to the review article of Simonsen, Hansen and Pedersen [27]. CO₂ rich stream with impurities significantly alters the thermodynamic and transport properties of CO₂, which will, in turn, impact the overall flow behavior, pipeline capacity, and operating window in CO₂ pipeline systems [157, 158]. Therefore, knowledge on thermodynamic and transport properties is indispensable to model the flow and phase behavior of impure CO₂ rich mixture within the operating window for the safe design and efficient operation of CO₂ transportation systems [35].

CO₂ captured from the stationary sources is compressed to a pressure higher than the critical pressure to avoid two-phase flows [37]. Due to planned maintenance or failure, transient processes such as startup,

Table 2.1. Description of all components used in this chapter. Long-range tail corrections for Lennard-Jones (LJ) interactions are used for all components.

Chemical name	Chemical formula	CAS number	Force field
Carbon dioxide	CO ₂	124-38-9	TraPPE [82]
Nitrogen	N ₂	7727-37-9	TraPPE [82]
Argon	Ar	7440-37-1	García-Pérez [116]
Hydrogen	H ₂	1333-74-0	Köster [115]
Methane	CH ₄	74-82-8	TraPPE [82]

shutdown, and depressurization are anticipated in CO₂ transportation systems, which plausibly lead to two-phase flows [37]. Hence, the operational window and conditions of CO₂ pipeline systems span a broad range of temperatures and pressures, encompassing transient processes from the wellhead to the bottom of the well. Operating conditions vary based on the geological location and reservoir characteristics [35]. High temperatures are limited considering the temperature limit of the pipeline coating material (< 50 °C) and cooling after compression stages [150]. The discharge pressure from the compressor to the pipeline is generally in the range of 100 to 200 bar [150]. The lowest temperature and pressure limit depends on geological conditions and phase behavior of CO₂ rich stream to maintain a dense phase [154]. Therefore, the temperature and the pressure range expected in CO₂ pipeline systems is assumed as −20 °C to 40 °C and 0 bar to 200 bar, respectively. The priori operational conditions considered in this chapter incorporate conditions at which transient events such as startup, shutdown, and depressurization are anticipated to occur [159]. Consequently, to ensure accurate modeling of transient processes, it is crucial to know the thermodynamic and transport properties for a wide range of temperatures and pressures expected within CO₂ transportation systems.

The thermodynamic and transport properties of impure CO₂ can be computed from thermodynamic models such as Equation of State (EoS) or other empirical correlations available in literature [28, 58, 62]. The validity of EoS predictions majorly depends on the interaction parameters that are obtained by fitting Vapor–Liquid Equilibrium (VLE) data obtained from experiments and assumptions used to develop EoS [160, 161]. Most EoS models accurately predict the thermodynamic properties related to first-order derivatives of the thermodynamic potentials (Gibbs energy, Helmholtz energy, enthalpy, and internal energy), i.e., the phase equilibria [94, 162]. The second-order derivative properties, such as isothermal compressibility, thermal expansion coefficient, Joule–Thomson coefficient, heat capacity, and speed of sound, are not predicted accurately by majority of EoS models [94, 162]. These properties serve as a basis for designing and

Table 2.2. CO₂ quality standards from the NETL [46], Dynamis [43], ISO [27], Porthos [19], and Teesside [47]. The impurity percentages with an asterisk (*) in the NETL [46] advised limits imply that the total impurity concentration should be $\leq 4\%$.

Reference		NETL [46]	Dynamis [43]	ISO 27913:2016 [27]	Porthos [19]	Teesside [47]
Component	CAS number	Concentration (in mol%)				
CO ₂	124-38-9	$\geq 95\%$	$> 95.5\%$	$\geq 95\%$	$\geq 95\%$	$\geq 95\%$
Ar	7440-37-1	4%*	$< 4\%$	-	$\leq 0.4\%$	1%
N ₂	7727-37-9	4%*	$< 4\%$	-	$\leq 2.4\%$	1%
H ₂	1333-74-0	4%*	$< 4\%$	-	$\leq 0.75\%$	1%
CH ₄	74-82-8	4%*	$< 4\%$	-	$\leq 1\%$	1%
O ₂	7782-44-7	4%*	$< 4\%$	-	40 ppm	10 ppm
CO	630-08-0	35 ppm	200 ppm	$< 2\%$	≤ 750 ppm	0.2 %
Total (Ar, N ₂ , H ₂ , CH ₄ , O ₂ , CO)	-	$\leq 4\%$	-	$\leq 4\%$	$\leq 4\%$	-

modeling pipeline transportation systems. Especially, knowledge on speed of sound is crucial in characterizing the state and structure of the fluid in pipeline transportation systems [163, 164]. Many literature studies predict the thermodynamic and transport properties of impure CO₂ using either a simple or an advanced EoS [64, 73], but no general agreement has been made to use a particular EoS with specific interaction parameters for CO₂ mixtures with small amount of impurities [158, 165].

Determination of the thermodynamic and transport properties from experiments is difficult due to low concentrations limits of impurities in CO₂ transportation systems (Table 2.2). Performing experiments for a wide range of compositions and conditions of a multicomponent CO₂ mixtures is costly and time-consuming [166]. Molecular simulations with classical force fields are widely used to compute the thermodynamic and transport properties of multicomponent gas systems [167–169]. Densities, viscosities, and phase equilibria calculated from molecular simulations provide reasonable and sometimes better predictions than EoS since calculations are based on accurate interaction potentials between atoms and molecules [168, 170, 171]. Simulations can efficiently compute heat capacities and speed of sound, both of which are essential for modeling transient phenomena like vapor collapse accurately [158, 163, 167, 172]. Using classical force field-based Monte Carlo (MC) simulations, Cresswell *et al.* [158] computed phase equilibria and densities of binary mixtures of CO₂ with Ar, N₂, H₂, and O₂ for a range of temperatures from 0 °C to 50 °C and pressure up to 200 bar. Aimoli, Maginn and Abreu [112] evaluated the performance of different force fields for computing density, isothermal compressibility, thermal expansion coefficient, heat capacity at constant volume and pressure, Joule–Thomson coefficient, viscosity, and speed of sound for pure CO₂ and CH₄ for a range of temperatures from –20 °C to 100 °C and pressure up to 1000 bar using Molecular Dynamics (MD) simulation. To the best of our knowledge, most molecular simulation studies are limited to binary CO₂ mixtures [112, 158]. Molecular simulations dedicated to multicomponent CO₂ mixtures are extremely limited [173, 174] or nonexistent.

In this chapter, we compute the thermodynamic and transport properties of pure and impure CO₂ streams for a range of impurity levels ranging from (1–10) mol% (which includes 12 CO₂ binary mixtures with 1 mol%, 5 mol%, and 10 mol% impurities which are shown in Section S14 of the Supporting Information for this chapter [147], and 36 multicomponent CO₂ mixtures with impurities ≤ 4 mol% which are shown in Sections S17 and S18 of the Supporting Information for this chapter [147]) using MC and MD simulation techniques. Simulations were carried out for a range of temperatures from –20 °C to 40 °C and pressure up to 200 bar. The main impurities, N₂, Ar, H₂, and CH₄, are selected to investigate the effect of impurities on the thermodynamic and transport properties. A comprehensive list of chemical components, CAS numbers, and force fields of all components used in this

chapter are shown in Table 2.1. Properties calculated within the operating window include density, isothermal compressibility, thermal expansion coefficient, heat capacity at constant volume and pressure, Joule–Thomson coefficient, shear viscosity, and speed of sound. We showed that, in comparison to pure CO₂ density, CO₂ rich mixture with molecular weight lower than pure CO₂ at a condition had lower densities. We also showed that impurities decrease the value of thermal expansion coefficients, isothermal compressibilities, heat capacities, and Joule–Thomson coefficients in the gas phase and increase the value of these properties in the liquid and supercritical phases. Conversely, impurities tend to increase speed of sound in the gas phase and decrease in the liquid and supercritical phases. Our results show that the order of influence of a particular impurity on a thermodynamic property other than density correlates with the critical temperature of that impurity.

This chapter is organized as follows. The methodology used to compute thermodynamic and transport properties from molecular simulations is described in Section 2.2, followed by the simulation details in Section 2.3. In Section 2.4, we present the validation and the results of the computed thermodynamic and transport properties. In Section 2.5, our key findings are summarized.

2.2. Theoretical Background

To compute the speed of sound (c), one requires other properties, which include heat capacity at constant pressure (C_P), heat capacity at constant volume (C_V), and isothermal compressibility (β_T) [163, 164]:

$$c(T, P) = \sqrt{\frac{vC_P(T, P)}{MC_V(T, V)\beta_T(T, P)}} \quad (2.1)$$

where v is the molar volume of the pure component or the mixture, and M is the molar mass of the pure component or the mixture. For a mixture, M can be calculated from the pure-component molar mass:

$$M = \sum_i^n x_i M_i \quad (2.2)$$

where n is the number of components present in the mixture, x_i and M_i are the mole fraction and molar mass of each component present in the mixture. To calculate C_V , C_P , and β_T , the derivatives of internal energy, volume, and enthalpy with respect to temperature and pressure have to be determined [118, 167].

$$C_V(T, V) = \left(\frac{\partial \langle U \rangle}{\partial T} \right)_V \quad (2.3)$$

$$C_P(T, P) = \left(\frac{\partial \langle H \rangle}{\partial T} \right)_P \quad (2.4)$$

$$\beta_T(T, P) = -\frac{1}{\langle V \rangle} \left(\frac{\partial \langle V \rangle}{\partial P} \right)_T \quad (2.5)$$

where U and H are the internal energy and enthalpy of the system, respectively. The angle brackets $\langle \dots \rangle$ denote an ensemble average, taken either in the NPT or in the NVT ensemble, which follows from the context. The internal energy and the enthalpy in Eq. 2.3 and Eq. 2.4 are defined as

$$U = U^{\text{internal}} + U^{\text{external}} \quad (2.6)$$

$$H = U^{\text{internal}} + U^{\text{external}} + K + PV \quad (2.7)$$

and therefore include the kinetic energy term (K) in addition to the potential energy contribution from intramolecular interactions (interactions inside molecules, denoted as U^{internal}) and intermolecular interactions (interactions between molecules, denoted as U^{external}). Hence, the C_V and C_P have been split into ideal and residual contributions. Following the work of Lagache *et al.* [118] we can write,

$$C_V(T, V) = C_V^{\text{ideal}}(T) + C_V^{\text{residual}}(T, V) = \left(\frac{\partial \langle U^{\text{ideal}} \rangle}{\partial T} \right)_V + \left(\frac{\partial \langle U^{\text{residual}} \rangle}{\partial T} \right)_V \quad (2.8)$$

$$C_P(T, P) = C_P^{\text{ideal}}(T) + C_P^{\text{residual}}(T, P) = \left(\frac{\partial \langle H^{\text{ideal}} \rangle}{\partial T} \right)_P + \left(\frac{\partial \langle H^{\text{residual}} \rangle}{\partial T} \right)_P \quad (2.9)$$

The ideal contributions $C_V^{\text{ideal}}(T)$ and $C_P^{\text{ideal}}(T)$ are the heat capacities that the system would have at the same temperature if it were an ideal gas. In an ideal gas, the molecules do not interact, so U^{ideal} and H^{ideal} contain only the kinetic energy and the intramolecular energy of isolated molecules and depend on temperature only. The ideal contributions are given by

$$C_V^{\text{ideal}}(T) = \left(\frac{\partial \langle U^{\text{ideal}} \rangle}{\partial T} \right)_V \quad (2.10)$$

$$C_P^{\text{ideal}}(T) = \left(\frac{\partial \langle H^{\text{ideal}} \rangle}{\partial T} \right)_P \quad (2.11)$$

and can be obtained from the standard thermodynamic databases [175, 176] or from quantum mechanical calculations. In this chapter, the latter approach was used to calculate $C_V^{\text{ideal}}(T)$ and $C_P^{\text{ideal}}(T)$ using the Gaussian09 software [177] with the B3LYP theory and a 6-31G(d,p) basis set. The derivatives $\frac{\partial \langle U^{\text{residual}} \rangle}{\partial T}$ and $\frac{\partial \langle H^{\text{residual}} \rangle}{\partial T}$ required to calculate C_V and C_P are

computed from fluctuations in NVT and NPT ensemble, respectively [118, 163, 164, 167].

$$C_V^{\text{residual}}(T, V) = \frac{1}{k_B T^2} \left(\langle U^{\text{external}} \hat{U} \rangle - \langle U^{\text{external}} \rangle \langle \hat{U} \rangle \right) \quad (2.12)$$

$$C_P^{\text{residual}}(T, P) = \frac{1}{k_B T^2} \left[\langle U^{\text{external}} \hat{H} \rangle - \langle U^{\text{external}} \rangle \langle \hat{H} \rangle + P(\langle V \hat{H} \rangle - \langle V \rangle \langle \hat{H} \rangle) \right] - N k_B \quad (2.13)$$

where k_B is the Boltzmann constant, N is the number of molecules in the system, $\hat{U}$ ($\hat{U} = U^{\text{internal}} + U^{\text{external}}$) is the configurational energy, and $\hat{H}$ ($\hat{H} = U^{\text{internal}} + U^{\text{external}} + PV$) is the configurational enthalpy of the system. The molar heat capacities (c_P and c_V) can be obtained from,

$$\frac{c_P}{C_P} = \frac{c_V}{C_V} = \frac{N_A}{N} \quad (2.14)$$

where N_A is Avogadro's number and N is the number of molecules in the system. The derivative $\left(\frac{\partial \langle V \rangle}{\partial P} \right)_T$ is required to calculate β_T defined in Eq. 2.5, and is computed using the following fluctuation formula [80],

$$\left(\frac{\partial \langle V \rangle}{\partial P} \right)_T = \frac{1}{k_B T} [\langle V^2 \rangle - \langle V \rangle^2] \quad (2.15)$$

Analogous to the computation of the speed of sound, the calculation of the Joule–Thomson coefficient requires heat capacity at constant pressure (c_P) and thermal expansion coefficient (α_P).

$$\mu_{JT} = \frac{V}{C_P} [T \alpha_P - 1] \quad (2.16)$$

The value of α_P is computed from the derivative of volume with respect to temperature,

$$\alpha_P(T, P) = \frac{1}{\langle V \rangle} \left(\frac{\partial \langle V \rangle}{\partial T} \right)_P \quad (2.17)$$

The derivative $\left(\frac{\partial \langle V \rangle}{\partial T} \right)_P$ which is required to calculate α_P is computed using [112, 118, 167],

$$\frac{\partial \langle V \rangle}{\partial T} = \frac{1}{k_B T^2} [\langle V \hat{H} \rangle - \langle \hat{H} \rangle \langle V \rangle] \quad (2.18)$$

It is important to note that C_V and α_P can be computed indirectly using the thermodynamic relations, [178]

$$C_V = C_P - \frac{T \langle V \rangle \alpha_P^2}{\beta_T} \quad (2.19)$$

$$\alpha_P = \sqrt{\left((C_P - C_V) \frac{\beta_T}{T \langle V \rangle} \right)} \quad (2.20)$$

The proofs of the mathematical derivation of equations to compute C_P , C_V , β_T , c , α_P , and μ_{JT} from simulations are provided in Sections S3-S8 of the Supporting Information for this chapter [147].

2.3. Simulation Details

All force field-based MC simulations were performed using the open-source software Brick, which uses the Continuous Fractional Component Monte Carlo (CFCMC) method [121–125] to calculate thermodynamics properties. Force field-based MD simulations were performed in the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) (version August 2023) package [179] to compute viscosities. All molecules were described with site-based conventional intermolecular potentials with point charges centered at the atom or at a dummy site. The pairwise-additive 12-6 Lennard-Jones (LJ) interaction potentials are used to model interactions:

$$u(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.21)$$

The Lorentz–Berthelot mixing rules were used to compute LJ interactions for dissimilar atoms [79, 80]:

$$\sigma_{ij} = \frac{1}{2} (\sigma_{ii} + \sigma_{jj}) \quad (2.22)$$

$$\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2} \quad (2.23)$$

Different force fields for CO_2 are available in the literature. These range from simple single-site force fields like the Higashi *et al.* model [180] and Statistical Associating Fluid Theory (SAFT)- γ [105] to more complex three-site force fields such as the Transferable Potentials for Phase Equilibria (TraPPE)-rigid [82], TraPPE-flex [181], Elementary Physical Model 2 (EPM2) [114], Zhang and Duan model [182], and Cygan, Romanov and Myshakin model [183]. Aimoli, Maginn and Abreu [112] investigated the performance of seven CO_2 force fields (TraPPE-rigid, TraPPE-flex, EPM2, Zhang model, Cygan model, Higashi model, and SAFT- γ) on density and second derivative thermodynamic properties of CO_2 up to 900 K and 1000 bar. Aimoli, Maginn and Abreu [112] found that TraPPE-rigid, EPM2, the Zhang model, and SAFT- γ produced nearly identical densities of CO_2 but the SAFT- γ force field predicted second derivative properties less accurately than TraPPE-rigid, EPM2, and Zhang model compared to reference data from National Institute of Standards and Technology (NIST). Aimoli, Maginn and Abreu [112] also investigated the TraPPE and SAFT- γ single-site force fields for CH_4 and found that the TraPPE force field performed the best compared to NIST reference data. Similar to CO_2 , several force fields for N_2 can be found in the literature, such as TraPPE [82], Köster, Thol and

Vrabec model [115], Murthy, O'Shea and McDonald model [184], Galassi and Tildesley model [185]. Force fields from Galassi and Tildesley [185] and Murthy, O'Shea and McDonald [184] were not optimized for VLE calculations [82]. The force fields TraPPE [82] and Köster, Thol and Vrabec [115] are three-site models with a quadrupole moment but differ in the parameterization. Rahbari *et al.* [167] compared the performance of five different force fields for H₂ (Cracknell [186], Buch [187], Hirschfelder, Curtiss and Bird [188], Marx and Nielaba [189], and Köster, Thol and Vrabec [115]). These authors found that the force field developed by Köster, Thol and Vrabec [115] best predicted the second derivative thermodynamic properties with the least deviation compared to data obtained from REFPROP [190]. The TraPPE [81, 82] force field was used in this chapter for CO₂, N₂, and CH₄ molecules since the primary objective in developing TraPPE force field is to predict the thermophysical properties for a wide range of state conditions and compositions. Hydrogen is simulated using the Köster, Thol and Vrabec [115] model. The force field from García-Pérez *et al.* [116] commensurate with tail corrections was used for monatomic nonpolar argon molecule. Molecular models of H₂, Ar, and CH₄ are single-site models consisting of a single LJ interaction site with no point charges, whereas CO₂ and N₂ are three-site models (including the dummy charge site for N₂) with point charges. The force field parameters of CO₂, Ar, N₂, H₂, and CH₄ are listed in Section S1 of the Supporting Information for this chapter [147].

All molecules were treated as rigid objects. LJ interactions in a simulation box have a cutoff radius of 12 Å with analytic tail corrections [79]. Periodic boundary conditions were imposed in all directions. The Ewald summation is used to compute the electrostatic energy due to point charges. To minimize computation expense, the cutoff radius of real space electrostatic interactions is chosen to limit the k -vectors (to a maximum of $k = 8$) in Fourier space, with a relative accuracy of 10^{-6} . For instance, we chose a cutoff radius in real space as 12 Å with a damping parameter of $\alpha = 0.2650 \text{ Å}^{-1}$ for a box of size 30 Å, whereas for a box of size 40 Å we chose cutoff radius in real space as 16 Å with a damping parameter of $\alpha = 0.1960 \text{ Å}^{-1}$. For each condition (concentration, temperature, and pressure), 10 independent simulations are performed, and each simulation starts with a different initial configuration. These 10 simulations are divided into 5 blocks from which average values and uncertainties of thermodynamic and transport properties are calculated. The mean and standard deviation of 5 blocks is the average value and uncertainty of a thermodynamic or transport property. The aforementioned force fields and simulation details used for MC and MD simulations are exactly the same.

2.3.1. Thermodynamic Properties

MC simulation of Gibbs Ensemble (GE) in *NVT* and *NPT* version is the most convenient way to perform phase equilibria calculations [120, 191, 192]. In GE, two simulation boxes are considered, one representing the liquid phase, and the other represents the gas phase. The simulation boxes are allowed to exchange energy, volume, and molecules. In a dense liquid simulation box, the insertion of a molecule in a single step is impeded due to the low probability of finding a cavity to accommodate a molecule, and the deletion of a molecule in a single step leaves the simulation box with a high energy penalty to form a new configuration [123]. The CFCMC method [121–125] overcomes this drawback by gradual insertion and removal of so-called fractional molecules by which the surrounding whole molecules can adapt simultaneously by performing trial moves related to fractional molecules besides thermalization trial moves such as translations, rotations, and, volume changes. In the Continuous Fractional Component Gibbs Ensemble (CFCGE), two simulation boxes with indistinguishable whole molecules and fractional molecules (fractional molecules can be in either of the simulation boxes but one per component type) are used for simulating the phase coexistence. The interaction of the distinguishable fractional molecule of a component type i with a whole molecule is scaled with a coupling parameter $\lambda_i \in [0, 1]$ [193]. The trial moves related to fractional molecules are randomly changing the value of λ while keeping the orientation and position of all the molecules constant, insertion of a fractional molecule in another simulation box at a randomly selected orientation and position while keeping the orientation and position of the whole molecules constant, and changing the identity of a fractional molecule in a simulation box to a whole molecule while simultaneously transforming the randomly selected whole molecule to a fraction molecule in another simulation box [194]. Further details specific to phase equilibria calculation in the CFCGE can be found elsewhere [123, 193]. The VLE of pure components (unary) systems CO_2 , Ar, N_2 , H_2 , and CH_4 , were computed in the *NVT* version of the CFCGE. The phase equilibria (Pxy diagram) of binary mixtures which include CO_2/Ar , CO_2/N_2 , CO_2/H_2 , and CO_2/CH_4 are computed in the *NPT* version of the CFCGE. The simulation box sizes and initial distribution of whole molecules between the simulation boxes were specified based on the experimental data of the state point for both the unary and binary systems. For example, phase equilibria computation of CO_2 and Ar binary system at 105 bar, we chose 360 CO_2 molecules and 140 Ar molecules for the intended liquid box of size 30 Å and 180 CO_2 molecules, and 320 Ar molecules for the intended gas box of size 40 Å. For a pressure of 35 bar, we chose 470 CO_2 molecules and 30 Ar molecules for the intended liquid box of size 30 Å and 325 CO_2 molecule and 175 Ar molecules for the intended gas box of size 40 Å. For simulating the phase coexistence of unary and binary systems, an equilibration run of 5×10^4 and 1×10^5 MC cycles was performed, respectively. The number of trial moves in an MC cycle in Brick-CFCMC

equals the total number of molecules in the simulation box with a minimum of 20. Following the equilibration run, a production run of 1×10^5 cycles was performed for unary systems and 2×10^5 cycles for binary systems to compute the coexistence densities and mole fractions of the components, respectively.

To calculate the thermodynamic properties, namely density (ρ), isothermal compressibility (β_T), thermal expansion coefficient (α_P), molar heat capacity at constant volume (c_V), molar heat capacity at constant pressure (c_P), Joule–Thomson coefficient (μ_{JT}), and speed of sound (c) MC simulations were performed without fractional molecules. The values of c_P , β_T , μ_{JT} , and α_P were computed using the fluctuation equations in the *NPT* ensemble defined in Eqs. 2.5, 2.17, 2.4, and 2.16, respectively. To compute the c_V and consequently the speed of sound (c), it is important to extract the ensemble-averaged volume ($\langle V \rangle$) from the *NPT* ensemble that reflects the same state for performing simulation in the *NVT* ensemble with the same number of molecules. The simulations were performed with 300 molecules, irrespective of the state and system. An equilibration of 5×10^4 MC cycles is performed to equilibrate the system successively, and 10^6 production runs are performed for each simulation.

2.3.2. Transport Properties

In MD simulations, initial configurations of molecules in the cubic simulation boxes are constructed using PACKMOL [195] and fftool [196]. Periodic boundary conditions are applied to simulation boxes in all directions. A cutoff radius of 12 Å is used for LJ interactions with analytic tail corrections. All molecules are treated as rigid bodies, and Newton's equations of motion are integrated using the velocity-Verlet algorithm with a time step of 0.5 fs. To thermostat and barostat the system, the Nosé-Hoover type is used with coupling constants of 0.1 ps and 1 ps, respectively. The Particle-Particle Particle-Mesh (PPPM) method is used to handle long-range electrostatic interactions with a cutoff radius of 12 Å and a relative accuracy of 10^{-6} . Shear viscosities (η) are calculated by performing an Equilibrium Molecular Dynamics (EMD) simulation by using the On-the-fly Computation of Transport Properties (OCTP) plugin in LAMMPS. The OCTP combines the Einstein relations with an order- n algorithm to calculate viscosity. Additional details about the OCTP's computation of transport properties can be found elsewhere [134]. System sizes chosen to evaluate η of pure and multicomponent mixtures were 100 molecules for very diluted gas phase, 300 molecules for the gas phase (40 to 80 bar), and 400 molecules for the liquid and supercritical phases (≥ 100 bar). The various steps involved in computing η for one independent simulation are as follows. First, a simulation is carried out in the *NPT* ensemble (0.5 ns equilibration run and 1 ns production run) to compute the ensemble average volume ($\langle V \rangle$). Next, the simulation box is scaled according to the computed value of $\langle V \rangle$, and this system is then used to perform simulation

in the *NVT* ensemble (0.5 ns equilibration run and 1 ns production run) to compute the average total energy of the system ($\langle E \rangle$). Finally, the ensemble average total energy is used to scale to the kinetic energy of the system to perform simulation in *NVE* ensemble. Viscosities are calculated in the *NVE* ensemble, ensuring that the thermostat and barostat have no effect on the results. In *NVE* ensemble, a production run of 5 ns is simulated to compute η .

2.4. Results and Discussion

This section shows and discusses the thermodynamic and transport properties of single and multicomponent CO₂ systems. All properties of interest computed in this chapter from molecular simulations are compared with datasets generated from the NIST REFPROP database version-10.0 [190] except when stated otherwise. EoS and correlation models used by REFPROP for computing pure components (CO₂, N₂, Ar, H₂, and CH₄) thermodynamic and transport properties are listed in Table S11 of the Supporting Information for this chapter [147]. For multicomponent systems, the Groupe Européen de Recherches Gazières (GERG-2008) EoS [58] is chosen in REFPROP. The GERG-2008 EoS is less accurate for unary systems [190]. Therefore, it is only used to compute the thermodynamic properties of multicomponent systems. The GERG-2008 EoS was originally developed for natural gas mixtures containing 21 components, including CO₂, N₂, Ar, H₂, and CH₄. Multicomponent mixtures containing high CO₂ concentrations and low levels of impurities were not the main focus for the development of the GERG-2008 EoS [58, 59]. The quality and quantity of experimental data used in developing the GERG-2008 EoS limits its accuracy [58]. Thermophysical properties computed from molecular simulations will help in optimizing the EoS [160]. Viscosities computed from MD simulations of multicomponent systems are compared with those obtained from the Extended Corresponding States (ECS) model available in REFPROP [197]. Further details specific to the ECS model can be found elsewhere [197]. Since the experimental data for multicomponent systems are scarce, and simulating an exact composition as experiments is impossible with a small system of only 300 molecules, we opted to compare results from molecular simulations with those obtained from EoS. The numerical data used to generate all plots is provided in Sections S3, S4, and S14-S17 of the Supporting Information for this chapter [147]. The deviations of properties computed from simulations with respect to REFPROP datasets are computed using:

$$\text{Relative Deviation \%} = \left| \frac{\chi^{\text{REFPROP}} - \chi^{\text{Simulation}}}{\chi^{\text{REFPROP}}} \right| \times 100\% \quad (2.24)$$

where $\chi^{\text{Simulation}}$ and χ^{REFPROP} are the property of interest computed from simulations and REFPROP, respectively. For the sake of clarity, plots for

temperatures 253 K and 313 K are shown for binary systems, and plots for a temperature 253 K are shown for multicomponent systems despite properties of interest were estimated for four different temperatures, namely 253 K, 273 K, 293 K, and 313 K.

2.4.1. Thermodynamic Properties

Phase Equilibria

The VLE of pure components (CO₂, Ar, N₂, H₂, and CH₄) computed from the CFCMC simulations in the *NVT* version are compared to REFPROP database. The computed VLE curves compared to REFPROP database are shown in Section S2 of the Supporting Information for this chapter [147]. Critical temperatures (T_c) and densities (ρ_c) of all pure components computed from simulations using the law of rectilinear diameters and EoS models are listed in Table 2.3. Our results show that the computed liquid and vapor densities are in excellent agreement with respect to REFPROP dataset, except for the liquid densities of hydrogen. Deviations of the computed liquid densities and, in turn, the T_c and ρ_c of hydrogen with respect to REFPROP are due to the domination of quantum effects at low temperatures [115]. This chapter focuses on temperatures significantly higher (> 250 K) than H₂ VLE temperatures, hence, only the gas densities were considered for the validation of H₂ force field.

Fig. 2.1 shows the phase equilibria (*Pxy* diagram) of four binary mixtures comprising of CO₂, Ar, N₂, H₂, and CH₄ computed from CFCGE simulations (in the *NPT* version) compared with experimental data and the GERG-2008 EoS. The *Pxy* diagram, i.e., the bubble and dew points of CO₂/Ar mixtures, is computed at 253.28 K and compared with experimental data of Coquelet *et al.* [198] and for CO₂/CH₄ mixtures, simulations were performed at 250 K and compared to experimental data of Wei *et al.* [200] and Davalos *et al.* [199]. Similarly, *Pxy* diagram of CO₂/H₂ and CO₂/N₂ mixtures are computed at 250 K and validated with experimental data of Tsang and Street [201] and Brown *et al.* [202], respectively. Fig. 2.1 (a) shows CO₂ liquid and gas mole fractions of CO₂/Ar mixtures, computed from the CFCGE simulations compared to experimental data of Coquelet *et al.* [198] and the GERG-2008 EoS. Our results match well with the GERG-2008 EoS compared to the experimental data of Coquelet *et al.* [198]. Bubble and dew points of CO₂/CH₄ mixtures (shown in Fig. 2.1 (b)) computed from simulations agrees well with experimental data of Davalos *et al.* [199] when compared to experimental data of Wei *et al.* [200] and the GERG-2008 EoS at high pressures. The phase equilibria of CO₂/H₂ binary system is shown in Fig. 2.1 (c). REFPROP fails to converge for pressures larger than 170 bar when using the GERG-2008 EoS. This discrepancy seen in the GERG-2008 EoS was also reported in the work of Shin *et al.* [165]. The bubble points obtained from the GERG-2008 EoS using REFPROP provide poor estimates when compared to the experimental data

Table 2.3. Comparison of critical temperatures ($T_c^{\text{SIM.}}$) and densities ($\rho_c^{\text{SIM.}}$) of pure components computed from simulations using the law of rectilinear diameters with critical temperatures ($T_c^{\text{REFP.}}$) and densities ($\rho_c^{\text{REFP.}}$) obtained from the REFPROP [190] database.

Component	CAS number	$T_c^{\text{REFP.}}$ / [K]	$T_c^{\text{SIM.}}$ / [K]	$\rho_c^{\text{REFP.}}$ / [kg m ⁻³]	$\rho_c^{\text{SIM.}}$ / [kg m ⁻³]
CO ₂	124-38-9	304.1	306.05	467.6	466.53
N ₂	7727-37-9	126.20	126.30	314.40	308.73
Ar	7440-37-1	150.65	148.93	536	539.70
H ₂	1333-74-0	33.18	33.36	31.04	37.39
CH ₄	74-82-8	190.6	191.51	162.1	160.67

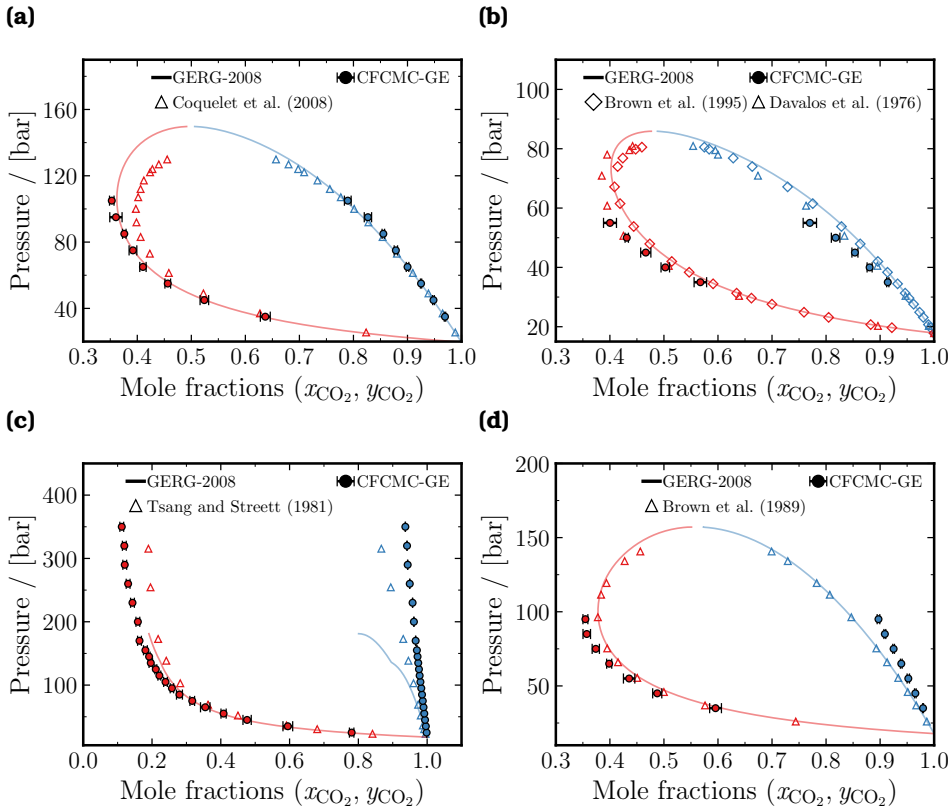


Figure 2.1. Comparison of binary VLE (Pxy diagram), i.e., the bubble points (blue symbols and lines) and dew points (red symbols and lines) of (a) CO₂/Ar, (b) CO₂/CH₄, (c) CO₂/H₂, and (d) CO₂/N₂ mixtures computed from CFCGE simulations with experimental data [198–202] and the GERG-2008 EoS [58]. The simulations are performed at 253.28 K for CO₂/Ar mixtures and at 250 K for CO₂/CH₄, CO₂/H₂, and CO₂/N₂ mixtures.

of Tsang and Street [201]. Mole fractions computed from simulations agree fairly with experimental data of Tsang and Street [201], and less than 5% relative deviation was noticed at high pressures. Dew points of CO₂/N₂ mixtures shown in Fig. 2.1 (d) have a reasonable agreement with EoS and experimental data of Brown *et al.* [202], but the computed bubble points have a maximum relative deviation of 3% at high pressures. At high pressures, the dew and bubble points computed from CFCGE simulations agree moderately with the GERG-2008 EoS and experiments for all systems shown in Fig. 2.1. This is because the mixing rules used in simulations did not include binary interaction parameters. Computing bubble and dew points close to the critical pressure is challenging, even with a system size of 1000 molecules (the system size chosen in this chapter to compute the phase equilibria). In principle, one can perform simulations with large system sizes for pressures in the neighborhood of the critical pressure at the cost of much larger computations. Nevertheless, it is impossible to compute accurate bubble and dew points very close to critical pressure since simulation boxes may switch identities, which complicates ensemble averaging. Similarly, the GERG-2008 EoS model fails to predict reliable results for pressures close to the critical pressure due to the unavailability of experimental data to develop the GERG-2008 EoS [59].

Densities (ρ)

Densities of pure CO₂ computed from MC and MD simulations for temperatures 253 K, 273 K, 293 K, and 313 K and pressures ranging from 20 bar to 200 bar are shown in Fig. 2.2 (a). MC and MD simulations have an excellent agreement with the Span and Wagner EoS for all temperatures. The computed densities from MC simulations have a maximum relative deviation of ca. 0.93% at 313 K and 200 bar, and MD simulations have a maximum relative deviation of ca. 0.89% at 293 K and 200 bar, excluding conditions close to the critical point. As expected, the density decreases with temperature to a large extent after the transition from gas to liquid or supercritical fluid. Fig. 2.2 (b) shows densities of different binary mixtures computed from MC simulations compared with densities obtained from the GERG-2008 EoS for 95 mol% of CO₂ and 5 mol% of one of impurities (CH₄, Ar, N₂, and H₂) for temperatures 253 K and 313 K. The computed densities of CO₂/N₂, CO₂/CH₄, CO₂/Ar and CO₂/H₂ binary mixtures from MC and MD are in close agreement with densities obtained from the GERG-2008 EoS and has a maximum relative deviation of ca. 4.4% (in MC simulation at 313 K and 60 bar), ca. 4.3% (in MD simulations at 253 K and 80 bar), ca. 2.1% (in MC simulations at 253 K and 40 bar), ca. 6.6% (in MD simulations at 313 K and 60 bar), respectively, excluding state points close to the critical point. A comprehensive analysis of the gas phase densities of binary mixtures with 5 mol% of one of the impurities found that impurities do not significantly change the gas phase densities. The influence of impurities on densities is consistent with the molecular weight

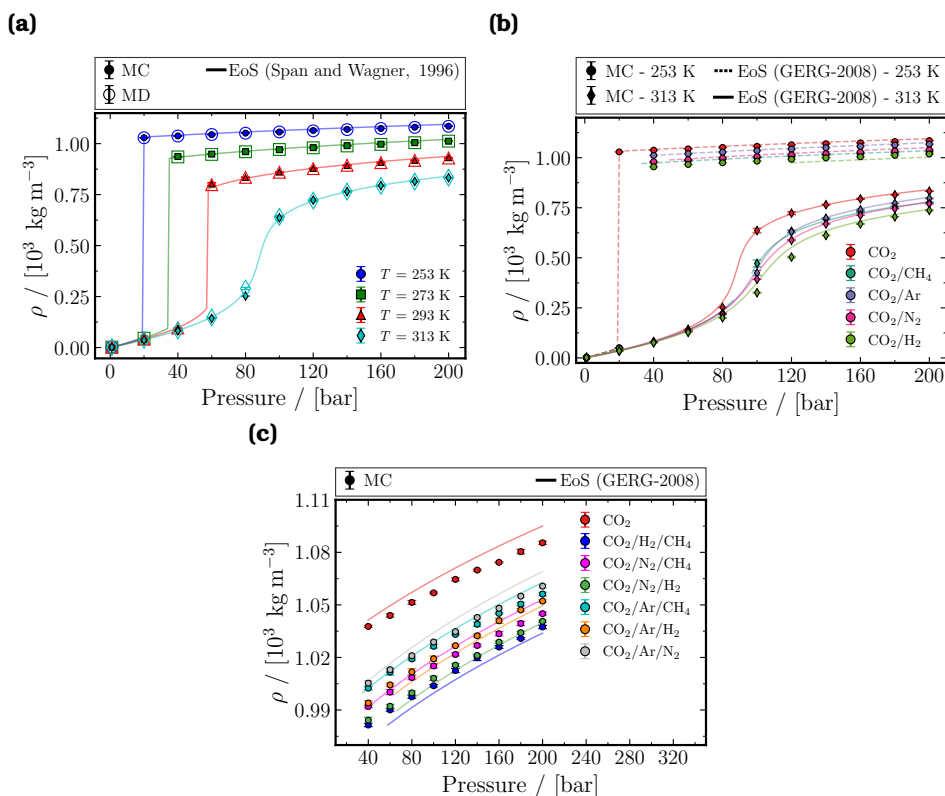


Figure 2.2. Computed densities as a function of temperature and pressure. (a) shows the calculated densities of pure CO_2 from MC simulations (closed symbols), MD simulations (open symbols), and the Span and Wagner EoS [28] (solid lines) for temperatures 253 K, 273 K, 293 K, and 313 K. (b) shows densities of binary mixtures with 95 mol% of CO_2 and 5 mol% of impurities (CH_4 , Ar, N_2 , and H_2) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with densities of pure CO_2 computed from MC simulations and the Span and Wagner EoS [28] for temperatures 253 K and 313 K. (c) shows densities of ternary mixtures with 96 mol% of CO_2 and 2 mol% impurities for each of two components (CH_4 , Ar, N_2 , and H_2) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with densities of pure CO_2 computed from MC simulations and the Span and Wagner EoS [28] at 253 K.

of the mixtures. A mixture with H_2 as an impurity decreases the density to a large extent, followed by Ar, N_2 , and CH_4 . Densities were also computed for 1 mol% and 10 mol% impurities at temperatures 253 K, 273 K, 293 K, and 313 K from MC simulations and MD simulations. The computed densities are provided in Tables S73-S167 of the Supporting Information for this

chapter [147]. Densities of all binary mixtures shown in Section S15 of the Supporting Information for this chapter [147], densities decrease with increasing mol% of impurity compared to densities of pure CO₂ irrespective of the temperature. Fig. 2.2 (c) shows the liquid densities of ternary mixtures with 96 mol% CO₂ and 2 mol% impurities for each of the two components (CH₄, Ar, N₂, and H₂) at 253 K. The liquid densities of ternary mixtures may appear to have some deviation when compared to the GERG-2008 EoS due to the reduced axis range seen in Fig. 2.2 (c). The maximum relative deviation of ternary mixtures liquid densities in Fig. 2.2 (c) was ca. 0.78% for CO₂/Ar/N₂ and CO₂/N₂/CH₄ mixtures at 200 bar. The liquid densities of ternary mixtures were observed to have good agreement with the GERG-2008 EoS. Similar to binary mixtures, ternary mixtures with the least molecular weight tend to have lesser densities when compared to other ternary mixtures. For example, a ternary mixture with 2 mol% H₂ and 2 mol% CH₄ as impurities, which have the lowest molecular weight among other ternary mixtures, have the lowest densities when compared to densities of other ternary mixtures.

Thermal Expansion Coefficients (α_P)

Thermal expansion coefficient (α_P) computed from MC simulations using Eq. 2.17 in the *NPT* ensemble for pure CO₂, binary CO₂ mixtures, and ternary CO₂ rich mixtures with different impurities are shown as a function of temperature and pressure in Fig. 2.3. α_P of pure CO₂ computed at temperatures 253 K, 273 K, 293 K, and 313 K are shown in Fig. 2.3 (a). α_P of pure CO₂ increases with temperature, whereas with increasing pressure, the value of α_P increases till it reaches a maximum value close to its saturation pressure and then decreases with increasing pressure. MC simulations closely predicted the peak of α_P . Thermal expansion coefficients of pure CO₂ computed at conditions close to its saturation pressure were found to have large uncertainties and deviations when compared to the Span and Wagner EoS. Excluding conditions close to the critical point, α_P of pure CO₂ computed from MC simulations were in good agreement with the Span and Wagner EoS for all temperatures with a maximum relative deviation of ca. 6.8% at 293 K and 100 bar. In binary mixtures containing 95 mol% of CO₂ and 5 mol% of one of the impurities (CH₄, Ar, N₂, and H₂), a similar pattern of α_P with a peak value near its saturation pressure was found, as shown in Fig. 2.3 (c) at temperatures 253 K and 313 K. The computed α_P of binary mixtures CO₂/N₂, CO₂/CH₄, CO₂/Ar and CO₂/H₂ agrees fairly with α_P obtained from the GERG-2008 EoS and has a maximum relative deviation of ca. 13.9% at 313 K and 120 bar, ca. 10.2% at 313 K and 60 bar, ca. 8.1% at 313 K and 60 bar, ca. 20.4% at 313 K and 120 bar, respectively, excluding state points close to the critical point. State points of binary mixtures close to the critical point were identified based on large uncertainties observed in simulations and large relative deviation with the GERG-2008 EoS. The uncertainties and relative

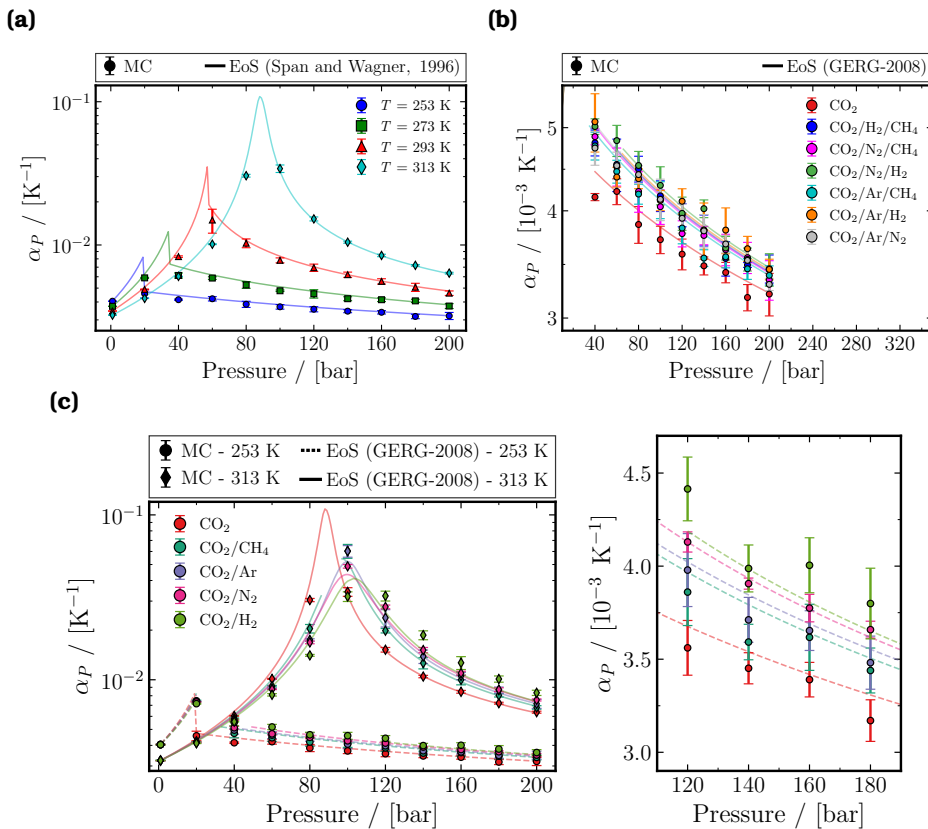


Figure 2.3. Computed thermal expansion coefficients as a function of temperature and pressure. (a) shows the calculated thermal expansion coefficients of pure CO_2 from MC simulations (closed symbols) and the Span and Wagner EoS [28] (solid lines) for temperatures 253 K, 273 K, 293 K, and 313 K. (b) shows thermal expansion coefficients of ternary mixtures with 96 mol% of CO_2 and 2 mol% impurities for each of two components (CH_4 , Ar, N_2 , and H_2) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with thermal expansion coefficients of pure CO_2 computed from MC simulations and the Span and Wagner EoS [28] at 253 K. (c) shows thermal expansion coefficients of binary mixtures with 95 mol% of CO_2 and 5 mol% of impurities (CH_4 , Ar, N_2 , and H_2) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with thermal expansion coefficients of pure CO_2 computed from MC simulations and the Span and Wagner EoS [28] for temperatures 253 K and 313 K.

deviation of CO₂/H₂ at 313 K decrease with increasing pressure. This suggests that the maximum relative deviation for CO₂/H₂ mixture observed in simulations at 313 K and 120 bar may be near its saturation/Widom line. The binary mixture containing 5 mol% of H₂ increases the value of α_P the most in liquid and supercritical phases when compared to α_P of pure CO₂, followed by binary mixtures containing 5 mol% of N₂, Ar, and CH₄. In the gas phase, the presence of H₂ as impurity decreases the value of α_P followed by CH₄, N₂, and Ar compared to α_P of pure CO₂. Thermal expansion coefficients were also computed for 1 mol% and 10 mol% impurities at temperatures 253 K, 273 K, 293 K, and 313 K from MC simulations. The computed thermal expansion coefficients provided in Tables S73-S167 of the Supporting Information for this chapter [147] show that the value of α_P are affected based on impurities and concentration level of impurities. Thermal expansion coefficients of ternary mixtures in the liquid phase with 96 mol% CO₂ and 2 mol% for each of the two impurities (CH₄, Ar, N₂, and H₂) are shown in Fig. 2.3 (b). The effect of a particular mixture on α_P was not discernible since the thermal expansion coefficients of all the ternary mixtures computed at 253 K were within the limits of computed uncertainty of other ternary mixtures. Similarly, no significant differences were found between the thermal expansion coefficients of ternary mixtures obtained from the GERG-2008 EoS.

Isothermal Compressibilities (β_T)

The isothermal compressibility (β_T) computed from MC simulations using Eq. 2.5 in the *NPT* ensemble for pure CO₂, binary CO₂ mixtures, and ternary CO₂ rich mixtures with different impurities are shown as a function of temperature and pressure in Fig. 2.4. β_T of pure CO₂ at temperature 253 K, 273 K, 293 K, and 313 K are shown in Fig. 2.4 (a). The maximum value of β_T was observed in the gas phase. This implies the volume change rate in response to the change in pressure is maximum when the fluid acts like an ideal gas. β_T rapidly decreases with increasing pressure for pressures lower than the saturation pressure of a particular temperature. For pressures away from the saturation/Widom line, the change in the value of β_T was insignificant for all temperatures, as seen in Fig. 2.4 (a). Fig. 2.4 (a) also shows that β_T is dependent on temperature and β_T increases with increasing temperature. β_T of pure CO₂ computed from MC simulations agrees qualitatively with the Span and Wagner EoS with a maximum relative deviation of ca. 19.9% at 293 K and 100 bar. β_T of binary mixtures with 95 mol% of CO₂ and 5 mol% of one of the impurities (CH₄, Ar, N₂, and H₂) at temperature 253 K, and 313 K are shown in Fig. 2.4 (c). Isothermal compressibilities of binary mixtures computed from MC simulations agree fairly with the GERG-2008 EoS. The maximum relative deviation of CO₂/N₂, CO₂/CH₄, CO₂/Ar and CO₂/H₂ binary mixtures were ca. 14.1% at 253 K and 80 bar, ca. 17.5% at 313 K and 120 bar, ca. 12.9% at 253 K and 120 bar, and ca. 22.5% at 313 K and 140 bar,

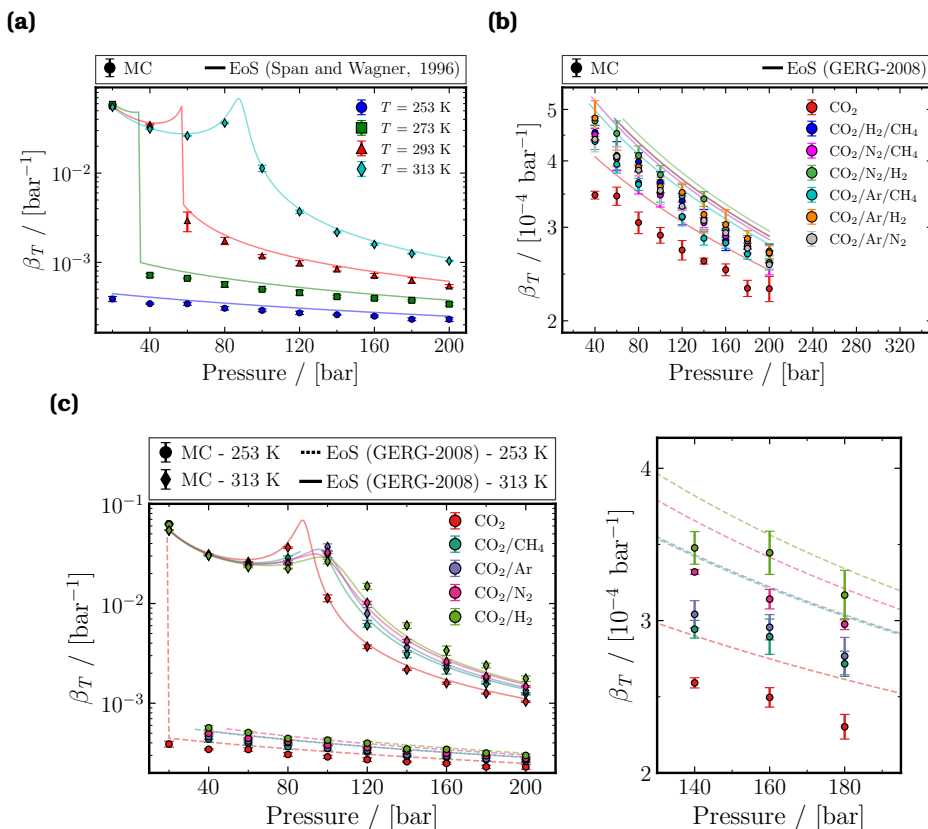


Figure 2.4. Computed isothermal compressibilities as a function of temperature and pressure. (a) shows the calculated isothermal compressibilities of pure CO₂ from MC simulations (closed symbols) and the Span and Wagner EoS [28] (solid lines) for temperatures 253 K, 273 K, 293 K, and 313 K. (b) shows isothermal compressibilities of ternary mixtures with 96 mol% of CO₂ and 2 mol% impurities for each of two components (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with isothermal compressibilities of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] at 253 K. (c) shows isothermal compressibilities of binary mixtures with 95 mol% of CO₂ and 5 mol% of impurities (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with isothermal compressibilities of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] for temperatures 253 K and 313 K.

respectively, excluding state points close to the critical point. Fig. 2.4 (c) shows that the presence of impurity increases β_T in the liquid and supercritical phases. In contrast, the presence of impurity decreases β_T in the gas phase. Comparing β_T of all binary mixtures shown in Fig. 2.4 (c), a binary mixture with H₂ as an impurity increases the value of β_T the most in the liquid and supercritical phases, followed by N₂, Ar, and CH₄. In the gas phase, a binary mixture with H₂ as an impurity decreases the value of β_T followed by N₂, Ar, and CH₄. This pattern of β_T with respect to phases remained the same for both higher (10 mol%) and lower (1 mol%) concentrations of impurities, which is provided in Tables S73-S167 of the Supporting Information for this chapter [147]. The liquid phase β_T of ternary mixtures with 96 mol% CO₂ and 2 mol% for each of the two impurities (CH₄, Ar, N₂, and H₂) at 253 K are shown in Fig. 2.4 (b). From Fig. 2.4 (b), it is obvious that the presence of impurities tends to increase the value of β_T . Similar to thermal expansion coefficients, the effect of a particular type of impurity on β_T was unclear since the change in the value of β_T due to the presence of impurities was limited. However, β_T computed from MC simulations for CO₂ rich ternary mixtures with different impurities agrees fairly with the GERG-2008 EoS. For instance, CO₂ rich ternary mixture with Ar and CH₄ as impurities results in the smallest increase in β_T , while CO₂ rich ternary mixture with N₂ and H₂ as impurities lead to the largest increase in β_T . The difference in the value of β_T due to the presence of different impurities combinations estimated from MC simulations was consistent with the GERG-2008 EoS.

Isobaric and Isochoric Heat Capacities

The constant pressure and volume heat capacities are computed using Eqs. 2.4 and 2.3, respectively. Residual heat capacities C_p^{residual} and C_V^{residual} are computed from MC simulations in the *NPT* and *NVT* ensemble respectively to obtain c_p and c_V . Figs. 2.5 and 2.6 show c_p of pure CO₂, binary CO₂ mixtures, and ternary CO₂ mixtures as a function of temperature and pressure. c_p of pure CO₂ computed from MC simulations compared to the Span and Wagner EoS at temperature 253 K, 273 K, 293 K, and 313 K are shown in Fig. 2.5 (a). Heat capacities at a constant pressure of pure CO₂ are sensitive to pressures closer to the saturation/Widom line, where a sudden surge in c_p is noticed. Fig. 2.5 (a) also shows that c_p increases with temperature. The maximum value of c_p observed in Fig. 2.5 (a) at 313 K is calculated using the Span and Wagner EoS. However, this peak value of c_p cannot be predicted by MC simulations unless one considers an extremely large system. This is because the correlation length, which measures the spatial extent of spontaneous density fluctuations, diverges near the critical point [79]. c_p of pure CO₂ computed from MC simulations resulted in minor systematic overprediction with a maximum relative deviation of ca. 6.9% at 313 K and 120 bar. At 20 bar, relative deviations of c_p with Span and Wagner EoS are less than 1% for all

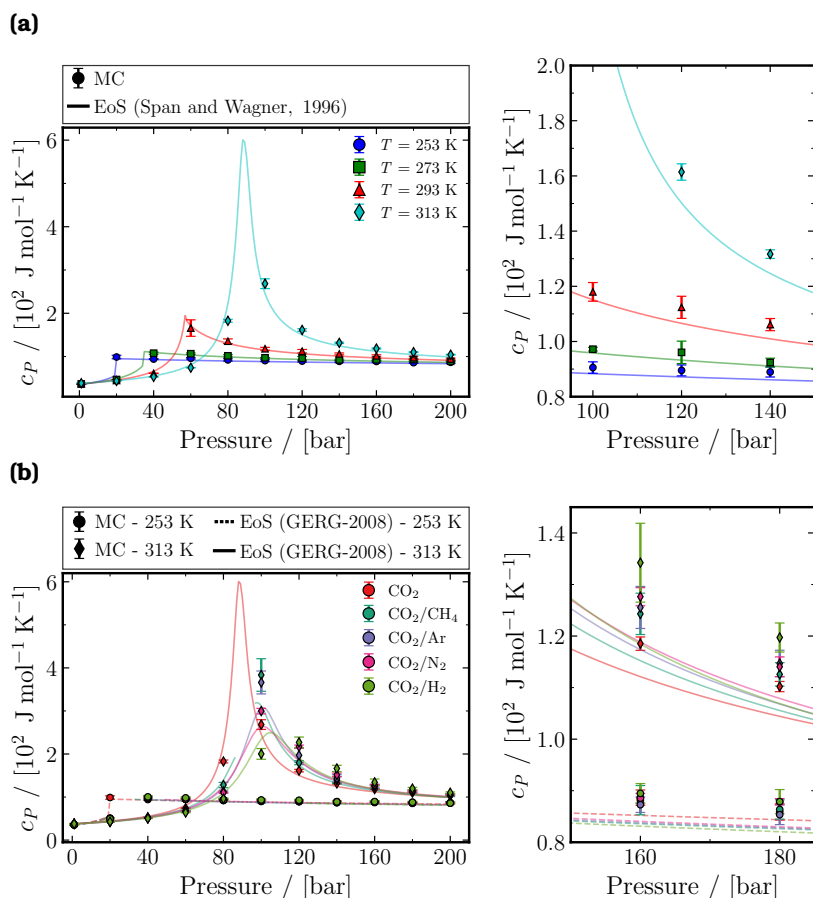


Figure 2.5. Computed isobaric heat capacities as a function of temperature and pressure. (a) shows the calculated isobaric heat capacities of pure CO₂ from MC simulations (closed symbols) and the Span and Wagner EoS [28] (solid lines) for temperatures 253 K, 273 K, 293 K, and 313 K. (b) shows isobaric heat capacities of binary mixtures with 95 mol% of CO₂ and 5 mol% of impurities (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with isobaric heat capacities of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] for temperatures 253 K and 313 K.

temperatures shown in Fig. 2.5 (a), except at 253 K. At 253 K and 20 bar, c_P has a relative deviation of ca. 3.8%, which is attributed to the close distance to the vapor-liquid phase transition. c_P of binary mixtures with 95 mol% of CO₂ and 5 mol% of one of the impurities (CH₄, Ar, N₂, and H₂) at temperature 253 K, and 313 K are shown in Fig. 2.5 (b). Similar to c_P of pure CO₂, systematic overprediction of c_P is observed in binary mixtures

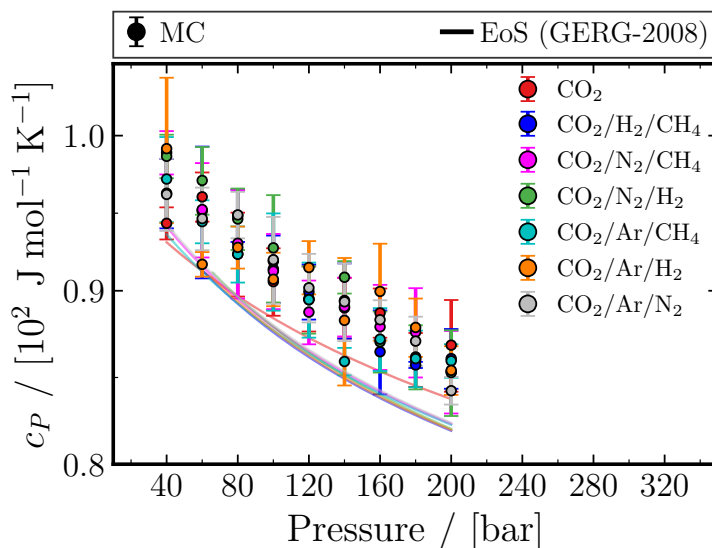


Figure 2.6. Isobaric heat capacities of ternary mixtures with 96 mol% of CO₂ and 2 mol% impurities for each of two components (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with isobaric heat capacities of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] at 253 K.

when compared to the GERG-2008 EoS. The maximum relative deviation of CO₂/N₂, CO₂/CH₄, CO₂/Ar and CO₂/H₂ binary mixtures were ca. 8.4% at 313 K and 140 bar, ca. 7.3% at 313 K and 160 bar, ca. 6.9% at 313 K and 180 bar, and ca. 15.7% at 313 K and 140 bar, respectively, excluding state points close to the saturation/Widom line. Although c_P of different binary mixtures may appear to be identical in Fig. 2.5 (b), impurities altered the value of heat capacities significantly compared to c_P of pure CO₂. The presence of impurities tends to increase the value of c_P for pressure larger than the pressure at which the heat capacity peaks. Conversely, when the pressure is lower than the pressure at which the heat capacity peaks, the presence of impurities tends to decrease the value of c_P . For instance, c_P of CO₂/H₂ binary mixture at 313 K and 160 bar were ca. 13.2% larger than c_P of pure CO₂. c_P of CO₂/H₂ binary mixture at 313 K and 60 bar were ca. 13.7% smaller than c_P of pure CO₂. At conditions away from the saturation/Widom line, a particular type of impurity did not influence c_P to a great extent. However, at conditions close to the saturation/Widom line, the impact of impurities on c_P becomes significant based on the type of impurity where binary mixture with H₂ as an impurity increase c_P to a great extent followed by N₂, Ar, and CH₄. Similar to binary mixtures with 5 mol% impurities, binary mixtures with 1 mol% and 10 mol% impurities

decreased and increased the value of c_P compared to c_P of pure CO_2 before and after the maxima heat capacity, respectively, see Tables S73-S167 of the Supporting Information for this chapter [147]. c_P of ternary mixtures with 96 mol% CO_2 and 2 mol% for each of the two impurities (CH_4 , Ar, N_2 , and H_2) at 253 K are shown in Fig. 2.6. From Fig. 2.6, it is clear that the presence of impurities tends to decrease c_P at 253 K. Fig. 2.6 also confirms that the effect of a particular type of impurity combination on c_P is insignificant.

Heat capacities at constant volume of pure CO_2 at temperature 253 K, 273 K, 293 K, and 313 K are compared with the Span and Wagner EoS in Fig. 2.7 (a). The computed c_V is in excellent agreement with c_V obtained from the Span and Wagner EoS with a maximum relative deviation of ca. 5% at 313 K and 120 bar, considering conditions away from the saturation/Widom line. c_V values of pure CO_2 vary insignificantly for pressures larger than its saturation pressure, irrespective of the temperature. Similarly, c_V of binary mixtures with 95 mol% of CO_2 and 5 mol% of one of the impurities (CH_4 , Ar, N_2 , and H_2) at temperature 253 K, and 313 K seen in Fig. 2.7 (b) have no significant difference in the value of c_V between 253 K, and 313 K for pressure larger than its saturation pressure. The presence of an impurity tends to decrease c_V hardly, irrespective of the type of impurity. c_V of ternary mixtures with 96 mol% CO_2 and 2 mol% for each of the two impurities (CH_4 , Ar, N_2 , and H_2) at 253 K are shown in Fig. 2.7 (c). In the reduced axis range seen in Fig. 2.7 (c), it is clear that the presence of impurities (in $\text{CO}_2/\text{Ar}/\text{H}_2$) decreases c_V maximum by ca. 3%. It is important to mention that c_V can be calculated from c_P using Eq. 2.19. However, the indirect computation of c_V using Eq. 2.19 is prone to high statistical uncertainties [203, 204], which will subsequently affect the calculations of speed of sound. Hence, constant volume heat capacities are computed by sampling Eq. 2.3 in the NVT ensemble with the average volume of the state obtained from the NPT simulation.

Joule–Thomson Coefficients (μ_{JT})

Joule–Thomson coefficients (μ_{JT}) of pure CO_2 , binary CO_2 mixtures, and ternary CO_2 rich mixtures computed using Eq. 2.16 as a function of temperature and pressure are shown in Figs. 2.8 and 2.9. μ_{JT} decreases as the temperature increases in the gas phase. Conversely, μ_{JT} increases with temperature in the liquid and supercritical phases. The computed μ_{JT} of pure CO_2 at temperature 253 K, 273 K, 293 K, and 313 K seen in Fig. 2.8 (a) agrees decently with μ_{JT} obtained from the Span and Wagner EoS. The deviations of μ_{JT} in the gas phase are larger when compared to the liquid and supercritical phases, and uncertainties of μ_{JT} are quite significant for extremely low pressures. This is because of Eq. 2.16 used for computing μ_{JT} . Joule–Thomson coefficients are computed indirectly using Eq. 2.16, and its uncertainties are computed using Eq. (S84) presented in Section S11 of the Supporting Information for this chapter [147]. The uncertainties

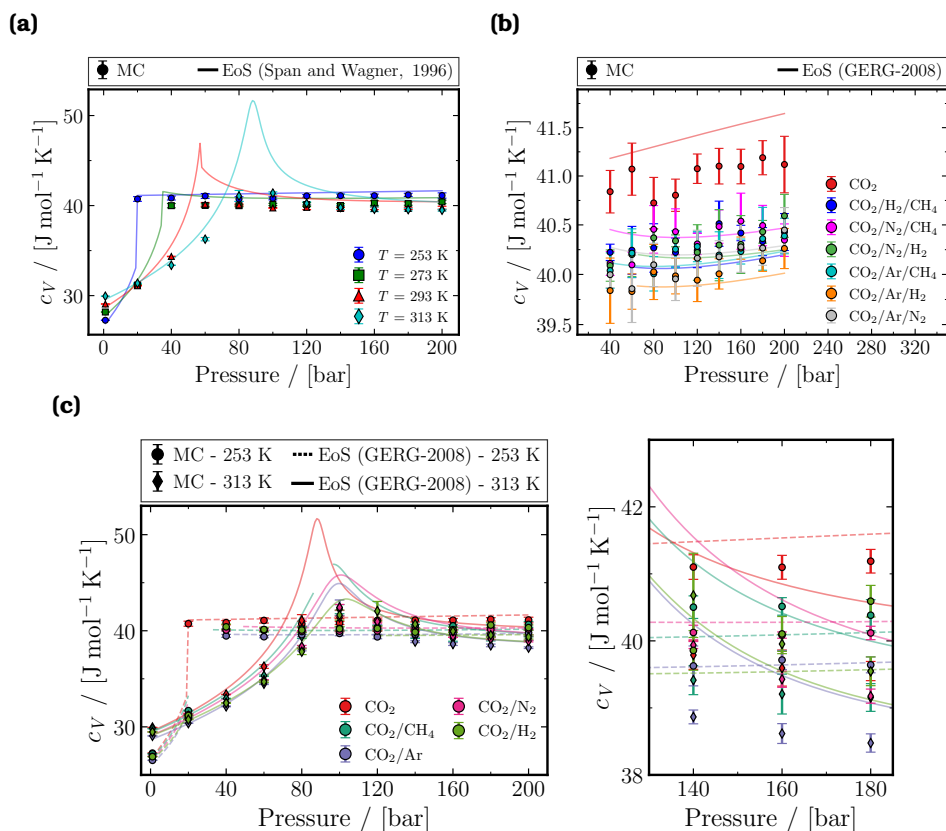


Figure 2.7. Computed isochoric heat capacities as a function of temperature and pressure. (a) shows the calculated isochoric heat capacities of pure CO₂ from MC simulations (closed symbols) and the Span and Wagner EoS [28] (solid lines) for temperatures 253 K, 273 K, 293 K, and 313 K. (b) shows isochoric heat capacities of ternary mixtures with 96 mol% of CO₂ and 2 mol% impurities for each of two components (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with isochoric heat capacities of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] at 253 K. (c) shows isochoric heat capacities of binary mixtures with 95 mol% of CO₂ and 5 mol% of impurities (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with isochoric heat capacities of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] for temperatures 253 K and 313 K.

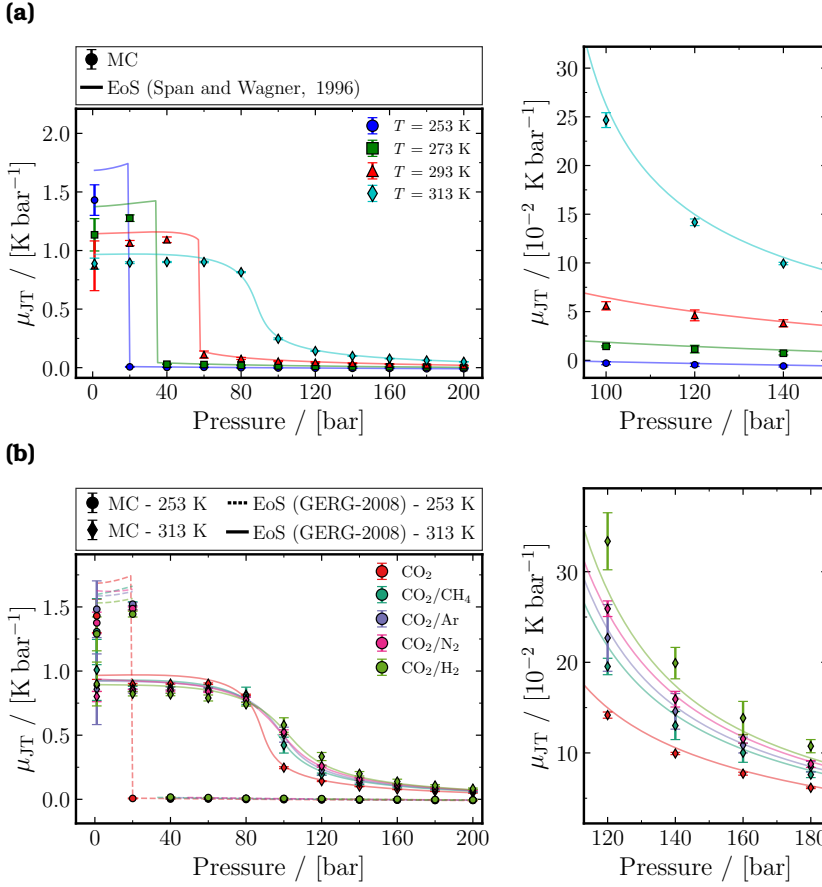


Figure 2.8. Computed Joule–Thomson coefficients as a function of temperature and pressure. (a) shows the calculated Joule–Thomson coefficients of pure CO_2 from MC simulations (closed symbols) and the Span and Wagner EoS [28] (solid lines) for temperatures 253 K, 273 K, 293 K, and 313 K. (b) shows Joule–Thomson coefficients of binary mixtures with 95 mol% of CO_2 and 5 mol% of impurities (CH_4 , Ar , N_2 , and H_2) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with Joule–Thomson coefficients of pure CO_2 computed from MC simulations and the Span and Wagner EoS [28] for temperatures 253 K and 313 K.

computed using Eq. (S84) depend on the precision and accuracy of α_P and c_P . In particular, the computation of μ_{JT} is highly sensitive to α_P . Even an insignificant error in α_P computation might result in high uncertainties and deviations in μ_{JT} . A similar drawback of using Eq. 2.16 is also reported in Ref. [205]. μ_{JT} of binary mixtures with 95 mol% of CO_2 and 5 mol% of

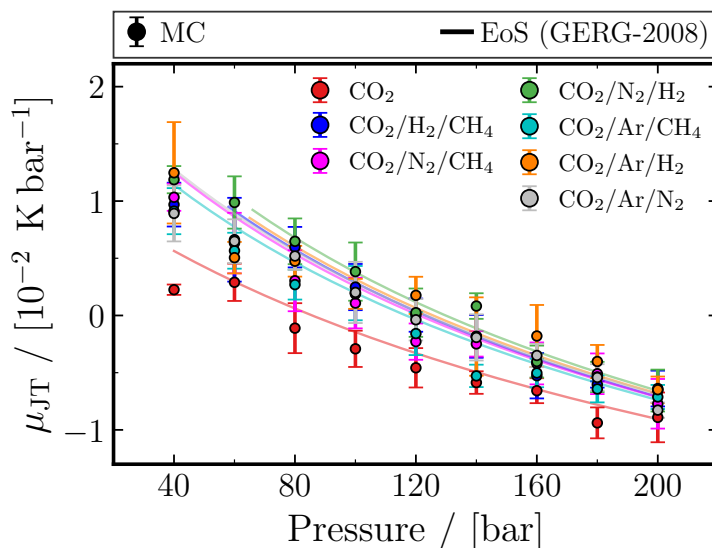


Figure 2.9. Joule–Thomson coefficients of ternary mixtures with 96 mol% of CO₂ and 2 mol% impurities for each of two components (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with Joule–Thomson coefficients of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] at 253 K.

one of the impurities (CH₄, Ar, N₂, and H₂) at temperature 253 K, and 313 K are shown in Fig. 2.8 (b). The value of μ_{JT} decreases due to the presence of impurities in the gas phase and increases in liquid and supercritical phases. H₂ has the most significant effect on the values of μ_{JT} , followed by N₂, Ar, and CH₄. μ_{JT} computed from MC simulations for binary mixtures rich in CO₂ with 1 mol%, 5 mol%, and 10 mol% of impurities (CH₄, Ar, N₂, and H₂) at 253 K, 273 K, 293 K, and 313 K listed in Tables S73–S167 of the Supporting Information for this chapter [147] shows the same qualitative trend with respect to impurity type. μ_{JT} listed in Tables S73–S167 of the Supporting Information for this chapter [147] also indicates that the impact of impurities increases with increasing concentration of impurities. μ_{JT} of ternary mixtures with 96 mol% CO₂ and 2 mol% for each of the two impurities (CH₄, Ar, N₂, and H₂) at 253 K are shown in Fig. 2.9. μ_{JT} of ternary mixtures computed from MC simulations at 253 K agrees well with μ_{JT} obtained from the GERG-2008 EoS, considering the uncertainty range. Fig. 2.9 also shows that the presence of impurities tends to increase μ_{JT} of ternary mixtures significantly. The impact of impurities is crucial and should be considered in the initial computations, especially at conditions where the inversion of μ_{JT} takes place. For instance, μ_{JT} inversion occurs

close to 80 bar for pure CO₂ at 253 K but for impure CO₂ ternary mixtures at 253 K, μ_{JT} inversion occurs at pressure larger than 120 bar.

Speed of Sound (c)

Speed of sound (c) of pure CO₂, binary CO₂ mixtures, and ternary CO₂ rich mixtures computed as a function of temperature and pressure using Eq. 2.1 from MC simulations are shown in Fig. 2.10. Since c is a function of ρ , β_T , c_P , and c_V , uncertainties associated with c are computed using the Eq. (S78) derived in Section S11 of the Supporting Information for this chapter [147]. Speed of sound of pure CO₂ computed at temperature 253 K, 273 K, 293 K, and 313 K results in overprediction compared to c obtained using the Span and Wagner EoS. Since c depends on ρ , β_T , c_P , and c_V , small deviations in these dependent properties will overestimate c . The overestimation of c is majorly due to the minor overestimation of c_P and underestimation of β_T . For example, c of pure CO₂ is overestimated by a relative deviation of ca. 6.7% at 253 K and 200 bar, while the overestimation of c_P is ca. 3.6%, and the underestimation of β_T is ca. 8.4%. The magnitude of overprediction was found to be less than 10% for all temperatures and pressures. Fig. 2.10 (a), shows that c decreases marginally till the pressure reaches its critical pressure and increases significantly for the pressure larger than the saturation pressure for a particular temperature. Fig. 2.10 (a), also shows that c is temperature dependent and decreases with increasing temperature. The impact of impurities on c are evaluated by comparing c computed from MC simulation and the GERG-2008 EoS for binary mixtures consisting of 95 mol% of CO₂ and 5 mol% of one of the impurities (CH₄, Ar, N₂, and H₂) as shown in Fig. 2.10 (b). In the liquid and supercritical phases, the presence of H₂ as an impurity is observed to have the largest impact on decreasing the value of c , followed by CH₄, Ar, and N₂. In the gas phase, the presence of H₂ as an impurity is observed to have the largest impact on increasing the value of c followed by CH₄, Ar, and N₂. The computed c of binary mixtures rich in CO₂ with 1 mol% and 10 mol% of impurity (CH₄, Ar, N₂, and H₂) for temperatures 253 K, 273 K, 293 K, and 313 K from MC simulations and the GERG-2008 EoS, as listed in Tables S73-S167 of the Supporting Information for this chapter [147] shows the same behavior with respect to type of impurity. Speed of sound of ternary mixtures with 96 mol% CO₂ and 2 mol% for each of the two impurities (CH₄, Ar, N₂, and H₂) at 253 K are shown in Fig. 2.10 (c). MC simulations overpredicted the value of c compared to the GERG-2008 EoS. However, MC simulations closely predicted the decrease in the value of c due to the presence of impurities. For instance, difference in the value of c at 200 bar between pure CO₂ and ternary CO₂ mixture with N₂ and H₂ as impurities is ca. 56 m/s from MC simulations and ca. 46 m/s from the GERG-2008 EoS.

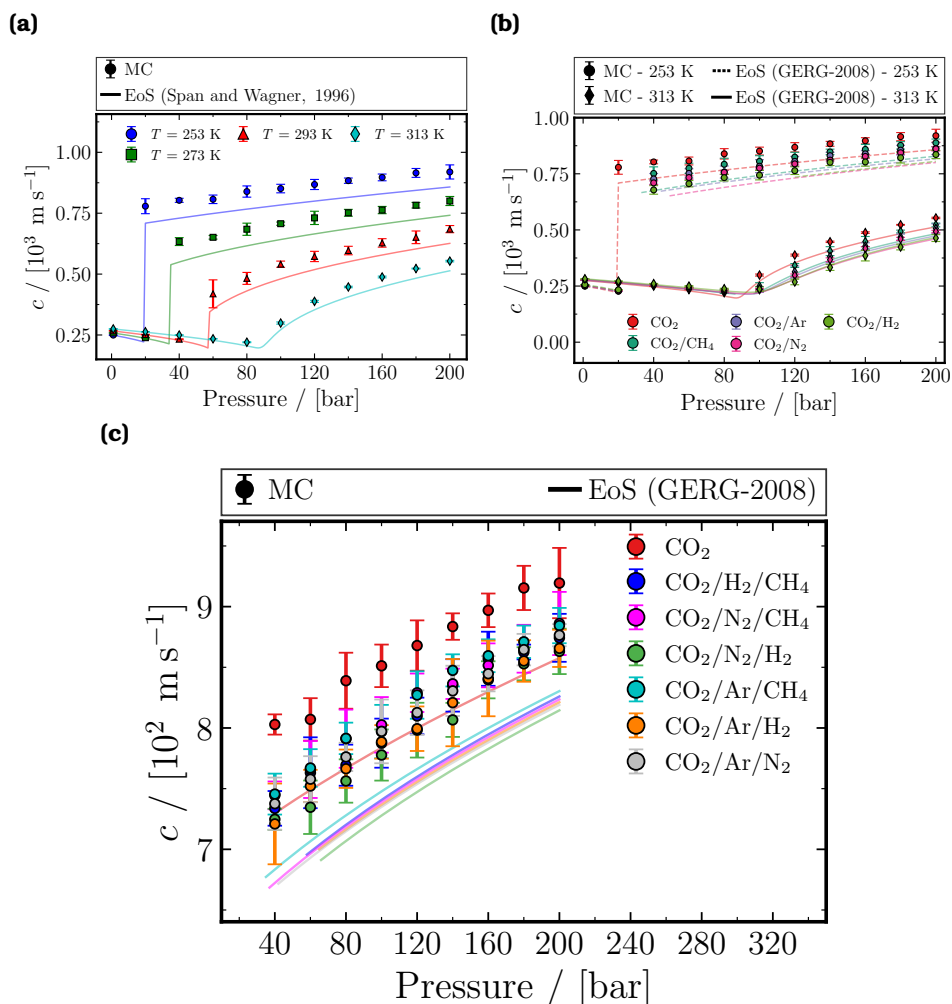


Figure 2.10. Computed speed of sound as a function of temperature and pressure. (a) shows the calculated speed of sound of pure CO₂ from MC simulations (closed symbols) and the Span and Wagner EoS [28] (solid lines) for temperatures 253 K, 273 K, 293 K, and 313 K. (b) shows speed of sound of binary mixtures with 95 mol% of CO₂ and 5 mol% of impurities (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with speed of sound of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] for temperatures 253 K and 313 K. (c) shows the speed of sound of ternary mixtures with 96 mol% of CO₂ and 2 mol% impurities for each of two components (CH₄, Ar, N₂, and H₂) computed from MC simulations (closed symbols) and the GERG-2008 EoS [58] (lines) compared with speed of sound of pure CO₂ computed from MC simulations and the Span and Wagner EoS [28] at 253 K.

2.4.2. Transport Properties

Viscosities (η)

Viscosities (η) of pure CO₂, binary CO₂ mixtures, and ternary CO₂ rich mixtures computed from MD simulations as a function of temperature and pressure are shown in Fig. 2.11. Viscosities of pure CO₂ were computed at temperatures 253 K, 273 K, 293 K, and 313 K. The computed viscosities are compared with those calculated from the correlation of Laesecke and Muzny [62]. Our results show a good agreement with the model for all temperatures as seen in Fig. 2.11 (a) with a maximum relative deviation of ca. 14.6% in the gas phase (at 313 K and 40 bar), excluding state points close to the saturation/Widom line. η at pressures below the saturation pressure for a temperature remains relatively constant, but for pressure far away from the saturation/Widom line, η increases with increasing pressure at a constant temperature. η is observed to be highly dependent on temperature compared to pressure as seen in Fig. 2.11 (a), where η decreases with increasing temperature. To analyze the effect of impurities, η of binary mixtures with 95 mol% of CO₂ and 5 mol% of one of the impurities (CH₄, Ar, N₂, and H₂) computed from MD simulations at 253 K and 313 K are compared with viscosities obtained from REFPROP in Fig. 2.11 (b). Computed η of binary CO₂ mixtures qualitatively correlate well with η obtained from REFPROP. The existence of impurity in a CO₂ mixture tends to reduce η compared to η of pure CO₂. Based on η computed from MD simulations and REFPROP, it is clear that a binary mixture with H₂ as impurity decreases η the most in liquid and supercritical phases. The uncertainties range of computed viscosities from MD simulations made it difficult to interpret a particular type of impurity that impacts η the most next to H₂. However, viscosities obtained from REFPROP indicate that a binary mixture with N₂ as an impurity decreases η after H₂ followed by CH₄, and Ar for pressure away from the saturation/Widom line at 253 K and 313 K. In addition to viscosities data of binary mixtures with 5 mol% impurities, binary mixtures with 1 mol% and 10 mol% of impurities (CH₄, Ar, N₂, and H₂) at temperatures 253 K, 273 K, 293 K, and 313 K computed from MD simulations are listed along with data obtained from REFPROP in Tables S73-S167 of the Supporting Information for this chapter [147]. To analyze the effect on η due to the presence of a particular combination of impurities in CO₂ rich mixtures, η of ternary mixtures were computed from MD simulations and compared with η obtained from REFPROP at 253 K, as shown in Fig. 2.11 (c). η of all ternary mixtures shown in Fig. 2.11 (c) has a concentration of 96 mol% CO₂ and 2 mol% for each of the two impurities (CH₄, Ar, N₂, and H₂). Comparing η of ternary mixtures obtained from REFPROP, it is clear that mixtures with H₂ as an impurity reduce the liquid viscosities of CO₂ mixtures to a larger extent compared to mixtures without H₂. A similar qualitative trend of η was also observed from simulations at 40, 60, and 80 bar. For higher pressures, considering the computed uncertainties, the effect of a particular combination of impurities

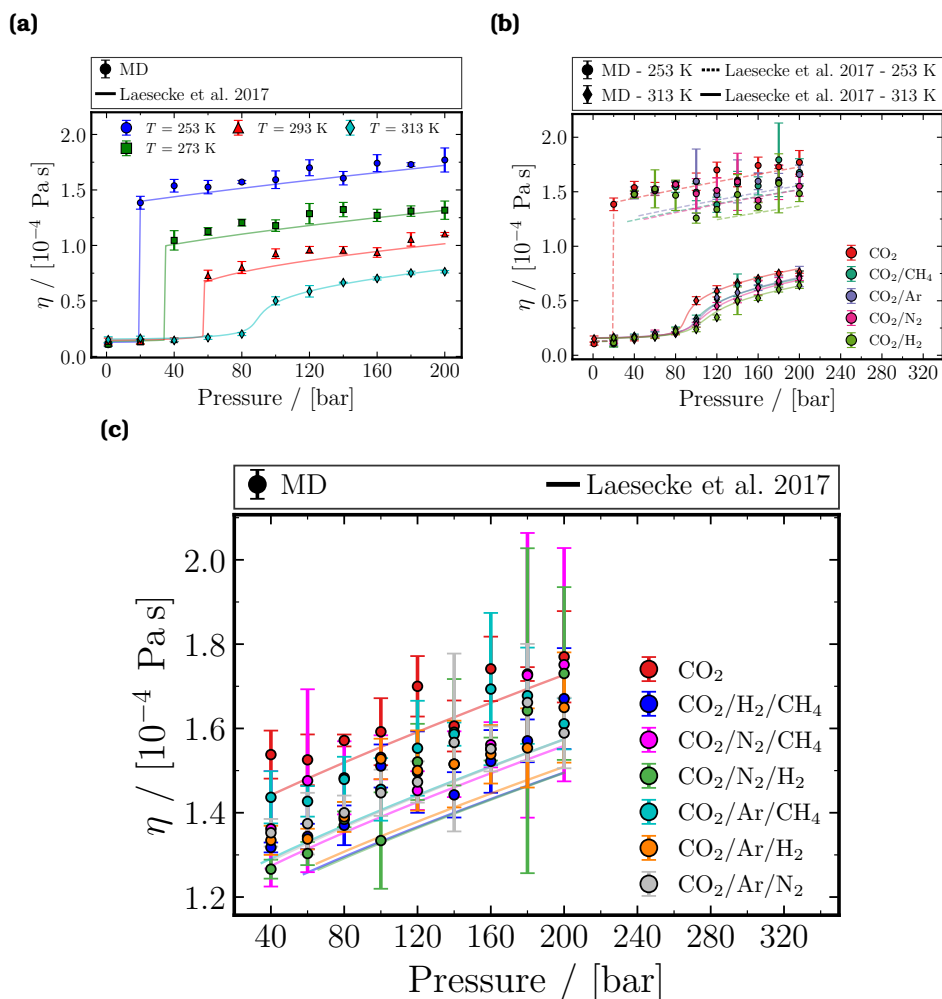


Figure 2.11. Computed viscosities as a function of temperature and pressure. (a) shows the calculated viscosities of pure CO₂ computed from MD simulations (closed symbols) compared with the correlation of Laesecke et al. (2017) [62] (solid lines) for temperatures 253 K, 273 K, 293 K, and 313 K. (b) shows viscosities of binary mixtures with 95 mol% of CO₂ and 5 mol% of impurities (CH₄, Ar, N₂, and H₂) computed from MD simulations (closed symbols) and REFPROP [190] (lines) compared with viscosities of pure CO₂ computed from MD simulations and REFPROP [190] for temperatures 253 K and 313 K. (c) shows viscosities of ternary mixtures with 96 mol% of CO₂ and 2 mol% impurities for each of two components (CH₄, Ar, N₂, and H₂) computed from MD simulations (closed symbols) and REFPROP [190] (lines) compared with viscosities of pure CO₂ computed from MD simulations and REFPROP [190] at 253 K.

was difficult to evaluate for a marginal decrease in the liquid viscosities. Uncertainties associated with viscosities can be reduced by performing multiple independent simulations, but this will increase computational costs [206]. We refrained from conducting additional simulations due to the marginal difference in liquid viscosities observed between ternary mixtures in Fig. 2.11 (c).

In summary, we performed molecular simulations of CO₂ rich mixtures with N₂, Ar, CH₄, and H₂ as impurities using force fields mentioned in Section S1 of the Supporting Information for this chapter [147]. The thermodynamic and transport properties were computed at temperature 253 K, 273 K, 293 K, and 313 K and pressure up to 200 bar for pure CO₂ mixture, binary mixtures rich in CO₂ with 1 mol%, 5 mol%, and 10 mol% impurities. The computed thermodynamic and transport properties were found to be in good agreement with EoS for pure and binary systems, except at conditions close to the saturation/Widom line. The thermodynamic and transport properties were also computed for 24 ternary and 12 quaternary CO₂ rich mixtures for various concentrations of impurities listed in Tables S169 and S362 of the Supporting Information for this chapter [147], respectively, for temperature 253 K, 273 K, 293 K, and 313 K and pressure up to 200 bar. Results of the thermodynamic and transport properties of ternary and quaternary mixtures are provided in Sections S16 and S17 of the Supporting Information for this chapter [147], respectively. The thermodynamic and transport properties of multicomponent CO₂ mixtures were compared with the GERG-2008 EoS [58] and the ECS model [197], respectively, and showed good agreement. We show that molecular simulations are a powerful tool to compute the thermodynamic and transport properties of multicomponent mixtures seen in CO₂ transportation with a smaller system of 300 molecules. These thermophysical properties will help in modeling and designing pipelines for CO₂ transportation, which will be the focus of further work.

2.5. Conclusions

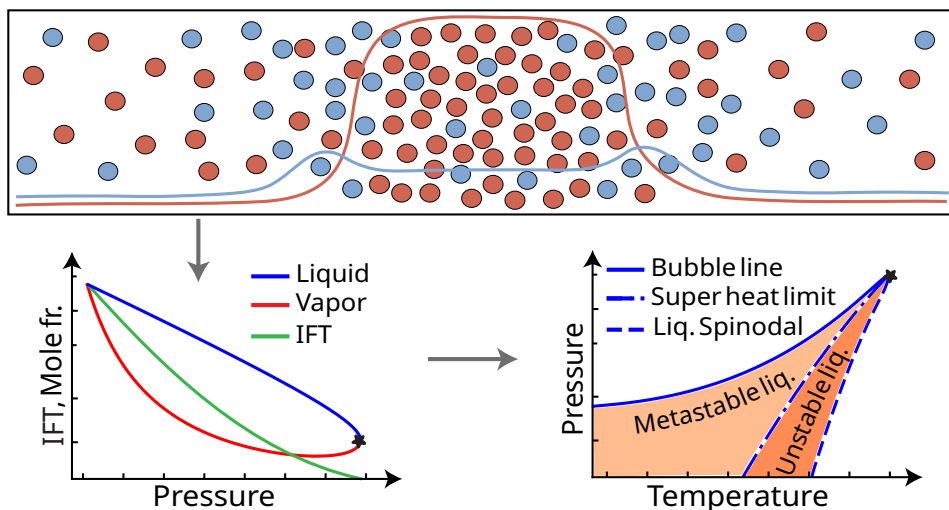
In this chapter, the effect of impurities in CO₂ rich mixtures on the value of thermodynamic and transport properties such as densities, thermal expansion coefficients, isothermal compressibilities, heat capacities at constant pressure, heat capacities at constant volume, Joule-Thomson coefficients, speed of sound, and viscosities were investigated using molecular simulations. The CFCMC method was used to compute the VLE of pure components such as CO₂, CH₄, Ar, N₂, and H₂ to validate force fields used in molecular simulations. The computed VLE of pure components showed an excellent agreement with EoS [28, 207–210]. The phase equilibria of CO₂/Ar, CO₂/CH₄, CO₂/N₂, and CO₂/H₂ binary mixtures were also computed using the CFCMC method and compared with the GERG-2008 EoS [58] and data from the literature [198–202], showing a good agreement.

The thermodynamic and transport properties were computed for pure CO₂ and binary and ternary mixtures rich in CO₂ at temperatures 253 K, 273 K, 293 K, and 313 K and for pressures ranging from 20–200 bar using MC and MD simulations. The computed thermodynamic and transport properties of pure CO₂ are in excellent agreement with corresponding values obtained from the Span and Wagner EoS [28]. Thermodynamic and transport properties of CO₂ rich binary mixtures with 1 mol%, 5 mol%, and 10 mol% concentrations of noncondensable impurities such as CH₄, Ar, N₂, and H₂ computed from simulations showed a good agreement with the GERG-2008 EoS [58]. The computed thermodynamic and transport properties of CO₂ rich ternary mixtures with various impurities were compared with the GERG-2008 EoS [58], showing a good agreement. The effect of different types of impurities on a specific thermodynamic and transport property was evaluated. Our findings show that CO₂ rich mixtures with impurities have low densities compared to densities of pure CO₂. The magnitude of reduction in densities of a CO₂ rich mixture depends strongly on the molecular weight of impurities present a mixture. Mixtures with molecular weight lower than pure CO₂ were observed to have reduced densities than pure CO₂. CO₂ rich mixtures containing H₂ as an impurity led to the most significant decrease in the value of thermal expansion coefficients, isothermal compressibilities, heat capacities at constant pressure, and Joule–Thomson coefficients followed by N₂, Ar, and CH₄ in the gas phase. In the liquid and supercritical phases, the presence of H₂ as an impurity led to the most significant increase in the value of thermal expansion coefficients, isothermal compressibilities, heat capacities at constant pressure, and Joule–Thomson coefficients followed by N₂, Ar, and CH₄. In contrast, the presence of H₂ as an impurity in CO₂ rich mixture increased the value of speed of sound in the gas phase and decreased in the liquid and supercritical phases. The order of effect due to a particular impurity on thermal expansion coefficients, isothermal compressibilities, heat capacities at constant pressure, Joule–Thomson coefficients, and speed of sound correlates with the critical temperature of impurities. In our investigation of heat capacities at constant volume, we found that the presence of impurities did not have a significant impact. Finally, differences in the value of viscosities in CO₂ rich mixtures due to the presence of impurities were evaluated. Our findings showed that mixtures containing H₂ as impurity significantly reduced viscosities in liquid and supercritical phases.

3

Vapor–Liquid Interfacial Properties of CO₂ Binary Mixtures

This chapter is based on the following publication: D. Raju, R. Skartlien, M. Ramdin and T. J. H. Vlugt. ‘Vapor–Liquid Interfacial Properties of CO₂ Mixtures for Sequestration Applications: Molecular Simulations, Classical Density Functional Theory, and Equations of State’. *Industrial & Engineering Chemistry Research* 65 (2026), pp. 5213–5238.



3.1. Introduction

Carbon dioxide CO₂ is a major greenhouse gas that drives global warming [148, 212]. Carbon capture and storage (CCS) has become an essential process to reduce CO₂ emissions [7, 9]. Efficient transportation of CO₂, either by pipelines or ships, plays a central role in connecting capture facilities with storage reservoirs [37]. The CO₂ obtained from capture processes is rarely pure and its composition depends on the capture technology and the source stream, often carrying trace components from flue gas or syngas [35]. Major impurities present in the CO₂ streams are noncondensable impurities such as Ar, N₂, H₂, and CH₄ [19, 27, 43, 46, 47]. These impurities can affect phase stability. For economic transportation, CO₂ streams must be maintained in a dense single phase [49, 53, 157]. This requirement is particularly challenging during transient operations, where pressure and temperature fluctuate and impurities can shift the phase boundaries [35, 213]. While steady-state operation targets a dense phase, systems often encounter startup, shutdown, depressurization, and planned maintenance, which can lead to the occurrence of two-phase flows in CO₂ transportation systems [214]. At such conditions, accurate knowledge of the thermodynamic and transport properties of impure CO₂ mixtures becomes indispensable for modeling their flow and phase behavior [35, 60].

Experimental determination of thermodynamic and transport properties for multicomponent mixtures is costly and time-consuming, resulting in limited and scarce experimental data for impure CO₂ mixtures [35, 36, 44, 215]. In addition to experimental measurements for these systems, molecular simulations with classical force fields have proven to be a powerful tool for computing thermodynamic and transport properties [147, 158, 169, 216]. Experiments and molecular simulations are complemented by models such as Equation of State (EoS) [28, 58, 65, 102], Extended Corresponding States (ECS) [77], friction theory [217], and other empirical correlations [141, 218, 219]. For instance, Lasala *et al.* [91] measured the phase equilibria of CO₂-N₂, CO₂-Ar, and CO₂-O₂ binary mixtures at 218–253 K and pressures up to 160 bar, and successfully modeled the phase equilibria with Peng–Robinson (PR) and Soave–Redlich–Kwong (SRK) cubic EoS using advanced mixing rules. In Chapter 2, we performed Monte Carlo (MC) and Molecular Dynamics (MD) simulations of multicomponent CO₂ mixtures to compute densities, thermal expansion coefficients, isothermal compressibilities, heat capacities, Joule–Thomson coefficients, speed of sound, and viscosities at 235–313 K and pressures between 20–200 bar, and compared with the Groupe Européen de Recherches Gazières (GERG-2008) EoS [58]. In addition to bulk thermodynamic and transport properties, interfacial properties of CO₂ mixtures have also been investigated in numerous studies [220–224]. To illustrate a few, Chow, Maitland and Trusler [222] measured interfacial tensions of H₂O–H₂ and H₂O–H₂–CO₂ mixtures at 298–448 K and pressures up to 450 bar and developed empirical correlations to calculate Interfacial Tensions (IFTs). In a previous

study, Chow *et al.* [220] measured IFTs of $\text{H}_2\text{O}-\text{CO}_2$, $\text{H}_2\text{O}-\text{Ar}$, $\text{H}_2\text{O}-\text{N}_2$, $\text{H}_2\text{O}-\text{CO}_2-\text{N}_2$, and $\text{H}_2\text{O}-\text{CO}_2-\text{Ar}$ mixtures at 298–473 K and pressures up to 600 bar and compared with Statistical Associating Fluid Theory (SAFT)-VR Mie EoS [69, 225] in combination with Square Gradient Theory (SGT) [111, 226, 227]. The SAFT-VR Mie EoS [69, 225] combined with SGT [111, 226, 227] predicted IFTs for systems containing H_2O , CO_2 , Ar, and N_2 in good agreement with experiments. Numerous studies have measured IFTs of binary and ternary systems such as CO_2 -water [220, 222–224], CO_2 -brine [221], H_2O -dilute gases (N_2 , Ar, and H_2) [222] and CO_2 -hydrocarbon mixtures [228], and compared with theoretical models such as classical Density Functional Theory (cDFT) [70, 108, 229, 230], Density Gradient Theory (DGT) [226, 231], SGT [111, 226, 227], Linear Gradient Theory (LGT) [232], and the Parachor model [233, 234] in combination with EoS. Most of these studies have focused on reservoir-relevant systems where water is the dominant component in the system. IFTs of pipeline-relevant mixtures have been scarcely investigated, where H_2O concentrations must remain below 500 ppm to prevent hydrate formation and corrosion inside CO_2 transportation systems [27, 235]. Experimental and molecular simulation data on IFT for binary CO_2 mixtures containing dilute gases are extremely limited. To the best of our knowledge, the only available study is by Schenk *et al.* [236], in which IFTs of CO_2-CH_4 mixtures at 223–233 K and 20–60 bar were computed using both experiments and molecular simulations, and good agreement was found between the two approaches.

The vapor-liquid interfacial tension (IFT) is an important property in governing bubble stability, droplet formation, and interphase heat and mass transfer [237–240] in CO_2 transportation systems. For example, during depressurization of a CO_2 -rich mixture through valves or nozzles, the liquid or dense-phase CO_2 -rich stream may experience a delayed phase change and remain in a metastable state instead of vaporizing spontaneously [241, 242]. In Computational Fluid Dynamics (CFD) simulations, neglecting this delayed phase change leads to inaccurate predictions of mass flow rates and pressure, which in turn results in incorrect evaluation of forces on pipelines and the associated risk of Running Ductile Fracture (RDF) [235, 243]. Phase change delay is modeled in CFD simulations by estimating the metastable limits of the fluid [57, 241]. These states exist between the binodal (saturation line) and the theoretical thermodynamic limit of stability, known as the spinodal [238]. Beyond the spinodal, the fluid can no longer remain metastable, and phase separation occurs spontaneously (unstable regime) [237, 238]. In practice, however, phase change typically sets in earlier than the spinodal, at the so-called Superheat Limit (SHL) for liquids and Supercool Limit (SCL) for vapors, which represent practical metastable limits [241]. Phase transitions in these metastable regions occur via either homogeneous or heterogeneous nucleation [237, 238, 242]. Heterogeneous nucleation is facilitated by surfaces and impurities, and its

rate is strongly affected by geometrical factors such as the surface roughness [237, 238]. It is typically represented in models through empirical correlations tailored to the system geometry [237, 238]. Homogeneous nucleation arises within the bulk fluid and is commonly described using Classical Nucleation Theory (CNT) [57, 235, 241, 244]. In CNT, the nucleation rate is expressed in an Arrhenius form, with the free energy barrier for forming a critical nucleus (bubble or droplet) evaluated using the capillarity approximation, which assumes that the IFT of a curved interface is equal to that of the planar interface [237]. The SHL and SCL are then estimated by applying a nucleation rate criterion, namely by identifying the conditions at which CNT predicts that at least one critical nucleus forms within the system volume over the relevant observation time [237]. Since this estimation relies strongly on the accuracy of IFT data [235], reliable prediction of both IFTs and metastability limits is essential for modeling the multiphase flow behavior in CO₂ transport systems.

In this chapter, we computed the vapor-liquid interfacial properties of pure CO₂ and four binary mixtures, namely CO₂-N₂, CO₂-Ar, CO₂-CH₄, and CO₂-H₂ using MD simulations, the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) EoS combined with the cDFT [245], and semi-empirical models [146, 219, 233]. To compute IFTs of pure components, the Guggenheim-Katayama equation [63, 219, 246] were used. The Parachor [233, 234] and Winterfeld-Scriven-Davis (WSD) [146, 247] models were used to compute IFTs of binary mixtures. To quantify the effect of impurities, the computed IFTs of the binary mixtures were compared with those of pure CO₂. The IFTs obtained from semi-empirical models were then used with phase equilibria to estimate the SHL and SCL for both pure CO₂ and the CO₂ binary mixtures using the CNT. Using this approach, phase equilibria, IFT, and metastable limits are examined by combining EoS based methods, cDFT, and MD simulations. This enables the influence of impurities on SHL and SCL to be quantified through direct comparison with pure CO₂ superheat and supercool limits. This chapter presents a comprehensive evaluation of models for computing IFTs, ranging from molecular simulations to simplified theoretical approaches, with direct relevance to CO₂ transport systems.

This chapter is organized as follows: the methodology used to compute phase equilibria and IFTs using MD simulations, PC-SAFT EoS with the cDFT, and semi-empirical models are explained in Section 3.2. In Section 3.3, we present the validation and the results of the computed phase equilibria and IFTs of pure CO₂ and binary CO₂ mixtures. We also analyze the effect of impurities on IFT and report the IFTs obtained from the Parachor and WSD models. In Section 3.4, our key findings are summarized.

3.2. Methodology

3.2.1. MD Simulations

MD simulations were performed with the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS: version August 2024) package [179] to compute IFTs of pure CO₂ and CO₂ mixtures with CH₄, N₂, Ar, and H₂ as impurities. The TraPPE [81, 82] force fields were used to model CO₂ and CH₄. N₂ and H₂ are simulated using force fields of Vrabec, Stoll and Hasse [113] and Köster, Thol and Vrabec [115], respectively. The force field from García-Pérez *et al.* [116] was used for Ar. CO₂ and N₂ are represented by three-site and two-site flexible models, respectively. For CO₂, bond-stretching and angle-bending terms were included to make the TraPPE [82] force field fully flexible, with parameters adopted from Perez-Blanco and Maginn [181]. The bond-stretching parameter for N₂ was taken from Wang, Hou and Heinz [248]. Force fields of H₂, Ar, and CH₄ are described by single-site models consisting of one Lennard-Jones (LJ) interaction site with no point charges. The force field parameters of CO₂, Ar, N₂, H₂, and CH₄ are listed in Section S1 of the Supporting Information for this chapter [211]. Intermolecular interactions were modeled by a pairwise-additive 12-6 Lennard-Jones (LJ) potential combined with long-range Coulombic interactions. The corresponding total potential energy U of the system is given as [79, 80]:

$$U = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}} \right) + \sum_{\text{bonds}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2, \quad (3.1)$$

where N denotes the number of atoms, ε_{ij} is the LJ energy parameter between atoms i and j , σ_{ij} is the LJ size parameter between atoms i and j , q_i and q_j are the partial charges of atoms i and j , respectively. ε_0 is the permittivity of free space. k_r and k_θ are the bond and angle force constants. r and θ are the bond length and bond angle. r_0 and θ_0 are the equilibrium bond length and equilibrium value of the angle. The first term in Eq. 3.1 corresponds to van der Waals interactions (intermolecular), and the second term represents Coulombic interactions (intermolecular). The third and fourth terms describe bond stretching and angle bending, modeled using harmonic potentials. For dissimilar atom pairs, the Lorentz-Berthelot (LB) [79, 80] mixing rules were used

$$\sigma_{ij} = (1 - l_{ij}) \frac{1}{2} (\sigma_{ii} + \sigma_{jj}), \quad (3.2)$$

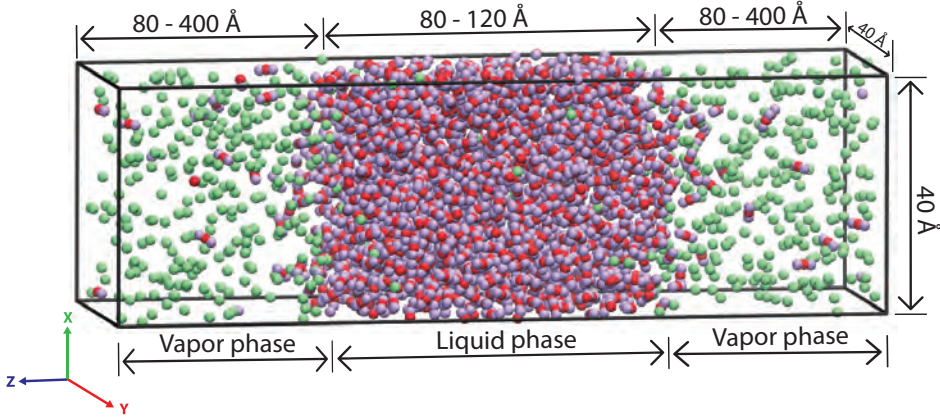
$$\varepsilon_{ij} = (1 - k_{ij}) (\varepsilon_{ii} \varepsilon_{jj})^{1/2}, \quad (3.3)$$

where the indices ii and jj denote interactions between atoms of the same type, while ij refers to cross interactions between atoms of different types.

The parameters l_{ij} and k_{ij} are the binary interaction parameters used in the mixing rules for σ_{ij} and ε_{ij} , respectively, for components i and j . In this chapter, l_{ij} is not considered ($l_{ij} = 0$), and only k_{ij} is considered. In MD simulations, binary interaction terms were set to zero, $k_{ij}^{\text{MD}} = 0$.

The equations of motion were integrated using the velocity-Verlet algorithm with a time step of 0.5 fs [133]. A cutoff radius of 16 Å was used for the short-range LJ and long-range Coulombic interactions. Since this value exceeds $5\sigma_{\text{CO}_2}$, no tail corrections were applied for energy and pressure [249]. This choice ensures an appropriate treatment of system inhomogeneity while maintaining MD simulations within an acceptable computational time. Long-range electrostatic interactions were handled using the Particle-Particle Particle-Mesh (PPPM) method with a relative accuracy of 10^{-5} [79, 250]. To compute the vapor-liquid IFT, simulations were carried out in a parallelepiped box containing a liquid slab at the center with vapor regions on either side [126] as shown in Fig. 3.1. The z -axis was chosen as the elongated dimension (normal to the vapor-liquid interface), with $L_x = L_y = 40$ Å, while L_z was varied between 5 and 25 times L_x , depending on liquid and vapor-phase densities. The liquid slab thickness ranged from 80 to 120 Å, and periodic boundary conditions were applied in all directions. Temperature and pressure were controlled using a Nosé-Hoover thermostat and barostat [79, 251, 252] with coupling constants of 0.1 ps and 1 ps, respectively. Initial configurations were generated in LAMMPS using liquid and vapor densities calculated from the multiparameter Helmholtz EoS of Span and Wagner [28] for pure CO₂ and the GERG-2008 [58] multiparameter Helmholtz EoS for binary CO₂ mixtures, as implemented in the REFPROP v10 [190]. According to the Gibbs phase rule [175], a pure-component system is characterized by a single independent thermodynamic variable, whereas a binary system requires two. In this chapter, IFTs of pure CO₂ were computed as a function of temperature with $N = 2000$ – 2500 molecules. For CO₂ mixtures IFTs were evaluated as functions of temperature and pressure with $N = 2500$ – 4000 molecules. Each simulation consisted of a 7–10 ns equilibration run followed by a 40 ns production run in the NVT ensemble for computing the IFTs of pure CO₂. For CO₂ mixtures, an equilibration run of 10–15 ns followed by a production run of 50–60 ns was performed in the $NAP_{zz}T$ ensemble for each simulation. In this ensemble, the interfacial area $A = L_x L_y$ is kept fixed, while the z -dimension of the simulation box is allowed to fluctuate to maintain a stable interface in the xy plane [126].

The IFT of a system is computed either from the ensemble-averaged diagonal components of the pressure tensor over the entire simulation box or from pressure profiles as a function of the z -coordinate. In this chapter, the ensemble-averaged diagonal pressure tensor components computed



3

Figure 3.1. Schematic representation of the MD simulation box used in this chapter to compute phase equilibria and interfacial properties of pure CO_2 and binary CO_2 mixtures. For pure CO_2 , the simulation box dimensions were fixed at $L_x = L_y = 40 \text{ \AA}$ and $L_z = 240 \text{ \AA}$, as the simulations were performed in the NVT ensemble. For binary mixtures, L_x and L_y were fixed, while L_z was allowed to vary since the simulations were conducted in the NPT ensemble. The example shown here illustrates the $\text{CO}_2\text{-H}_2$ mixture at 220 K and 100 bar. The visualization was rendered in orthographic view using the Visual Molecular Dynamics (VMD) software. [253]

over the whole volume of the simulation box were used to calculate IFTs, where the IFT (γ) is given by

$$\gamma = \frac{\langle L_z \rangle}{2} \left[\langle P_{zz} \rangle - \frac{\langle P_{xx} \rangle + \langle P_{yy} \rangle}{2} \right]. \quad (3.4)$$

The brackets $\langle \dots \rangle$ denote an ensemble average. $\langle P_{xx} \rangle$ and $\langle P_{yy} \rangle$ are the ensemble-averaged tangential components of the pressure tensor, while $\langle P_{zz} \rangle$ is the ensemble-averaged normal component. All components are evaluated over the entire simulation box volume with $\langle P_{zz} \rangle$ corresponding to the saturation pressure. $\langle L_z \rangle$ represents the box length averaged in the NPT ensemble and is used to compute the IFTs of binary systems. For pure systems, L_z is fixed (constant value) since simulations are performed in the NVT ensemble. As an alternative to Eq. 3.4, IFTs of pure and binary systems can be computed using the Irving-Kirkwood formulation [126, 128]:

$$\gamma = \frac{1}{2} \int_{-\infty}^{+\infty} \left[\langle P_{zz}(z) \rangle - \frac{\langle P_{xx}(z) \rangle + \langle P_{yy}(z) \rangle}{2} \right] dz, \quad (3.5)$$

$\langle P_{xx}(z) \rangle$, $\langle P_{yy}(z) \rangle$, and $\langle P_{zz}(z) \rangle$ denote the ensemble-averaged pressure tensor profiles as a function of z -coordinate. To compute these profiles, the

simulation box was divided into 100 to 300 bins along the z -coordinate. The factor of 2 in Eqs. 3.4 and 3.5 accounts for the two interfaces present in the simulation box. Eqs. 3.4 and 3.5 yield identical IFTs for a given thermodynamic state, and this has been verified for pure CO₂ in Section S5 of the Supporting Information for this chapter [211]. The same binning approach was also applied to calculate the density profiles, $\rho(z)$, along the z axis. The bulk liquid density (ρ^L) and vapor density (ρ^V) were computed by fitting $\rho(z)$ to the following hyperbolic tangent function [126]:

3

$$\rho(z) = \rho^V + \frac{\rho^L - \rho^V}{2} \left[\tanh\left(\frac{z - z_R}{D}\right) - \tanh\left(\frac{z - z_L}{D}\right) \right] \quad (3.6)$$

where z_R and z_L denote the position of the vapor-liquid interface (Gibbs-dividing surface) on the left and right sides [126]. D is the interface width, and it is related to the 10/90 effective interfacial thickness (L_{10}^{90}) definition proposed by Lekner and Henderson [254] via the relation $L_{10}^{90} \approx 2.19D$ [249]. The parameter L_{10}^{90} is defined as the distance between the positions where the local density reaches 10% and 90% of the bulk density differences:

$$L_{10}^{90} = z(\rho_{90}) - z(\rho_{10}), \quad (3.7)$$

$$\rho_{10} = \rho^V + 0.1(\rho^L - \rho^V), \quad (3.8)$$

$$\rho_{90} = \rho^V + 0.9(\rho^L - \rho^V). \quad (3.9)$$

Computing bulk densities of binary mixtures by fitting the density profiles $\rho_i(z)$ to Eq. 3.6 can be inaccurate due to the interfacial enrichment near the interface [255]. To overcome this limitation, we used a plateau-detection method. In this approach, the spatial derivative of the average $\rho_i(z)$ is evaluated as a function of the z coordinate. Bulk regions are identified as plateaus where the derivative $d\rho/dz$ remains close to zero within a relative tolerance of 10^{-5} , corresponding to constant density in either the vapor or liquid phase. By isolating and averaging these plateau regions, the saturated vapor and liquid densities are obtained without being biased by possible interfacial enrichment. The computed bulk densities were used to calculate the mass fractions of the liquid and vapor phases [255]. For a binary mixture with components $i = 1, 2$, the mass fraction w^k of component i in phase k is defined as

$$w_i^k = \frac{\rho_i^k}{\sum_{j=1}^2 \rho_j^k}, \quad k \in \{L, V\}, \quad (3.10)$$

where ρ_i^k is the bulk density of component i in phase k . The mole fractions were obtained from the liquid and vapor mass fractions using

$$x_i^L = \frac{w_i^L/M_i}{\sum_{j=1}^2 w_j^L/M_j}, \quad y_i^V = \frac{w_i^V/M_i}{\sum_{j=1}^2 w_j^V/M_j}, \quad (3.11)$$

where M_i is the molar mass of the component i .

3.2.2. Classical Density Functional Theory + PC-SAFT

The interfacial properties of pure CO₂ and CO₂ mixtures were computed using the FeO_s (FeO_s) v0.8 framework [245], in which EoS models are integrated with cDFT. EoS models are used to compute thermodynamic properties of homogeneous phases (phase equilibria) [36, 215], but the modeling of inhomogeneous systems with density gradients, such as fluid interfaces, is not possible [60, 127]. By coupling EoS with cDFT, the FeO_s [245] approach enables the modeling of spatial inhomogeneities. In cDFT, the liquid–vapor interface is represented as a mesoscopic region where the density smoothly varies from the bulk liquid to the bulk vapor value [70, 245]. The theory is formulated by expressing the grand potential as a functional of the density profile, which incorporates both ideal and nonideal interaction contributions [70, 245]. The equilibrium density profile is then obtained by minimizing this functional, i.e., by setting its functional derivative with respect to the density to zero, which enforces a constant chemical potential across the interface [70, 245]. Once the equilibrium profile is known, interfacial properties such as density gradients, interfacial thicknesses, and IFTs can be directly computed [70, 245]. In this chapter, the PC-SAFT [68, 102] was used as the EoS model. The PC-SAFT EoS is a molecular-based approach that represents fluids as chains of spherical segments and accounts for dispersion and association interactions. For pure CO₂ and binary CO₂ mixtures, a simplified PC-SAFT EoS formulation without explicit quadrupolar contributions was employed. In this formulation, quadrupolar effects are implicitly implemented through fitting of the pure-component PC-SAFT parameters to experimental data [256]. This simplified treatment may contribute to deviations in the predicted critical properties [103]. For more details regarding the cDFT and PC-SAFT EoS, the reader is referred to Refs. [68, 70, 108, 229, 230].

The phase equilibria of pure CO₂ and CO₂ binary mixtures were computed using the PC-SAFT EoS [68, 102]. For binary mixtures, the optimized binary interaction parameters, k_{ij} , were required in the Lorentz–Berthelot (Eq. 3.3) mixing rules to accurately model the experimental vapor–liquid equilibrium (P_{xy}) data available in the literature [91, 198, 200–202, 257, 258]. In FeO_s [245], the optimization was carried out by adjusting k_{ij} using nonlinear least-squares minimization, where the calculated equilibrium pressures and compositions were fitted to reproduce the experimental data. The quality of each fit was evaluated through the Mean Absolute Percentage Deviation (MAPD), defined as

$$\text{MAPD} = \frac{1}{N} \sum_{n=1}^N \left| \frac{q_n^{\text{EoS}} - q_n^{\text{exp}}}{q_n^{\text{exp}}} \right| \times 100\%, \quad (3.12)$$

where q denotes the target property (pressure, composition distance, or chemical potential) and N is the total number of data points. The k_{ij} value corresponding to the lowest MAPD was selected and used in the FeO_s framework to compute IFTs of CO₂ mixtures. The PC-SAFT parameters of

CO₂, CH₄, N₂, Ar, and H₂ used to compute Pxy diagram, IFTs are listed in Section S2 of the Supporting Information for this chapter [211].

3.2.3. Parachor Model

IFTs of pure fluids were evaluated using the Guggenheim–Katayama equation [190, 219, 259, 260] (as implemented in the REFPROP v10 program by the National Institute of Standards and Technology) expressed as

$$\gamma(T) = \sum_{i=0}^{k-1} s_i \left(1 - \frac{T}{T_c}\right)^{n_i}, \quad (3.13)$$

where k denotes the total number of summation terms, s_i and n_i are fluid-specific coefficients, and T_c is the critical temperature of the pure component [63, 246]. The fluid-specific parameters k , s_i , n_i , and T_c used in this chapter were taken from the works of Cachadiña, Mulero and Tian [246] and Mulero, Cachadiña and Parra [63], which are listed in Section S3 of the Supporting Information for this chapter [211]. IFTs of mixtures (γ_{mix}) was calculated using the Parachor model [145, 233, 234, 261], expressed as

$$\gamma_{\text{mix}} = \left(\rho_{\text{mix}}^L \sum_{j=1}^N \sum_{i=1}^N x_i x_j \mathcal{P}_{ij} - \rho_{\text{mix}}^V \sum_{j=1}^N \sum_{i=1}^N y_i y_j \mathcal{P}_{ij} \right)^n, \quad (3.14)$$

with the cross-Parachor term defined as

$$\mathcal{P}_{ij} = (1 - \delta_{ij}) \frac{\mathcal{P}_i + \mathcal{P}_j}{2}, \quad (3.15)$$

where N is the number of components in the mixture, n is the exponent constant of the Parachor model ($n = 3.87$) [190, 234], ρ_{mix}^L and ρ_{mix}^V are the liquid and vapor-phase densities of the mixture in mol/cm³, and x_i and y_i denote the mole fractions of component i in the liquid and vapor phases, respectively, $\mathcal{P}_i$ is the Parachor number of component i , and δ_{ij} represents the binary interaction parameter of the Parachor model [190, 234]. The parameter $\mathcal{P}_i$ is obtained from the relation

$$\gamma^{1/n} = \mathcal{P}_i (\rho^L - \rho^V), \quad (3.16)$$

where ρ^L and ρ^V are the liquid and vapor densities of the pure fluid, expressed in mol/cm³, at the mixture temperature (T_{mix}), and γ is the pure fluid IFT computed using Eq. 3.13. For numerical stability, if $T_{\text{mix}} > 0.9 T_c$, $\mathcal{P}_i$ is computed at $0.9 T_c$, following the approach implemented in REFPROP v10 [190]. The calculation of IFTs using the Parachor model requires the EoS [234] to compute the necessary thermodynamic properties, including ρ^L , ρ^V , $\mathcal{P}_i$, x , and y . In this chapter, the GERG-2008 EoS [58] was used to obtain these properties.

3.2.4. Winterfeld–Scriven–Davis Model

The Winterfeld–Scriven–Davis (WSD) model presents an approximate theory for determining the planar interfacial tension of multicomponent, polyatomic fluids [146]. It extends the Fowler–Kirkwood–Buff framework by Kirkwood, Buff and Fowler [128, 262] by simplifying the rigorous statistical-mechanical expression for calculating IFT through the assumption of pairwise-additive intermolecular potentials [128, 146, 262]. The WSD model provides a practical approach for estimating IFTs of nonaqueous mixtures using IFTs of pure components and can be generalized to multicomponent systems as

$$\gamma_{\text{mix}} = \sum_{\alpha, \beta}^N \left(\frac{\rho_{\alpha}^L - \rho_{\alpha}^V}{\rho_{\alpha}^{0,L} - \rho_{\alpha}^{0,V}} \right)^2 \gamma_{\alpha}^0 \delta_{\alpha, \beta} + \Phi_{\alpha, \beta} \left[\left(\frac{\rho_{\alpha}^L - \rho_{\alpha}^V}{\rho_{\alpha}^{0,L} - \rho_{\alpha}^{0,V}} \right) \left(\frac{\rho_{\beta}^L - \rho_{\beta}^V}{\rho_{\beta}^{0,L} - \rho_{\beta}^{0,V}} \right) \right] \sqrt{\gamma_{\alpha}^0 \gamma_{\beta}^0}, \quad (3.17)$$

where γ_{mix} is the interfacial tension of the mixture, indices α and β labels for the chemical components in the mixture, and the summation in Eq. 3.17 runs over all component pairs ($\alpha, \beta = 1, 2, \dots, N$). The parameters γ_{α}^0 and γ_{β}^0 represent IFTs of pure components α and β , respectively. Similarly, $\rho_{\alpha}^{0,L}$, $\rho_{\alpha}^{0,V}$, $\rho_{\beta}^{0,L}$, and $\rho_{\beta}^{0,V}$ correspond to the equilibrium liquid and vapor molar densities in mol/cm³ of the pure components α and β , respectively. The parameters ρ_{α}^L , ρ_{α}^V , ρ_{β}^L , and ρ_{β}^V denote the liquid and vapor molar densities of α and β components in the mixture [146]. The term $\Phi_{\alpha, \beta}$ is the cross-interaction or mixing parameter that accounts for nonideal interactions between dissimilar molecules, and for nonaqueous, weakly polar fluids, $\Phi_{\alpha, \beta} \simeq 1$ [146]. The Kronecker delta $\delta_{\alpha, \beta}$ equals one when $\alpha = \beta$ (pure-component contribution) and zero otherwise [146].

The WSD model provides accurate predictions for nonaqueous mixtures and serves as a reliable semi-empirical framework in cases where molecular correlation data are unavailable [146, 247]. However, the formulation becomes less accurate when the mixture temperature exceeds the critical temperature of one or more components present in the mixture. For a binary system, Eq. 3.17 equals to

$$\gamma_{\text{mix}} = \left(\frac{\rho_1^L - \rho_1^V}{\rho_1^{0,L} - \rho_1^{0,V}} \right)^2 \gamma_1^0 + \left(\frac{\rho_2^L - \rho_2^V}{\rho_2^{0,L} - \rho_2^{0,V}} \right)^2 \gamma_2^0 + 2\Phi \left[\left(\frac{\rho_1^L - \rho_1^V}{\rho_1^{0,L} - \rho_1^{0,V}} \right) \left(\frac{\rho_2^L - \rho_2^V}{\rho_2^{0,L} - \rho_2^{0,V}} \right) \right] \sqrt{\gamma_1^0 \gamma_2^0}, \quad (3.18)$$

which further reduces to

$$\gamma_{\text{mix}} = \left(\frac{\rho_1^L - \rho_1^V}{\rho_1^{0,L} - \rho_1^{0,V}} \right)^2 \gamma_1^0, \quad (3.19)$$

when the mixture temperature is higher than the critical temperature of component 2. Eq. 3.19 implies that the supercritical component has no effect on IFT, which is physically unrealistic. In practice, the presence of a supercritical component is expected to decrease the IFT between the liquid and vapor phases of the subcritical component, due to the weakening of cohesive forces between unlike molecules, an effect analogous to that of a surfactant. To account for this behavior, an empirical correction term is introduced in this chapter, which should decrease IFT as the concentration of the supercritical component increases. The correction factor can be conveniently expressed in terms of the liquid-phase mole fraction of the supercritical component as $1 - x_2^L$, where x_2^L is the mole fraction of component 2 in the liquid phase, equivalently given by $x_2^L = \rho_2^L / (\rho_1^L + \rho_2^L)$. The corrected expression for the IFT of a binary mixture is given by

$$\gamma_{\text{mix}} = \left(\frac{\rho_1^L - \rho_1^V}{\rho_1^{0,L} - \rho_1^{0,V}} \right)^2 \gamma_1^0 \left(1 - \frac{\rho_2^L}{\rho_1^L + \rho_2^L} \right). \quad (3.20)$$

3.2.5. Classical Nucleation Theory

The Superheat Limit (SHL) and Supercool Limit (SCL) for CO₂ and binary CO₂ mixtures were evaluated using Classical Nucleation Theory (CNT), which relates the work required to form a critical bubble or droplet to the nucleation rate J [237]. The nucleation rate per unit volume and time is written in an Arrhenius form [237],

$$J = K \exp \left[-\frac{\Delta G^*}{k_B T} \right], \quad (3.21)$$

where ΔG^* is the free energy barrier at the critical radius r_c of the bubble or droplet, k_B is the Boltzmann constant, T is the absolute temperature, and K is the kinetic prefactor. Within the capillarity approximation of CNT, i.e., by assuming a planar IFT (γ) and bulk densities and chemical potentials inside the nucleus, ΔG^* is defined as,

$$\Delta G^* = \frac{4\pi \gamma r_c^2}{3}. \quad (3.22)$$

For superheated liquid (bubble nucleation), the nucleation driving force is the liquid tension, which is the pressure deficit of the liquid relative to equilibrium.

$$\Delta P = P^{\text{sat}}(T, z) - P_c > 0, \quad (3.23)$$

where $P^{\text{sat}}(T, z)$ is the saturation pressure of the fluid at temperature T and composition z , and P_c is a metastable liquid pressure at which J is evaluated. At the critical radius r_c of the bubble, the liquid tension exactly balances the Laplace pressure $2\gamma/r$, yielding the standard CNT expression for r_c ,

$$r_c^{\text{bubble}} = \frac{2\gamma}{P^{\text{sat}}(T, z) - P_c}. \quad (3.24)$$

For supersaturated vapor (droplet nucleation), the driving force is the bulk free energy density $\rho_n^L k_B T \ln(P_c/P^{\text{sat}}(T, z)) > 0$, where ρ_n^L is the liquid number density. Balancing the Laplace pressure and bulk free energy density terms at r_c leads to

$$r_c^{\text{droplet}} = \frac{2\gamma}{\rho_n^L k_B T \ln(P_c/P^{\text{sat}}(T, z))}. \quad (3.25)$$

For a detailed derivation, see Ref. [237]. The kinetic prefactor K for bubble formation in liquid and droplet formation in vapor is calculated by following the approximate expressions used by Aursand *et al.* [244] and Skarsvåg *et al.* [235],

$$K^{\text{bubble}} \approx \rho_n^L \sqrt{\frac{2\gamma}{\pi m}} \quad ; \quad K^{\text{droplet}} \approx \frac{(\rho_n^V)^2}{\rho_n^L} \sqrt{\frac{2\gamma}{\pi m}}, \quad (3.26)$$

where ρ_n^V is the vapor number density and m is the mass per molecule. For binary mixtures, we adopt the effective mass, m_{eff} instead of m in Eq. 3.26

$$\sqrt{1/m_{\text{eff}}^{\text{bubble}}} = \sum_{i=1}^{N_c} \frac{x_i^L}{\sqrt{m_i}} \quad ; \quad \sqrt{1/m_{\text{eff}}^{\text{droplet}}} = \sum_{i=1}^{N_c} \frac{y_i^V}{\sqrt{m_i}}, \quad (3.27)$$

where x_i^L is the liquid mole fraction and y_i^V is the vapor mole fraction of component i . N_c denotes the number of components in a mixture. The P_{SHL} and P_{SCL} limits are obtained by solving $J = J_{\text{cr}}$ for the metastable pressure P_c at a given T and z . The resulting P_c equals the P_{SHL} for bubbles in liquid and P_{SCL} for droplets in vapor. In this chapter, we adopt $J_{\text{cr}} = 10^{12} \text{ m}^{-3} \text{ s}^{-1}$, which is the same value of J_{cr} used by Aursand *et al.* [244] and Skarsvåg *et al.* [235] to estimate P_{SHL} and P_{SCL} of pure CO_2 and multicomponent CO_2 mixtures.

3.3. Results and discussion

In this section, the Vapor-Liquid Equilibrium (VLE) and interfacial properties of pure CO_2 and binary CO_2 mixtures computed from MD simulations, cDFT, EoS, and semi-empirical models are discussed. The computed phase equilibria from MD simulations and the PC-SAFT EoS were compared with the multiparameter Helmholtz EoS of Span and Wagner [28] for pure CO_2 , as implemented in REFPROP v10 [190] and with experimental data [91, 198,

200–202, 257, 258] for binary CO₂ mixtures. For binary CO₂ mixtures, IFTs computed from MD simulations and the PC-SAFT EoS + cDFT are presented combined with the phase equilibria (P_{xy} diagram) for three different temperatures. The P_{xy} diagrams and IFTs of CO₂ binary mixtures were computed from the PC-SAFT EoS + cDFT, for both $k_{ij} = 0$ and $k_{ij} \neq 0$. The k_{ij} values were determined by fitting the phase equilibria to experimental data [91, 198, 200–202, 257, 258] at each temperature and are therefore temperature dependent. The methodology used for obtaining k_{ij} and the experimental data used for fitting are presented in Sections S7 and S8 of the Supporting Information for this chapter [211]. The values of k_{ij} used in this chapter to compute the phase equilibria and interfacial properties are listed in Table 3.1. To obtain reliable statistics from MD simulations, 10 independent simulations with different initial configurations were performed for each temperature, and the results were analyzed using block averaging over 5 blocks to compute mean values and uncertainties of VLE and interfacial properties of pure CO₂. For binary systems, due to the computational costs, 2 independent simulations were performed at different initial configurations to compute the mean and uncertainty values of liquid mole fractions, vapor mole fractions, and interfacial properties at a given temperature and pressure. For clarity, the uncertainties in IFT and in the phase equilibria of the binary mixtures are omitted from the P_{xy} diagrams and are instead provided in Section S9 of the Supporting Information for this chapter [211]. To evaluate the accuracy of the computed bulk and interfacial properties of pure CO₂ and binary CO₂ mixtures, the Relative Percentage Deviation (RPD) and Maximum Absolute Relative Deviation (MARD) are used in addition to MAPD (Eq. 3.12). The RPD at a given state point i (temperature, pressure, or composition) is defined as

$$\text{RPD}_i = \frac{\chi_i^{\text{Comp.}} - \chi_i^{\text{Ref.}}}{\chi_i^{\text{Ref.}}} \times 100\%, \quad (3.28)$$

where $\chi_i^{\text{Comp.}}$ is the value obtained from MD simulations, EoS, or semi-empirical models, and $\chi_i^{\text{Ref.}}$ is the corresponding reference value (experiment, simulation, or semi-empirical model). The MARD is then

$$\text{MARD} = \max_i \left| \frac{\chi_i^{\text{Comp.}} - \chi_i^{\text{Ref.}}}{\chi_i^{\text{Ref.}}} \right| \times 100\%, \quad (3.29)$$

which represents the largest deviation across all state points, for example, IFTs of pure CO₂ over a full temperature range. The interfacial thickness of binary mixtures computed using MD simulations and the PC-SAFT EoS + cDFT are provided and discussed in Section S10 of the Supporting Information for this chapter [211], as their detailed presentation involves numerous figures. The numerical values of phase equilibria and interfacial properties of pure CO₂ and binary CO₂ mixtures computed from MD

simulations, PC-SAFT EoS + cDFT, and semi-empirical models are provided in Sections S6 and S11 of the Supporting Information for this chapter [211], respectively. The numerical values of SHL and SCL of pure CO₂ and binary CO₂ mixtures computed using CNT are presented in Section S12 of the Supporting Information for this chapter [211].

3.3.1. Pure CO₂

MD simulations were performed in the NVT ensemble using a two-phase system with a liquid slab between two vapor slabs to compute VLE and interfacial properties of pure CO₂, similar to the setup shown in Fig. 3.1. Density profiles $\rho(z)_{\text{MD}}$ of pure CO₂, averaged over 40 ns, are shown in Fig. 3.2a for 220–290 K in 10 K increments. Two system sizes were used: $L_z = 240$ Å with 2000 CO₂ molecules for 220–270 K, and $L_z = 320$ Å with 3000 CO₂ molecules for $T = 280$ and 290 K. Larger system sizes at $T = 280$ and 290 K were necessary to account for the significant density fluctuations noticed in MD simulations [79]. Coexistence densities were determined by fitting Eq. 3.6 to the $\rho(z)_{\text{MD}}$ profiles shown in Fig. 3.2a. The coexistence densities were also computed using the PC-SAFT EoS [68, 102] (computed using FeO_s[245]) and the Span and Wagner EoS [28] (computed using REFPROP v10 [190]). As shown in Fig. 3.2b, liquid and vapor densities from MD simulations and the PC-SAFT EoS were found to be in close agreement over the temperature range 220 K to 290 K, when compared to the Span and Wagner EoS [28]. Vapor densities predicted by the PC-SAFT EoS were underestimated up to 290 K and overestimated at higher temperatures, while liquid densities were underestimated up to 280 K and overestimated at higher temperatures. For the temperature range 220 K to 290 K, we observed maximum RPD values of approximately –5.5% at 290 K and –0.7% at 220 K for vapor densities, relative to the Span and Wagner EoS [28]. The MARD of the PC-SAFT EoS, relative to the Span–Wagner EoS [28], was ca. 0.3% for liquid densities and ca. 2.5% for vapor densities. Liquid densities were underestimated in MD simulations, with a maximum RPD of ca. 1% at 220 K, and were overestimated from 280 K to T_c . Vapor densities were underestimated up to 280 K and overestimated at higher temperatures, with a maximum RPD of ca. –6% at 220 K. The MARD was ca. 1% for liquid densities and ca. 3% for vapor densities. From the coexistence densities computed from MD simulations, T_c and ρ_c were calculated using the scaling law [263] and the law of rectilinear diameter [232] (see Section S4 of the Supporting Information for this chapter [211]). Compared to the Span and Wagner EoS, MD predicts a slightly lower value of T_c (ca. 0.2%), whereas the PC-SAFT EoS yields a higher estimate (ca. 1.7%), consistent with previous studies that showed the PC-SAFT EoS tends to overestimate critical properties [102, 264]. Fig. 3.2c presents the comparison of CO₂ density profiles at 220 K obtained from MD and PC-SAFT EoS + cDFT, and it also includes a representative snapshot of the MD simulation (top), showing the liquid

Table 3.1. Binary interaction parameters (k_{ij}) of CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ at different temperatures to compute phase equilibria (Pxy diagram) and interfacial properties from the PC-SAFT EoS [68, 102] combined with cDFT. The binary interaction parameters were obtained by fitting to experimental data reported in the literature [91, 198, 200–202, 257, 258]. Experimental data for the CO₂-CH₄ mixtures were taken from Wei *et al.* [200] at 230, 250, and 270 K. For the CO₂-Ar mixtures, the experimental data were taken from Lasala *et al.* [91] at 223.07 K and from Coquelet *et al.* [198] at 253.28 and 273.26 K. Experimental data for the CO₂-N₂ mixtures at 220, 250, and 273.12 K were obtained from Brown *et al.* [202], Al-Sahhaf, Kidnay and Sloan [258], and Lasala *et al.* [91], respectively. Experimental data for the CO₂-H₂ mixtures were taken from Tsang and Street [201] at 220 and 250 K and from Fandiño, Trusler and Vega-Maza [257] at 273.1 K.

Binary mixture	T / [K]	k_{ij} / [-]
CO ₂ -CH ₄	230.0	0.0519
	250.0	0.0541
	270.0	0.0566
CO ₂ -Ar	223.07	0.0317
	253.28	0.0295
	273.26	0.0240
CO ₂ -N ₂	220.0	-0.0164
	250.0	-0.0294
	273.12	-0.0699
CO ₂ -H ₂	220.0	-0.0594
	250.0	0.0474
	273.1	0.0538

phase in equilibrium with its vapor phases. To enable direct comparison with the density profiles from the PC-SAFT EoS + cDFT, $\rho(z)_{\text{MD}}$ (computed with $L_z = 240 \text{ \AA}$, i.e., $120 \text{ \AA}$ per interface) was scaled to a reference length of $110 \text{ \AA}$ so that both methods could be compared on the same spatial basis. Both approaches agree well with each other in describing the bulk liquid, vapor, and interfacial regions, while the metastable region provides a smooth transition between the two phases. Multiparameter EoS models, such as the Span and Wagner [28] and GERG-2008 [58], are known to produce multiple artificial Maxwell loops [60], which hinder a smooth phase transition. For this reason, the PC-SAFT EoS [68, 102] was selected for estimating interfacial properties of CO₂ and binary CO₂ mixtures.

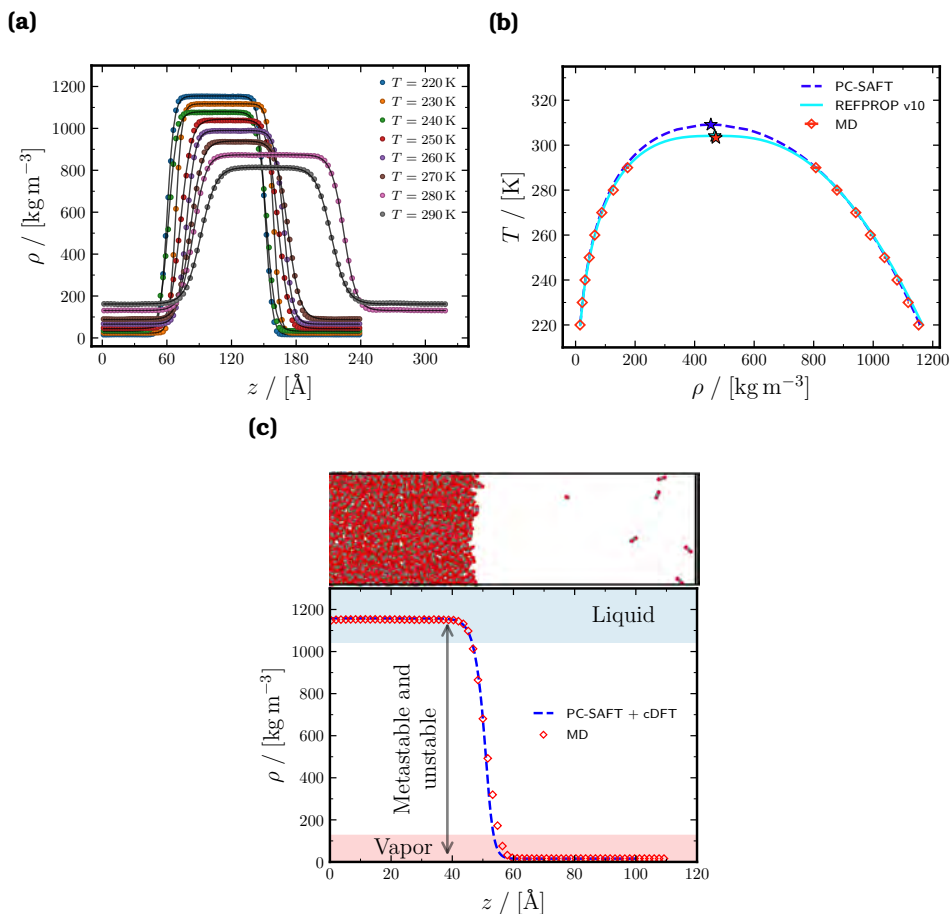


Figure 3.2. (a) Density profiles of pure CO_2 obtained from MD simulations, $\rho(z)_{\text{MD}}$, at temperatures 220 K to 290 K in 10 K increments. (b) Comparison of CO_2 VLE densities computed from MD simulations (red symbols) with the PC-SAFT EoS + cDFT [68, 102] (blue line) and REFPROP v10 (Span and Wagner EoS [28, 190]) (cyan line). Star symbols denote the critical temperatures T_c estimated from MD simulations, the PC-SAFT EoS, and REFPROP v10. (c) Snapshot of the MD simulation box representing the coexistence of liquid and vapor phases (top) and comparison of the CO_2 density profile at $T = 220 \text{ K}$ obtained from MD simulations with the PC-SAFT EoS + cDFT [68, 102]. The liquid, vapor, metastable, and unstable density regions were determined using the definition of L_{10}^{90} (Eq. 3.7). The visualization was rendered using the Open Visualization Tool (OVITO) software [265].

Fig. 3.3a shows IFTs of pure CO₂ as a function of temperature computed from MD simulations and the PC-SAFT EoS + cDFT and their comparison with experimental measurements of Quinn [266] and Jasper [267], REFPROP v10 [190] (Guggenheim–Katayama equation [219]), and MD simulations of Stephan [127]. The coefficients required to compute IFTs using the Guggenheim–Katayama equation [219] are fitted with 125 experimental values by Cachadiña, Mulero and Tian [246] which also includes the experimental measurements of Quinn [266] and Jasper [267]. The fitted coefficients suggested by Cachadiña, Mulero and Tian [246], which is used in REFPROP v10 [190], are reported to have a maximum RPD of 11.7% at 280.15 K. Using the Guggenheim–Katayama equation (Eq. 3.13) [219] with the coefficients of Cachadiña, Mulero and Tian [246] as reference, the IFTs computed from MD simulations with the TraPPE force field [82] were overestimated, consistent with findings of Stephan [127] where the force field of Merker *et al.* [268] was used for IFTs computation. The PC-SAFT EoS + cDFT underestimates IFTs up to 274 K and tends to overestimate at higher temperatures ($T > 274$ K). Fig. 3.3b shows the RPD of MD and the PC-SAFT EoS + cDFT computed using Eq. 3.28. The maximum uncertainty reported by Cachadiña, Mulero and Tian [246] was taken as the uncertainty range for comparison. As shown in Fig. 3.3b, IFTs computed from MD simulations with the TraPPE force field [82] remain within the uncertainty range except at 290 K, where large density fluctuations cause deviations larger than 11.7%. Similarly, the RPD of the PC-SAFT EoS + cDFT lies within the uncertainty range for temperatures less than 290 K, and diverges at temperatures larger than 290 K. The RPD values from MD simulations of Stephan [127] are not included in Fig. 3.3b since all points fall outside the uncertainty range. We computed the interfacial thickness of pure CO₂ from MD simulations with the TraPPE force field [82] and from the PC-SAFT EoS + cDFT using the definition of L_{10}^{90} [254] (Eq. 3.7), and compared the results with values reported by Stephan [127] (Merker *et al.* [268] force field), as shown in Fig. 3.3c. Stephan [127] computed the interfacial thickness from ensemble-averaged density profiles of MD simulations using the same L_{10}^{90} definition [254]. In this chapter, we applied two approaches: (1) using Eq. 3.6 to compute L_{10}^{90} ($L_{10}^{90} \approx 2.19D$ [249]), and (2) sampling L_{10}^{90} every 500 fs (i.e., every 1000 time steps), since we observed the lateral movement of the liquid slab. The results from both approaches are shown in Fig. 3.3c, with Eq. 3.6 represented by red symbols and the L_{10}^{90} values sampled every 1000 time steps represented by green symbols. In all cases, the interfacial thickness increases with temperature and diverges at T_c . The values obtained from Eq. 3.6 and those reported by Stephan [127] overestimate L_{10}^{90} at all temperatures, whereas the values sampled every 500 fs agree well with the PC-SAFT EoS + cDFT, except at 290 K. We did not report RPD and MARD, since no reference experimental data or model to compute the L_{10}^{90} for pure CO₂ were available in literature. Using the same simulation setup as for the computation of coexistence densities, IFTs, and

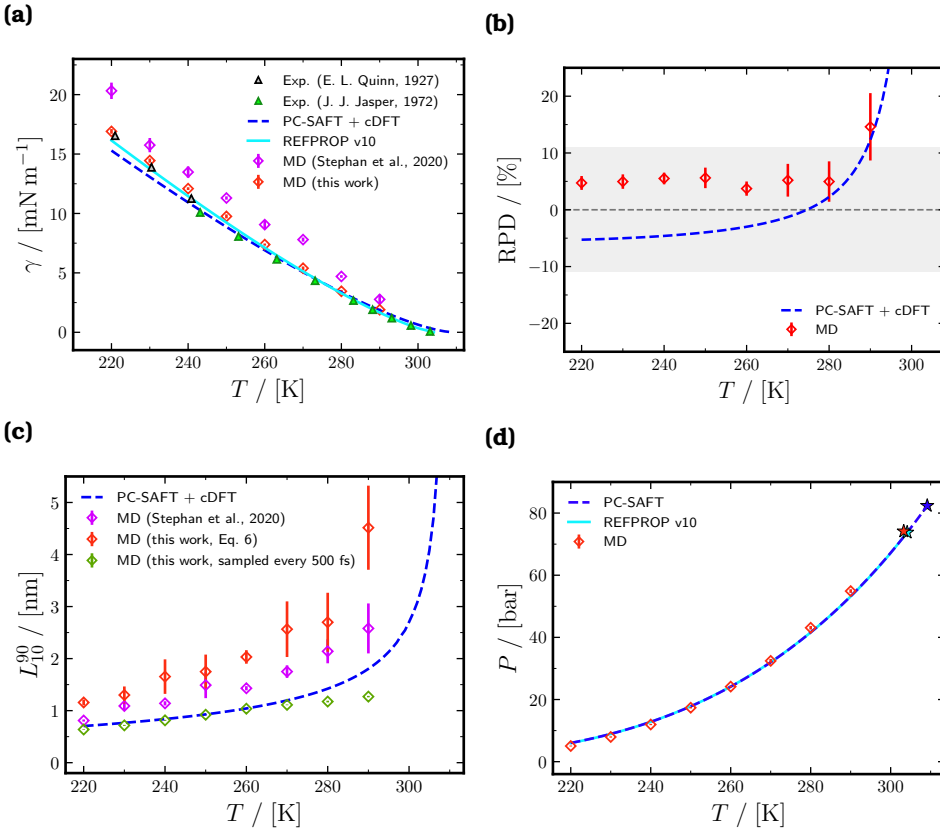


Figure 3.3. (a) Interfacial Tension (IFT) of CO₂ obtained from MD simulations and the PC-SAFT EoS + cDFT [68, 102] as a function of temperature and comparison with REFPROP v10 [190], MD simulations of Stephan, Fleckenstein and Hasse [269] using the force field of Merker *et al.* [268] and experimental measurements of Quinn [266] and Jasper [267]. (b) Relative Percentage Deviation (RPD) of IFTs computed from MD simulations and the PC-SAFT EoS [68, 102]. The uncertainty region of 11.7% reported by Cachadiña, Mulero and Tian [246] is shown as a gray band. (c) Comparison of interfacial thickness L_{10}^{90} of pure CO₂ computed from MD simulations and the PC-SAFT EoS + cDFT [68, 102] with MD simulations of Stephan, Fleckenstein and Hasse [269]. In this chapter, L_{10}^{90} of pure CO₂ were computed using two approaches: (1) using Eq. 3.6 and (2) by sampling the density profiles $\rho(z)_{\text{MD}}$ every 500 fs. (d) Comparison of CO₂ saturation pressures P^{sat} computed from MD simulations (red symbols) and PC-SAFT EoS [68, 102] (cyan line) with REFPROP v10 [190] (blue line). Star symbols denote the critical pressure P_c estimated by MD using the Antoine equation [270], PC-SAFT EoS [68, 102], and REFPROP v10 [190].

interfacial thickness of pure CO₂, we computed the saturation pressures P_{sat} as a function of temperature. Fig. 3.3d compares P_{sat} obtained from MD simulations and the PC-SAFT EoS with the Span and Wagner EoS [28]. MD simulations underestimated P_{sat} below 260 K and overestimated it above this temperature, yielding a maximum RPD of ca. 16% at 220 K and a MARD of ca. 6%. The large RPD at 220 K arises from relatively low P_{sat} at this low temperature. Conversely, the PC-SAFT EoS overpredicted up to 234 K and underpredicted above that temperature, with a maximum RPD of ca. 0.5% at 220 K and a MARD of ca. 0.2% in the range 220 K to 290 K. The saturated vapor pressure at T_c was estimated by fitting the $P_{\text{sat}}(T)$ data from the MD simulations with the Antoine equation [270] and extrapolating the fit to T_c (see Section S4 of the Supporting Information for this chapter [211]). The saturated vapor pressure at T_c computed from both MD simulations and the PC-SAFT EoS were found to be overestimated by ca. 0.4% and ca. 12%, respectively, when compared to the Span and Wagner EoS [28].

3.3.2. Binary CO₂ mixtures

The phase equilibria (Pxy diagrams) and IFTs for four binary mixtures – CO₂–CH₄, CO₂–Ar, CO₂–N₂, and CO₂–H₂ computed from MD simulations and from the PC-SAFT EoS combined with cDFT using both $k_{ij} = 0$ and $k_{ij} \neq 0$ at different temperatures are shown in comparison with experimental Pxy data [91, 198, 200–202, 257, 258] in Figs. 3.4, 3.5, 3.6, and 3.7. MD simulations at low-pressure limit of the phase envelope (i.e., at high vapor mole fraction) were not feasible because the vapor phase became too dilute to maintain a stable vapor-liquid interface within a single simulation box. Hence, MD simulations were considered at pressures where stable two-phase equilibrium could be maintained for computing Pxy diagrams and IFTs. The experimental Pxy data reported in the literature for binary mixtures considered in this chapter were measured at different temperatures, namely 230 K for CO₂–CH₄, 223.07 K for CO₂–Ar, and 220 K for CO₂–N₂. Hence, IFTs and Pxy diagrams of binary mixtures were computed using MD simulations and the PC-SAFT EoS + cDFT, at the experimentally available temperatures as shown in Table 3.1. The k_{ij} s of binary mixtures obtained by fitting the experimental Pxy data [91, 198, 200–202, 257, 258] at different temperatures are also presented in Table 3.1. It is clear from all the Pxy diagrams (Figs. 3.4, 3.5, 3.6, and 3.7) that the phase equilibria computed from the PC-SAFT EoS with $k_{ij} \neq 0$ has a good agreement with experimental measurements [91, 198, 200–202, 257, 258] compared to the PC-SAFT EoS with $k_{ij} = 0$. To quantify the deviations in liquid and vapor mole fractions predicted by the PC-SAFT EoS relative to experimental data, the mean absolute percentage deviation defined in Eq. 3.12 (MAPD) is used. Similarly, MAPD is used to compare the liquid and vapor mole fractions and IFTs predicted by the PC-SAFT EoS relative to those obtained from MD simulations. IFTs computed using the PC-SAFT EoS + cDFT

with $k_{ij} \neq 0$ are underestimated for all binary mixtures compared to MD simulations. The MAPD is an appropriate measure for quantifying these deviations than RPD. The analysis of Pxy data between MD simulations and experimental measurements was not quantified because the experimental data points were not available exactly at the same discrete pressure values used in the simulations. The experimental IFTs of $\text{CO}_2\text{-CH}_4$ mixture at 233 K of Schenk *et al.* [236] were used to compare IFTs computed from MD and the PC-SAFT EoS + cDFT at 230 K. Except for the experimental measurements reported by Schenk *et al.* [236], experimental data on IFTs of binary mixtures discussed in literature are very limited, to the best of our knowledge. This scarcity has also been highlighted by Skarsvåg *et al.* [235] In addition to the Pxy diagrams, we also computed the density profiles of $\text{CO}_2\text{-CH}_4$, $\text{CO}_2\text{-Ar}$, $\text{CO}_2\text{-N}_2$, and $\text{CO}_2\text{-H}_2$ from MD simulations, and the PC-SAFT EoS + cDFT with $k_{ij} = 0$ and $k_{ij} \neq 0$. We compared the density profiles for a single thermodynamic state point (at a particular temperature and pressure) for each binary mixture. For binary mixtures, the enrichment of the Light Boiling Component (LBC), such as CH_4 , Ar, N_2 , and H_2 , was noticed in both MD simulations and the PC-SAFT EoS + cDFT. The enrichment of the LBC at the interface was evaluated using the definition [271]

$$E_i = \frac{\max(\rho_i(z))}{\max(\rho_i^L, \rho_i^V)}, \quad (3.30)$$

where, $\rho_i(z)$, ρ_i^L , and ρ_i^V denote the density profile, liquid density, and vapor density of component i , respectively. Eq. 3.30 provides the ratio between the maximum interfacial density of a component i and its corresponding maximum bulk phase densities [271]. By definition, $E_i > 1$ indicates interfacial enrichment, whereas $E_i = 1$ signifies presence of no interfacial enrichment. For all the binary mixtures investigated in this chapter, E_1 (E_{CO_2}) is equal to 1, indicating that CO_2 does not accumulate at the interface. In contrast, E_2 represents the enrichment of the LBC, and values greater than 1 indicate that the LBC becomes more concentrated at the interface than in either bulk phases. Enrichment is strongest at lower pressures with a high liquid mole fraction and decreases as pressure increases [271]. In this chapter, we present the density profiles in Figs. 3.4d, 3.5d, 3.6d, and 3.7d at relatively low pressures compared to the critical pressure of the mixture. Density profiles are shown at 250 K for $\text{CO}_2\text{-CH}_4$, $\text{CO}_2\text{-N}_2$, and $\text{CO}_2\text{-H}_2$, and at 253.28 K for $\text{CO}_2\text{-Ar}$, enabling a consistent assessment of enrichment across different LBC within the same temperature range. The bulk liquid and vapor densities of the binary mixtures studied in this chapter are not reported. Readers are referred to Chapter 2 for details and discussion on the computation of bulk densities using Monte Carlo [124, 125, 167] simulations in the Gibbs ensemble [120, 191], rather than by simulating two coexisting phases within a single simulation box.

CO₂-CH₄

The Pxy diagrams of CO₂-CH₄ mixtures at 230, 250, and 270 K are shown in Fig. 3.4 and compared with the experimental data of Wei *et al.* [200]. The liquid mole fractions predicted by the PC-SAFT EoS with $k_{ij} \neq 0$ at 230, 250, and 270 K show good agreement with the experimental data of Wei *et al.* [200], yielding a MAPD below 5%, excluding the points near the critical pressures (i.e., $P < 65$, 75, and 80 bar for 230, 250, and 270 K, respectively). The critical pressures predicted by the PC-SAFT EoS are generally overestimated compared to experimental values [68, 102]; therefore, data near the critical region were excluded from the analysis. Similarly, the liquid mole fractions obtained using the PC-SAFT EoS with $k_{ij} \neq 0$ maintain a MAPD of less than 5% across the considered pressure range. When comparing the MD simulations with the PC-SAFT EoS with $k_{ij} \neq 0$ as a reference state point, the MAPD of liquid mole fractions and vapor mole fractions were ca. 2% for 230, 250, and 270 K. The MAPD of vapor mole fraction was also found to be ca. 5% for 250, and 270 K, whereas at 230 K it was ca. 7%. Along with the Pxy diagrams of CO₂-CH₄ presented in Fig. 3.4, the IFT values computed using the PC-SAFT EoS + cDFT, with $k_{ij} = 0$ and $k_{ij} \neq 0$, are compared with the results from MD simulations. The results clearly show that the IFT decreases with increasing pressure and decreasing liquid mole fraction for 230, 250, and 270 K. The PC-SAFT EoS combined with cDFT and $k_{ij} \neq 0$ generally underpredicted the IFTs, except at 270 K and 65 bar. For all temperatures, the MAPD remained below 14%. Data points near the critical pressure were excluded from the analysis because evaluating the MAPD at very low IFT values ($< 1 \text{ mN m}^{-1}$) leads to disproportionately large errors. In addition to the comparison of IFTs computed from MD simulations and the PC-SAFT EoS combined with cDFT, the experimental measurements of Schenk *et al.* [236] at 233 K are shown in Fig. 3.4a. IFTs predicted by the MD simulations show good agreement with the experimental data reported by Schenk *et al.* [236]. A slight underestimation is observed, which we attribute to the temperature difference between the IFTs measured at 233 K and those computed at 230 K.

We computed the density profiles of the CO₂-CH₄ mixture at 250 K and 50 bar using MD simulations and the PC-SAFT EoS coupled with cDFT, and present them in Fig. 3.4d. A comparison of these density profiles shows that the PC-SAFT EoS + cDFT with $k_{ij} = 0.0541$ provides better agreement with the MD simulation than with $k_{ij} = 0$. The density profile from the MD simulation also accurately reproduces the interfacial thickness, which was averaged over 60 ns. The plateau regions in the MD density profile, corresponding to the bulk liquid and vapor phases, show good agreement with the densities predicted by the PC-SAFT EoS. However, the liquid density of CO₂ obtained from the MD simulation is underestimated by a RPD of ca. 2% compared to the PC-SAFT EoS + cDFT results with $k_{ij} = 0.0541$. A slight interfacial enrichment ($E_{\text{CH}_4} \approx 1.1$) is observed in

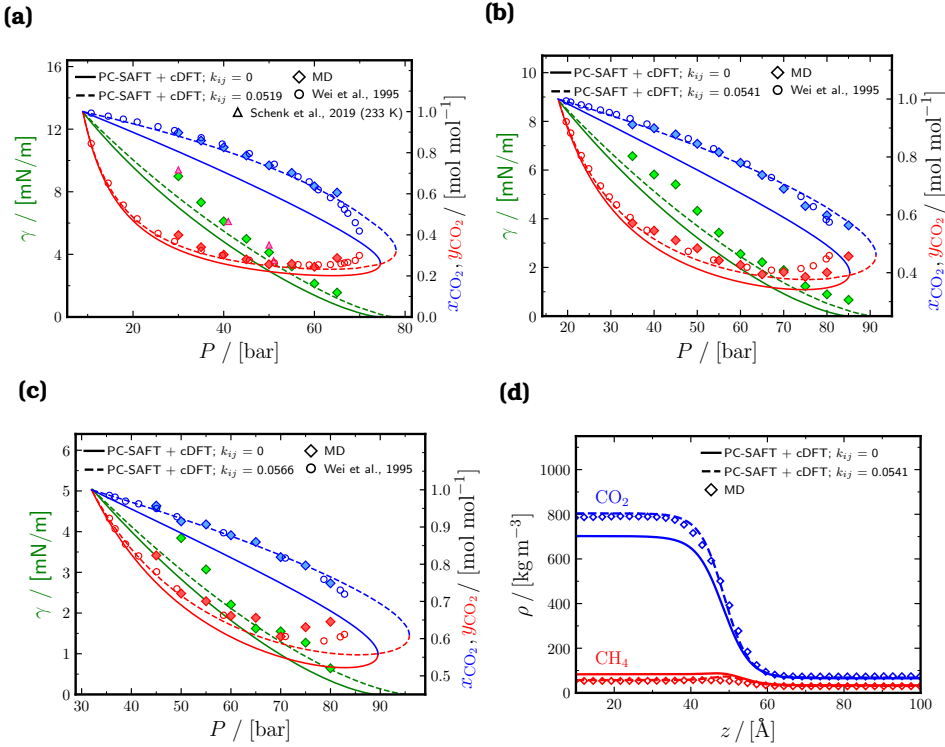


Figure 3.4. Pxy diagrams and IFTs of CO₂-CH₄ mixtures at (a) 230 K, (b) 250 K, and (c) 270 K computed using the PC-SAFT EoS [68, 102] combined with cDFT, with $k_{ij} = 0$ and $k_{ij} \neq 0$, compared with experimental data [200, 236] and MD simulations. In subfigures (a)-(c), the left vertical-axis (green) represents the interfacial tension γ , while the right vertical-axis (red and blue) corresponds to the vapor and liquid mole fractions of CO₂, respectively. (d) Comparison of density profiles at 250 K and 50 bar, computed using the PC-SAFT EoS [68, 102] combined with cDFT for $k_{ij} = 0$ and $k_{ij} \neq 0$, and from the MD simulation.

the density profiles obtained from the MD simulation. An enrichment of $E_{\text{CH}_4} \approx 1.3$ is observed in the PC-SAFT EoS combined with cDFT using $k_{ij} = 0.0541$ at 250 K and 50 bar.

CO₂-Ar

The Pxy diagrams of CO₂-Ar mixtures at 223.07, 253.28, and 273.26 K, computed from the PC-SAFT EoS and MD simulations, are shown in Fig. 3.5. The computed Pxy diagrams were compared with the experimental data of Lasala *et al.* [91] for 223.07 K and Coquelet *et al.* [198] for 253.28 K and 273.26 K. The Pxy diagrams computed from the PC-SAFT EoS with

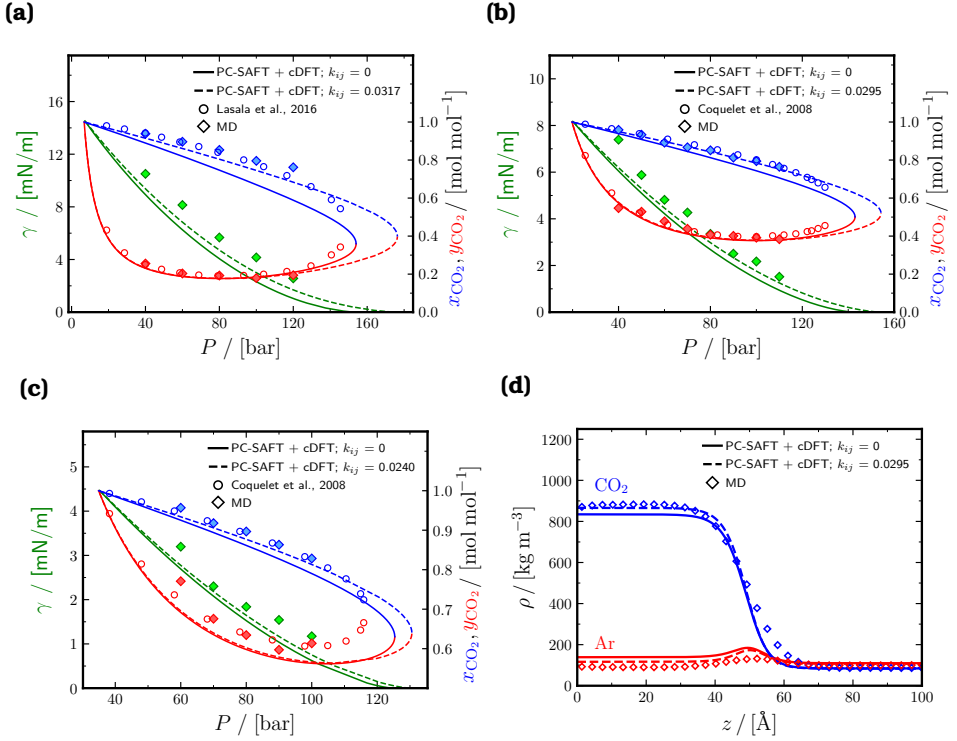


Figure 3.5. Pxy diagrams and IFTs of CO₂-Ar mixtures at (a) 223.07 K, (b) 253.28 K, and (c) 273.26 K computed using the PC-SAFT EoS [68, 102] combined with cDFT, with $k_{ij} = 0$ and $k_{ij} \neq 0$, compared with experimental data [91, 198] and MD simulations. In subfigures (a)-(c), the left vertical-axis (green) represents the interfacial tension γ , while the right vertical-axis (red and blue) corresponds to the vapor and liquid mole fractions of CO₂, respectively. (d) Comparison of density profiles at 250 K and 70 bar, computed using the PC-SAFT EoS [68, 102] combined with cDFT for $k_{ij} = 0$ and $k_{ij} \neq 0$, and from the MD simulation.

fitted k_{ij} agreed well with the experimental Pxy data [91, 198] at all three temperatures. The liquid and vapor mole fractions estimated by the PC-SAFT EoS with fitted k_{ij} exhibited MAPDs of less than 2% and 5%, respectively, for both 253.28 and 273.26 K. However, the MAPDs were slightly higher at 223.07 K, with ca. 3% for the liquid mole fractions and 6% for the vapor mole fractions. This analysis excluded data points close to the critical pressure, corresponding to P values below 120, 110, and 100 bar for 223.07, 253.28, and 273.26 K, respectively, to ensure consistency across conditions. A similar comparison was also performed for the Pxy diagrams and IFTs computed from MD simulations. The liquid

mole fractions computed from MD simulations agreed closely with the experimental measurements of Coquelet *et al.* [198] and Lasala *et al.* [91], consistent with the results obtained from the PC-SAFT EoS with fitted k_{ij} . The MAPDs of liquid and vapor mole fractions computed from the PC-SAFT EoS with $k_{ij} \neq 0$ were 2% and 6%, respectively, at 253.28 and 273.26 K compared to MD simulations, whereas at 223.07 K, the MAPDs were ca. 6% and 7% for the liquid and vapor mole fractions, respectively. The higher MAPDs of the PC-SAFT EoS with $k_{ij} \neq 0$ at 223.07 K were also reflected in the estimation of IFTs, where the MAPD was close to 18%, whereas at 253.28 and 273.26 K, the MAPDs were ca. 12% and 14%, respectively.

We computed the density profile of the CO₂–Ar mixture at 253.28 K and 70 bar using MD simulations and the PC-SAFT EoS with cDFT. The results are shown in Fig. 3.5d. From the comparison of the density profiles, it is clear that the density profiles computed from the PC-SAFT EoS + cDFT with $k_{ij} = 0.0295$ and $k_{ij} = 0$ have a better agreement with the MD simulation. The density profile computed from MD simulation (averaged for 60 ns), shown in Fig. 3.4d overestimated the interfacial thickness compared to the PC-SAFT EoS + cDFT with $k_{ij} = 0.0295$ and $k_{ij} = 0$. The MD simulation accurately predicted the CO₂ liquid and vapor densities and Ar vapor density, whereas the Ar liquid density was underpredicted by ca. 7% compared to the PC-SAFT EoS + cDFT with $k_{ij} = 0.0295$. At the interface, a slight enrichment of Ar is noticed in Fig. 3.5d in both MD simulation and the PC-SAFT EoS + cDFT with $k_{ij} = 0.0295$ and $k_{ij} = 0$. The MD simulations and the PC-SAFT EoS + cDFT yielded E_{Ar} values of approximately 1.3 and 1.5, respectively.

CO₂–N₂

The P_{xy} diagrams of CO₂–N₂ mixtures at 220, 250, and 273.12 K, computed using the PC-SAFT EoS and MD simulations, are compared with the experimental data of Brown *et al.* [202], Al-Sahhaf, Kidnay and Sloan [258], and Lasala *et al.* [91] in Fig. 3.6. Compared to CO₂–CH₄ and CO₂–Ar, CO₂–N₂ binary interaction parameters are negative, which means the liquid and vapor mole fractions, IFTs and density profiles shown in Fig. 3.6 are overestimated compared to $k_{ij} = 0$. To quantify the deviations between the experimental measurements [91, 202, 258] and the PC-SAFT EoS with $k_{ij} \neq 0$, as well as between the MD simulations and the PC-SAFT EoS with $k_{ij} \neq 0$, we excluded data points near the critical region. This analysis was carried out to avoid disproportionate errors caused by the overestimation of the critical pressure by the PC-SAFT EoS [68, 102] and from IFTs values lower than 1 mNm^{–1}. The excluded pressure ranges correspond to values below 160, 150, and 100 bar for 220, 250, and 273.12 K, respectively. The MAPD of liquid and vapor mole fractions computed from the PC-SAFT EoS with $k_{ij} \neq 0$ were found to be ca. 1% and 2%, respectively, compared to experimental measurements [91, 202, 258] for all temperatures. Comparing the liquid and vapor mole

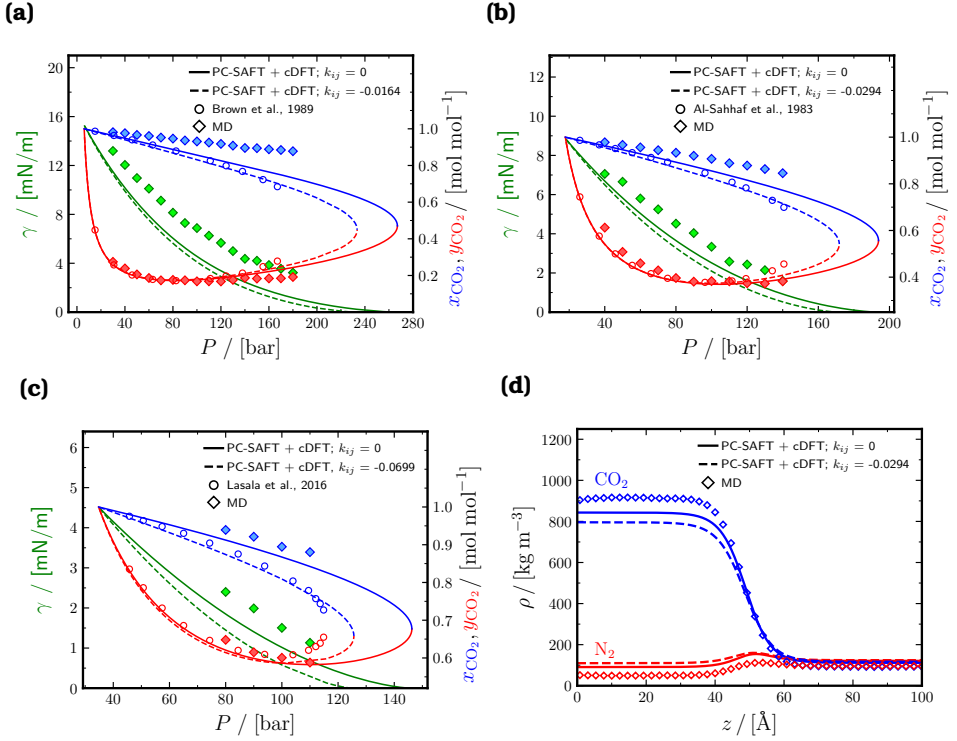


Figure 3.6. Pxy diagrams and IFTs of CO₂-N₂ mixtures at (a) 220 K, (b) 250 K, and (c) 273.12 K computed using the PC-SAFT EoS [68, 102] combined with cDFT, with $k_{ij} = 0$ and $k_{ij} \neq 0$, compared with experimental data [91, 202, 258] and MD simulations. In subfigures (a)-(c), the left vertical-axis (green) represents the interfacial tension γ , while the right vertical-axis (red and blue) corresponds to the vapor and liquid mole fractions of CO₂, respectively. (d) Comparison of density profiles at 250 K and 90 bar, computed using the PC-SAFT EoS [68, 102] combined with cDFT for $k_{ij} = 0$ and $k_{ij} \neq 0$, and from the MD simulation.

fractions between MD simulations and the PC-SAFT EoS with $k_{ij} \neq 0$, a large MAPD up to ca. 9% and 8%, respectively, were observed for all the temperatures. Large deviation in the liquid mole fraction computed from MD simulations compared to experimental measurements [91, 202, 258] and the PC-SAFT EoS observed in Fig. 3.6 is due to exclusion of the binary interaction parameter ($k_{ij}^{\text{MD}} = 0$) in MD simulations for CO₂-N₂ mixture at 220, 250, and 273.12 K. Due to the use of $k_{ij}^{\text{MD}} = 0$, the IFTs computed from MD simulations were overestimated, with MAPDs of ca. 36%, 35%, and 50% at 220, 250, and 273.12 K, respectively. The overprediction of both liquid mole fractions and IFTs clearly indicates that

MD simulations deviate significantly from the PC-SAFT EoS with $k_{ij} \neq 0$. To achieve accurate predictions of liquid mole fractions and IFTs, the appropriate k_{ij}^{MD} values should be determined by constructing a k_{ij}^{MD} matrix and validating the resulting thermodynamic properties with experimental data. The inclusion of k_{ij}^{MD} is particularly important for the CO₂–N₂ mixture because the combination of force fields used for CO₂ [82] and N₂ [113] does not accurately capture the unlike interactions when applying the LB mixing rules. It should also be noted that k_{ij}^{MD} differs from k_{ij}^{EoS} , and therefore, the same value cannot be assumed for both. In this chapter, no k_{ij}^{MD} corrections were applied due to the high computational cost associated with their determination. Assuming $k_{ij}^{\text{MD}} = 0$ is reasonable for the other mixtures examined in this chapter, however, for the CO₂–N₂ mixture, it results in a noticeable overprediction of IFTs.

We computed the density profile of the CO₂–N₂ mixture at 250 K and 90 bar using MD simulations and the PC-SAFT EoS coupled with cDFT, and present the results in Fig. 3.6d. The density profiles obtained from the PC-SAFT EoS + cDFT with $k_{ij} = -0.0294$ show good agreement with the MD simulation, except for an overprediction of the liquid-phase densities by ca. 11% for CO₂ and 35% for N₂. This overprediction can be attributed to the use of $k_{ij}^{\text{MD}} = 0$. Since the MD simulations overpredict the bulk densities, the resulting density difference between the two phases, which is an important factor governing the IFT [233, 262], is also overestimated. This contributes directly to the overprediction of IFTs noticed in Fig. 3.6. The density profile obtained from MD simulations, averaged over 80 ns, was found to accurately reproduce the interfacial thickness when compared with the results from the PC-SAFT EoS + cDFT using $k_{ij} = 0$ and $k_{ij} = -0.0294$. A minor enrichment of N₂ is observed at the interface, with E_{N_2} values of ca. 1.05% and 1.2% predicted by the MD simulation and the PC-SAFT EoS + cDFT with $k_{ij} = -0.0294$, respectively.

CO₂–H₂

The Pxy diagrams of the CO₂–H₂ mixture at 220, 250, and 273.1 K, computed using the PC-SAFT EoS and MD simulations, are shown in Fig. 3.7 along with the experimental data from Tsang and Street [201] and Fandiño, Trusler and Vega-Maza [257]. Experimental data [201, 257] were limited to 354, 884, and 148.67 bar at 220, 250, and 273.1 K, respectively. MD simulations were performed up to 400 bar, but comparisons of liquid and vapor mole fractions and IFTs were restricted to pressures below the experimental limits. At 220 K, the limited data yielded a negative value of k_{ij} , leading to poor predictions of Pxy diagrams and IFTs relative to both experiments and the PC-SAFT EoS results with $k_{ij} = 0$. In sharp contrast, MD simulations matched experimental Pxy data well, although mole fractions could not be directly compared due to differing pressure points. At 250 K, the PC-SAFT EoS with $k_{ij} \neq 0$ showed good agreement

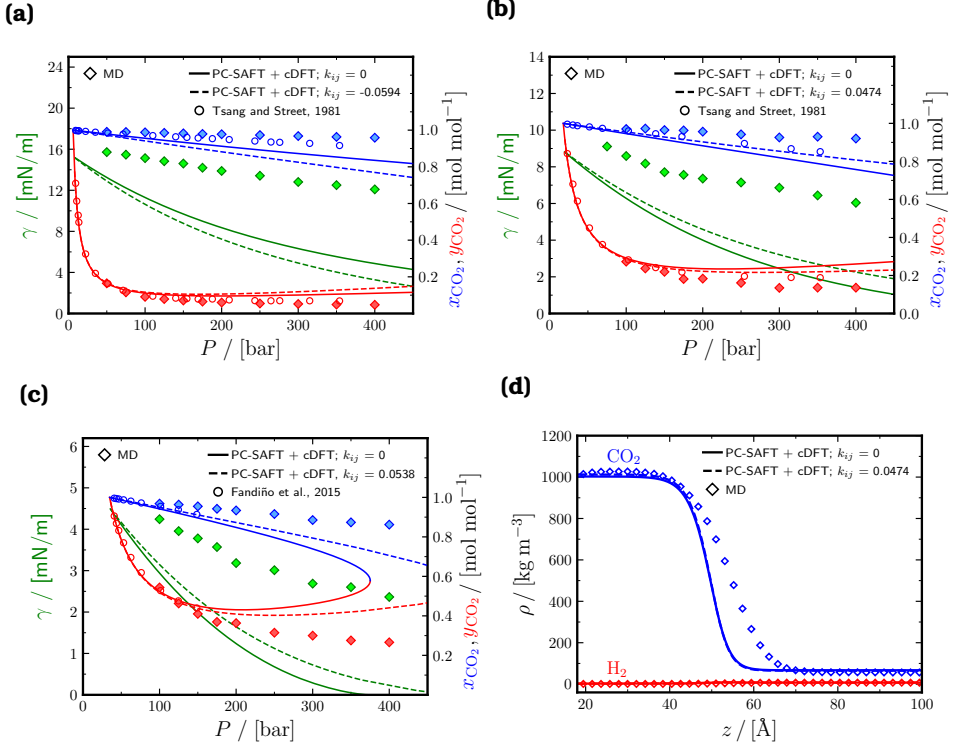


Figure 3.7. Pxy diagrams and IFTs of CO₂-H₂ mixtures at (a) 220 K, (b) 250 K, and (c) 273.1 K computed using the PC-SAFT EoS [68, 102] combined with cDFT, with $k_{ij} = 0$ and $k_{ij} \neq 0$, compared with experimental data [91, 201] and MD simulations. In subfigures (a)-(c), the left vertical-axis (green) represents the interfacial tension γ , while the right vertical-axis (red and blue) corresponds to the vapor and liquid mole fractions of CO₂, respectively. (d) Comparison of density profiles at 250 K and 100 bar, computed using the PC-SAFT EoS [68, 102] combined with cDFT for $k_{ij} = 0$ and $k_{ij} \neq 0$, and from the MD simulation.

with both experimental and MD results, yielding liquid and vapor mole fractions with MAPDs below 1% and 7%, respectively, compared to the experimental data[201]. MD simulations overpredicted the liquid and vapor mole fractions by 6% and 32%, respectively, compared to the PC-SAFT EoS with $k_{ij} \neq 0$, showing RPDs exceeding 50% for y_{CO_2} values lower than 0.2 at pressures above 250 bar. At 273.1 K, the PC-SAFT EoS with $k_{ij} \neq 0$ accurately reproduced the experimental data, with MAPDs of 0.3% and 2% for the liquid and vapor mole fractions, respectively, whereas the MD simulations slightly overpredicted these values by 3% and 4%. However, at 273.1 K and pressures above 150 bar, the Pxy diagram changes

noticeably when $k_{ij} = 0.0538$ compared to the results with $k_{ij} = 0$. This deviation arises from the limited experimental data [257], which impacts the reliability of k_{ij} and, consequently, the predicted phase equilibria and IFTs. Although the experimental liquid and vapor mole fractions agree reasonably well with the PC-SAFT EoS + cDFT and MD simulations at 220 and 250 K (Figs. 3.7a and 3.7b), IFTs estimated by the two methods differ substantially, with MAPDs of ca. 40% at 220 K and 33% at 250 K. This suggests that the observed deviations in IFTs may arise either from limitations in the PC-SAFT EoS + cDFT model or from the force field used for H₂, even though the liquid and vapor mole fractions were predicted accurately. The discrepancy becomes more pronounced at 273.1 K, likely due to the combined influence of uncertainties in the IFT predictions (from both the PC-SAFT EoS + cDFT and the force field) and the k_{ij} values.

The density profile of CO₂-H₂ mixture at 250 K and 100 bar computed from MD simulation and the PC-SAFT EoS combined with cDFT are shown in Fig. 3.7d. From the density profiles comparison, it is clear that the density profiles computed from the PC-SAFT EoS + cDFT with $k_{ij} = 0.0541$ and $k_{ij} = 0$ are consistent with the MD simulation. The density profile of MD simulations obtained by averaging 80 ns, shown in Fig. 3.7d, overpredicts the interfacial thickness due to the presence of fluctuation and slight lateral movement of the liquid slab. The MD simulation accurately predicted the vapor densities of both CO₂ and H₂, and the liquid densities of CO₂ and H₂ were overestimated by ca. 1.4% and underestimated by ca. 50%, respectively, compared to the PC-SAFT EoS + cDFT. The MD simulation accurately predicted the vapor densities of both CO₂ and H₂. The liquid density of CO₂ was overestimated by ca. 1.4%, while that of H₂ was underestimated by about 50% compared to the PC-SAFT EoS combined with cDFT. The apparent 50% underestimation of the H₂ liquid density in both the PC-SAFT EoS + cDFT and MD simulations is primarily due to its very low absolute density values, corresponding to a relative difference between 1 kg m⁻³ and 2 kg m⁻³. The difference in density profiles computed from the PC-SAFT EoS + cDFT with $k_{ij} = 0.0474$ and $k_{ij} = 0$ was insignificant. The MD simulation show no noticeable interfacial enrichment ($E_{H_2} \approx 1$). The PC-SAFT EoS + cDFT with $k_{ij} = 0.0474$ predicts a minor enrichment ($E_{H_2} \approx 1.1$) at 250 K and 100 bar.

To summarize, the PC-SAFT EoS with binary interaction parameters (k_{ij}) obtained from experimental P_{xy} data provided excellent agreement for the phase equilibria of the CO₂-CH₄, CO₂-Ar, and CO₂-N₂ mixtures, except for the well-known overprediction of critical pressures by PC-SAFT [68, 102]. The CO₂-H₂ system exhibited noticeable deviations due to the limited availability of experimental data for determining accurate k_{ij} values. MD simulations showed good agreement with experimental phase equilibria for the CO₂-CH₄, CO₂-Ar, and CO₂-H₂ mixtures but deviated for CO₂-N₂, where the absence of a binary interaction parameter in the MD simulations resulted in overpredicted CO₂ liquid-phase compositions. For the phase

equilibria at the corresponding temperatures and pressures, IFTs of the CO₂-CH₄ and CO₂-Ar mixtures obtained from MD simulations and from the PC-SAFT EoS combined with cDFT were in good agreement. In sharp contrast, for the CO₂-N₂ and CO₂-H₂ mixtures, the IFTs values predicted by the two approaches differ significantly. Additionally, the density profiles of the CO₂-CH₄, CO₂-Ar, and CO₂-N₂ mixtures obtained from MD simulations and from the PC-SAFT EoS combined with cDFT were compared at similar IFT magnitude ($\gamma = 4 \text{ mN m}^{-1}$) 250 K, except for the CO₂-Ar system, which was evaluated at 253.28 K. This showed that the enrichment at the interface is dependent on the impurity (LBC) component present in the mixture. It is expected to have larger enrichment at low pressures and high IFT [271], but for CO₂-H₂ mixture the enrichment of H₂ was observed to be insignificant at $\gamma = 8 \text{ mN m}^{-1}$ at 250 K and 100 bar. The enrichment of the LBC in the CO₂-CH₄, CO₂-Ar, and CO₂-N₂ mixtures, obtained from both MD simulations and the PC-SAFT EoS combined with cDFT for $k_{ij} = 0$ and $k_{ij} \neq 0$, is presented in Section S10 of the Supporting Information for this chapter [211]. Interfacial thicknesses for all mixtures, computed from MD simulations and from the PC-SAFT EoS combined with cDFT using $k_{ij} = 0$ and $k_{ij} \neq 0$, are provided in Section S10 of the Supporting Information for this chapter [211]. The results show that the interfacial thickness ranges from 0.35 nm to 2 nm for all binary mixtures, except in the vicinity of the critical pressure.

3.3.3. Effect of impurities on IFT

The effect of impurities on IFT was evaluated by comparing IFTs of CO₂ binary mixtures with those of pure CO₂. The IFTs of the binary mixtures were obtained using the PC-SAFT EoS [68, 102] combined with cDFT and the k_{ij} parameters listed in Table 3.1. At temperatures where the k_{ij} values are not reported in Table 3.1, namely 220 K for CO₂-CH₄, 220 K, 250 K, and 270 K for CO₂-Ar, and 270 K for CO₂-N₂ and CO₂-H₂, the k_{ij} parameters were determined using a third-order polynomial fit to the values listed in Table 3.1. Details of the fitting procedure and the corresponding plots are provided in Section S8 of the Supporting Information for this chapter [211]. Results at 220, 250, and 270 K for CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ are presented in Fig. 3.8 with the pure CO₂ IFTs. Pure CO₂ IFTs were computed using the Eq. 3.13 and using the coefficients of Cachadiña, Mulero and Tian [246]. The pure CO₂ IFTs are plotted as a function of $\gamma(P^{\text{sat}})$ in Fig. 3.8. The interfacial tension values $\gamma(P^{\text{sat}})$ for pure CO₂ were determined by converting the temperature dependent $\gamma(T)$ into a function of the saturated vapor pressure using $P^{\text{sat}}(T)$ values calculated from the Span and Wagner EoS [28]. Using $\gamma(P^{\text{sat}})$ of pure CO₂ as a reference, it can be observed from Fig. 3.8 that IFTs of the binary mixtures are found over the entire two-phase region, from the onset of vapor-liquid phase separation up to the critical pressure of each binary system. At the pressure where the two-phase region starts for each binary mixture at a given temperature,

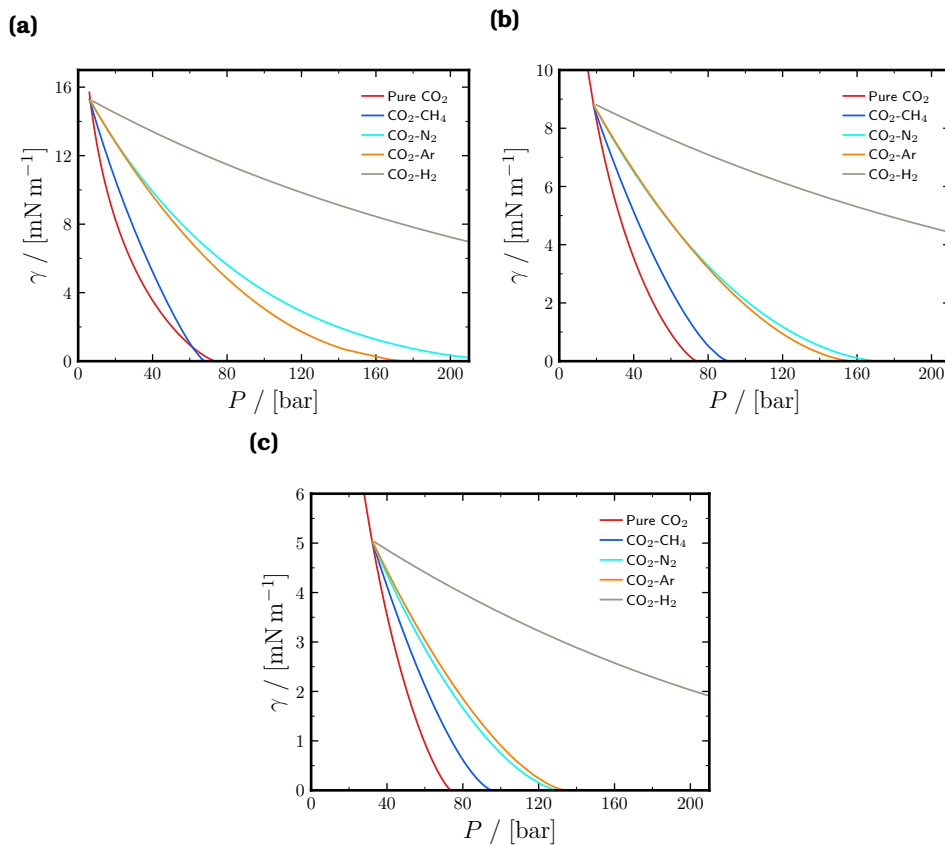


Figure 3.8. Comparison between pure CO_2 IFTs evaluated at the saturation pressure, $P^{\text{sat}}(T)$, and the IFTs of $\text{CO}_2\text{-CH}_4$, $\text{CO}_2\text{-N}_2$, $\text{CO}_2\text{-Ar}$, and $\text{CO}_2\text{-H}_2$ mixtures evaluated at the imposed pressure, P , for (a) 220 K, (b) 250 K, and (c) 270 K. All IFTs are computed using the PC-SAFT EoS [68, 102] with $k_{ij} \neq 0$ in combination with cDFT.

IFT coincides with that of pure CO_2 . This is because the liquid phase remains highly enriched in CO_2 , with only negligible impurity content, as observed in the Pxy diagrams of Figs. 3.4, 3.5, 3.6, and 3.7. In this dilute impurity point, the impurity components (CH_4 , Ar, N_2 , and H_2) do not noticeably affect the IFT, and the mixture IFT values coincide with those of pure CO_2 . At the dilute impurity point, the IFT can be approximated by the IFT of pure CO_2 evaluated at the saturation pressure of the binary mixture at that temperature. For example, in the $\text{CO}_2\text{-Ar}$ system at 220 K, the two-phase region begins at ca. 8 bar (Fig. 3.5a), where the liquid phase is nearly pure CO_2 . At this pressure and temperature, the IFT of the $\text{CO}_2\text{-Ar}$ mixture matches the pure CO_2 $\gamma(P^{\text{sat}})$, as shown in Fig. 3.8a.

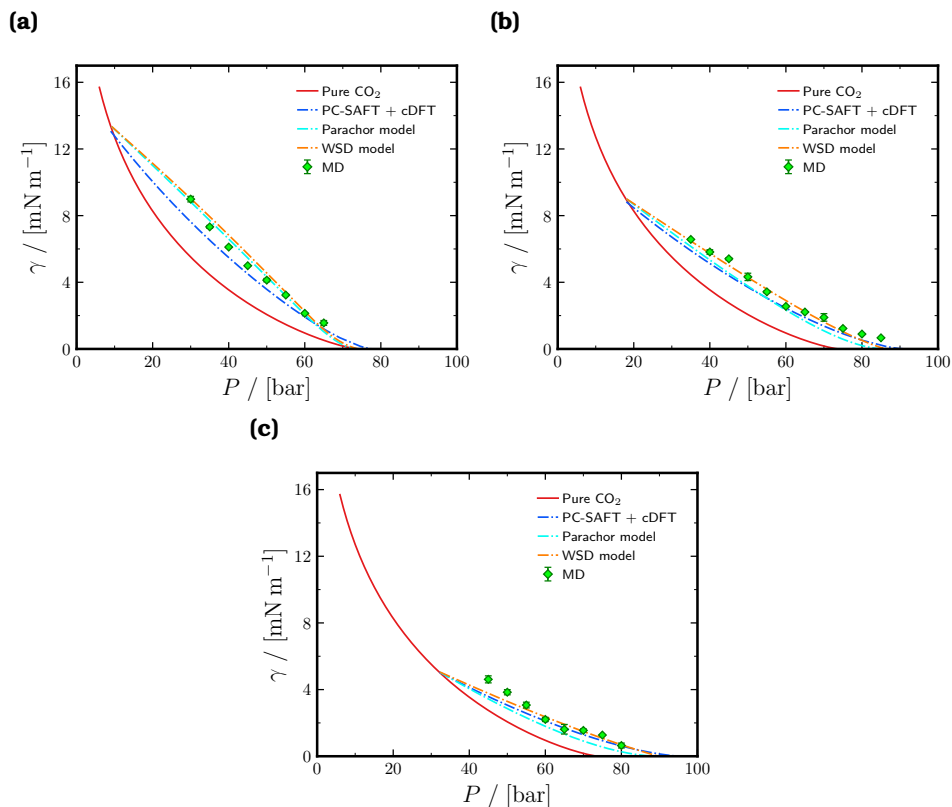


Figure 3.9. Comparison of IFTs computed from the PC-SAFT EoS [68, 102] with k_{ij} s combined with cDFT, Parachor model [233], Winterfeld–Scriven–Davis (WSD) model [146], and MD simulations of CO₂–CH₄ mixtures at (a) 230 K, (b) 250 K, and (c) 270 K, CO₂–Ar mixtures at (d) 223 K, (e) 253.3 K, and (f) 273.3 K, CO₂–N₂ at (g) 220 K, (h) 250 K, and (i) 273.12 K and, CO₂–H₂ mixtures at (j) 220 K, (k) 250 K, and (l) 273.12 K.

Similar behavior is observed for other mixtures. The CO₂–N₂ and CO₂–H₂ systems enter the two-phase region at ca. 7 bar, while the CO₂–CH₄ mixture begins at ca. 10 bar at 230 K. At these pressures, the mixture IFTs coincides with the pure CO₂ IFTs. The same trend is observed at higher temperatures, as shown in Fig. 3.8b and Fig. 3.8c, with the dilute impurity saturation pressures of the mixtures shifting to higher values as temperature increases.

Fig. 3.8 show that, as pressure increases at constant temperature, the CO₂-rich liquid phase is diluted by impurities, resulting in higher IFTs for the mixtures compared to pure CO₂. The magnitude of this increase depends on the volatility of the impurity component. Impurities with

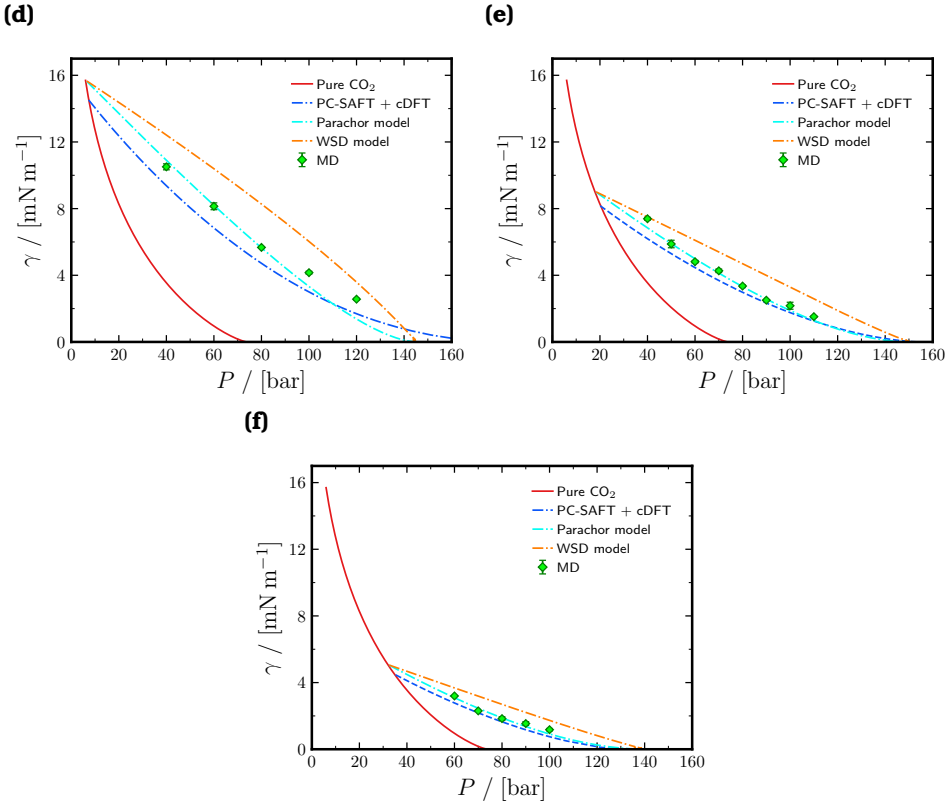


Figure 3.9. (continued)

very low boiling points, such as H_2 and Ar, dissolve only weakly in the liquid phase and therefore exhibit the largest increase in IFT. Moderately volatile species, such as N_2 , yield an intermediate response, whereas the most condensable impurity examined here, CH_4 , mixes more readily with liquid CO_2 and consequently produces the smallest enhancement in IFT. In general, binary mixtures follow the trend $\text{CO}_2\text{-H}_2$ shows the largest increase in IFT, followed by $\text{CO}_2\text{-Ar}$ and $\text{CO}_2\text{-N}_2$, which exhibit similar behavior, while $\text{CO}_2\text{-CH}_4$ produces the smallest increase in IFT relative to pure CO_2 . In addition, increasing the temperature reduces IFTs for both pure CO_2 and binary CO_2 mixtures, because larger temperatures weaken the interface and reduce the distinction between the liquid and vapor phases. In other words, the bulk density difference ($\rho^{\text{L}} - \rho^{\text{V}}$) decreases as the temperature increases. This leads to a decrease in the IFT [233].

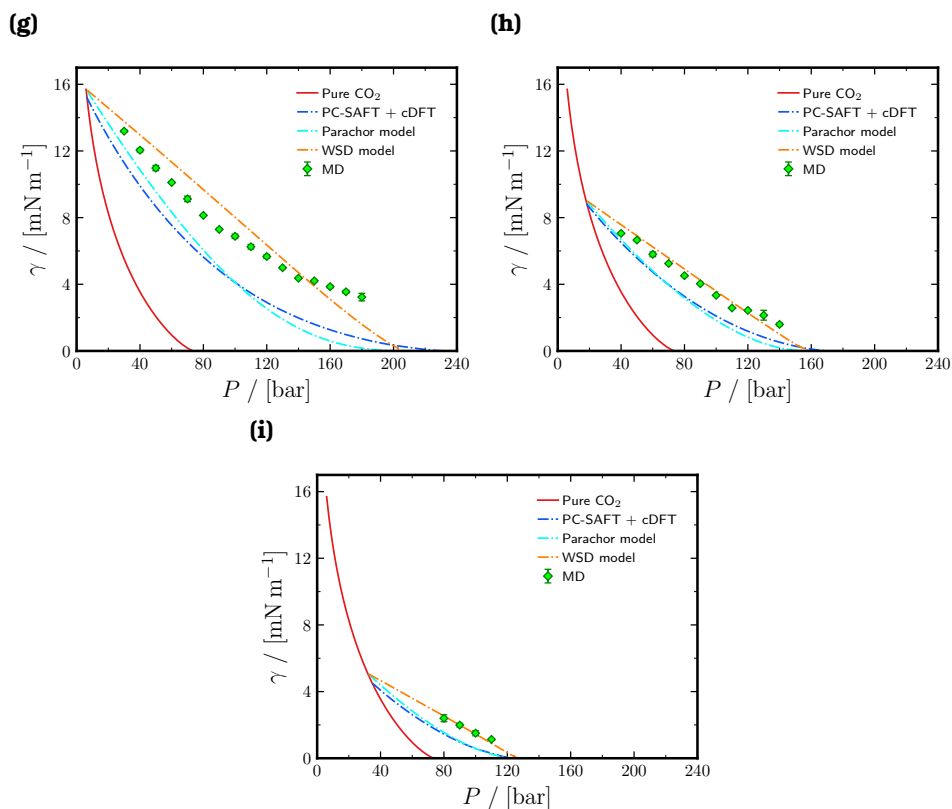


Figure 3.9. (continued)

3.3.4. Computation of IFT from Parachor and WSD models

We computed IFTs for the CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ binary mixtures using the semi-empirical Parachor [233] and WSD [146] models defined in Eq. 3.14 and Eq. 3.20, respectively. IFTs computed from the Parachor and WSD were compared with IFTs obtained from the PC-SAFT EoS coupled with cDFT [68, 102, 245] using the k_{ij} values listed in Table 3.1 and MD simulations. These comparisons were performed for the temperatures and mixtures given in Table 3.1, and the results are presented in Fig. 3.9, where pure CO₂ IFTs were also included as a reference using the procedure described in Section 3.3.3. The Parachor and WSD models yield identical IFT values at the lowest pressure where the two phases coexist for the corresponding binary mixture and temperature, as well as at the critical pressure (where $\gamma \rightarrow 0$) for all binary systems shown in Fig. 3.9. In contrast, the PC-SAFT EoS + cDFT tends to underestimate the IFTs near the low-pressure limit at the onset of the two-phase region when compared with the Parachor and WSD models. This deviation arises because the Parachor

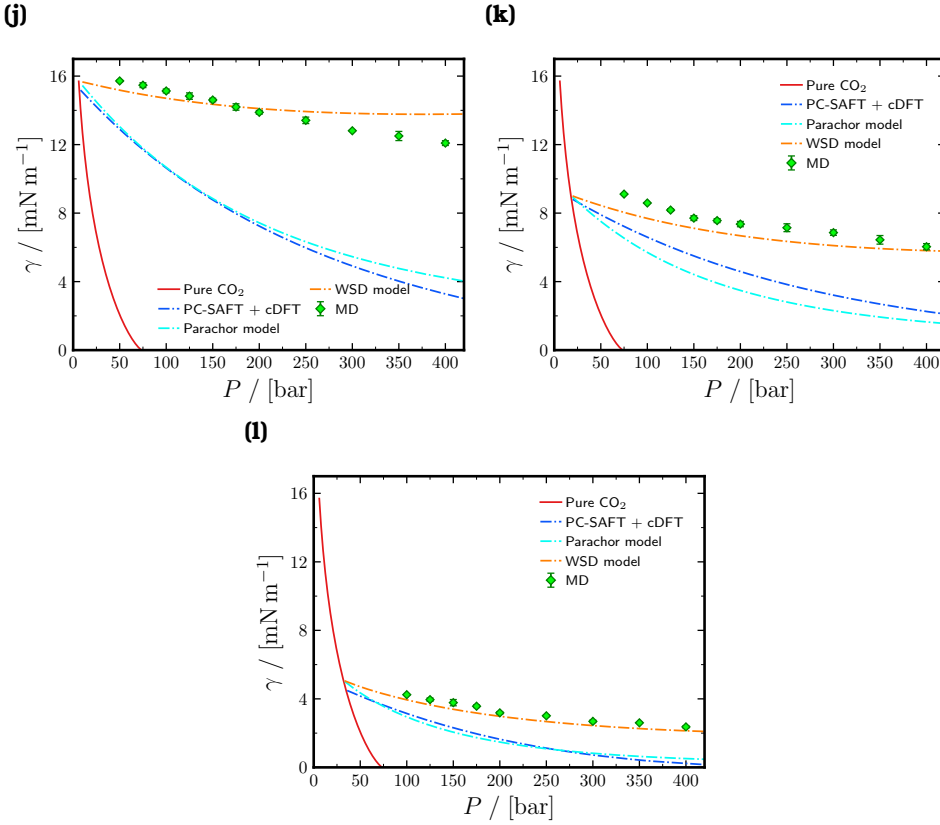


Figure 3.9. (continued)

and WSD models are based on the GERG-2008 EoS [58], whereas the cDFT approach uses the PC-SAFT EoS. In addition to the underestimation at the low-pressure limit, the PC-SAFT EoS also tend to overestimate the two-phase pressure range [68, 102]. Therefore, only the Parachor and WSD models were compared quantitatively with the MD obtained IFT values, while the PC-SAFT EoS combined with cDFT was used for quantitative analysis. Additionally, IFT values at pressures close to the critical pressure were neglected for analysis to avoid disproportionate errors. Fig. 3.9 show that the Parachor and WSD models reproduce the expected decrease in IFT with increasing pressure. However, both models deviate from MD and PC-SAFT EoS + cDFT results, particularly in capturing the curvature of the IFT-pressure relationship. To assess the predictive performance of the Parachor and WSD models for all the binary mixtures shown in Fig. 3.9, the MAPD values were evaluated relative to MD simulations for each binary system listed in Table 3.1.

Figs. 3.9a, 3.9b, and 3.9c show IFTs of CO₂-CH₄ system at 230, 250, and 270 K. The Parachor model deviates by MAPDs of ca. 5%, 13%, and 30% at 230, 250, and 270 K, respectively, whereas the WSD model has MAPDs of ca. 8%, 4%, and 12% at 230, 250, and 270 K, respectively. At 230, 250, and 270 K, the WSD model generally predicts higher IFTs than the PC-SAFT EoS + cDFT for the entire pressure range, with deviations more pronounced at 230 K and diminishing near the critical region. The Parachor model aligns more closely with PC-SAFT EoS + cDFT than the WSD model. While both models exhibit a similar decreasing trend with pressure that parallels PC-SAFT EoS + cDFT, the WSD model consistently produces slightly higher IFTs than the Parachor model compared to the PC-SAFT EoS + cDFT. The WSD model shows better agreement with MD simulations than the PC-SAFT EoS + cDFT. Figs. 3.9d, 3.9e, and 3.9f show IFTs of CO₂-Ar mixture at 223.07, 253.28, and 273.26 K. The WSD deviated by a MAPDs of ca. 34%, 31%, and 36% at 223.07, 253.28, and 273.26 K, respectively. The Parachor model had considerably less MAPDs of ca. 6%, 5%, 6% at 223.07, 253.28, and 273.26 K, respectively. Compared to the PC-SAFT + cDFT, the Parachor model had a better agreement than the WSD model. At pressures less than 40 bar (Figs. 3.9d, 3.9e, and 3.9f), the Parachor model differed significantly with PC-SAFT + cDFT, while MD simulations had a better agreement. Figs. 3.9g, 3.9h, and 3.9i show IFTs of CO₂-N₂ system at 220.07, 250, and 273.12 K. Compared to the PC-SAFT + cDFT, the Parachor model had a better agreement than the WSD model. The Parachor model insignificantly overestimated IFTs at pressures lower than ca. 80 bar in Figs. 3.9g, 3.9h, and 3.9i and underestimated at pressures larger than ca. 80 bar. The Parachor model deviated by MAPDs of ca. 33%, 28%, and 49% compared to MD simulations at 220.07, 250, and 273.12 K, respectively, while the Parachor model had MAPDs of ca. 13%, 8%, and 3%. It is important to note that the MD simulations systematically overpredicted the IFTs, liquid mole fractions, and bulk densities relative to the PC-SAFT + cDFT results with $k_{ij} \neq 0$ in Fig. 3.6, due to the use of $k_{ij}^{\text{MD}} = 0$. Consequently, the Parachor model showed higher MAPDs than the WSD model when compared with IFTs obtained from MD simulations. Figs. 3.9j, 3.9k, and 3.9l show IFTs of CO₂-H₂ system at 220, 250, and 273.1 K. The WSD model shows good agreement with the IFTs obtained from MD simulations, with MAPDs of approximately 3.7%, 9.7%, and 9.5% at 220, 250, and 273.1 K, respectively. In contrast, the Parachor model exhibits significantly larger deviations, with MAPDs approaching 50% relative to the MD results. The Parachor model had a better agreement with PC-SAFT + cDFT compared to MD simulations. To summarize, the Parachor and WSD models yielded better IFT predictions for the CO₂-CH₄ mixtures, whereas the Parachor model provided more accurate IFT predictions for the CO₂-Ar and CO₂-N₂ mixtures. For the CO₂-H₂ mixtures, the WSD model provided better IFT predictions than both the Parachor model and the PC-SAFT + cDFT approach, particularly considering the limited experimental phase

equilibria data available for determining the k_{ij} values at 220, 250, and 273.1 K.

3.3.5. Metastable phase boundaries of CO₂ and CO₂-binary mixtures

The metastable boundaries of pure CO₂ and impure CO₂ mixtures are essential for identifying when the fluid departs from equilibrium. These boundaries indicate the onset of mechanical and chemical nonequilibrium between phases during depressurization, decompression, or nozzle flow [243, 272, 273]. A liquid can be expanded past its equilibrium boundary and still remain liquid for a short time [56, 57, 235]. The point where it can no longer remain liquid is known as the SHL. Similarly, a vapor can be cooled or compressed past its equilibrium boundary and remain vapor until it reaches the SCL [57]. In this chapter, we compute the homogeneous metastable limits (SHL and SCL) of pure CO₂ and binary CO₂ mixtures using CNT [237, 238]. In practice, fluids undergo both homogeneous and heterogeneous nucleation. However, this chapter focuses on homogeneous nucleation since it is expected to dominate at larger temperatures, whereas heterogeneous nucleation is strongly influenced by surface properties and system geometry [56]. Compared to the other parameters discussed in Section 3.2.5 that are required to compute the SHL and SCL, the critical nucleation rate J and the interfacial tension (IFT) γ show the strongest sensitivity [235]. In particular, IFT has a dominant influence because the free energy barrier ΔG^* scales as γ^3 , which makes the SHL and SCL sensitive to small changes in IFT. In this chapter, the SHL and SCL were computed using a critical nucleation rate of $J = 10^{12}/\text{m}^3 \cdot \text{s}$, consistent with the approach of Skarsvåg *et al.* [235] and Aursand *et al.* [244]. Skarsvåg *et al.* [235] demonstrated that changing the critical nucleation rate J by six orders of magnitude produced a very small change in the predicted SHL of CO₂ mixtures at low temperatures, where heterogeneous nucleation dominates. At high temperatures, where homogeneous nucleation dominates the nucleation process, the predicted SHL remained essentially unchanged. In this chapter, the SHL and SCL of pure CO₂ and CO₂ binary mixtures with 95 mol% CO₂ and 5 mol% CH₄, Ar, N₂, or H₂ are computed by solving the CNT using a constant nucleation rate of $J = 10^{12}/\text{m}^3 \cdot \text{s}$. IFT values of pure CO₂ required to compute the SHL and SCL were estimated using MD simulations, the PC-SAFT EoS [68, 102], and the Guggenheim-Katayama equation (Eq. 3.13) [190, 219, 246] with coefficients reported by Cachadiña, Mulero and Tian [246] (REFPROP v10). For CO₂-CH₄, CO₂-Ar, and CO₂-N₂ mixtures, we computed IFT values with the Parachor model [233], which agrees well with both MD simulations and the PC-SAFT EoS [68, 102] with $k_{ij} \neq 0$, as detailed in Section 3.3.4. Similarly, IFT values for CO₂-H₂ were obtained using the WSD model [146], which agrees well with MD simulations. Although extending MD simulations to

a broader temperature range would allow IFT values at saturation to be determined for each mixture and composition, such an extension is computationally intensive. The PC-SAFT EoS combined with cDFT also requires temperature-dependent values for $k_{ij}(T)$. Binary interaction parameters are typically expressed as temperature-dependent correlations obtained by fitting a limited number of k_{ij} values, such as the three values listed for each mixture in Table 3.1. Constructing a $k_{ij}(T)$ relationship from few data points can introduce substantial uncertainty at temperatures where no experimental measurements exist. For example, if k_{ij} is only available at 220 K and 250 K, the interpolated k_{ij} at an intermediate temperature such as 225 K could be unreliable. For these reasons, we use the semi-empirical Parachor and WSD models in this chapter to estimate IFT values for computing the SHL and SCL. We computed the saturation pressures, liquid and vapor number densities, and liquid and vapor mole fractions needed for the SHL and SCL estimation, using the Span and Wagner EoS [28] for pure CO₂ and the GERG-2008 EoS [58] for binary mixtures. In addition to the SHL and SCL, we also present the spinodal limits for pure CO₂ and binary CO₂ mixtures, computed using the PC-SAFT EoS. The spinodal marks the absolute thermodynamic stability limit of a single phase and corresponds to the points where the pressure-density curve reaches a local maximum or minimum, i.e., where $\frac{\partial P}{\partial \rho} = 0$. Beyond this boundary, which corresponds to the region enclosed by the liquid and vapor spinodal lines, the homogeneous fluid becomes thermodynamically unstable, cannot suppress infinitesimal density fluctuations, and undergoes spontaneous phase separation through spinodal decomposition [237, 238]. The PC-SAFT EoS was chosen because multiparameter models such as Span and Wagner and GERG-2008 can introduce artificial Maxwell loops that distort the true thermodynamic stability boundary [60]. Nevertheless, thermodynamic properties in metastable and unstable regions, including spinodal lines and IFTs, can also be computed from multiparameter EoSs using extrapolation techniques, such as the isochoric pressure extrapolation with second-order temperature expansion proposed by Hammer, Aasen and Wilhelmsen [274, 275]. At realistic conditions, fluids do not typically reach the spinodal limit because nucleation initiates phase change earlier at the SHL and SCL [56]. This early onset of phase change has also been demonstrated experimentally for pure CO₂ [56, 276]. For a detailed description of spinodal lines estimation from EoS, the reader is referred to Keller, Langenbach and Hasse [277] and Aursand *et al.* [244].

Fig. 3.10a presents the PT phase diagram of pure CO₂ including the saturation line and spinodal lines computed from the Span and Wagner and PC-SAFT EoS, respectively. The critical points of the Span and Wagner and PC-SAFT EoS are shown as a closed and open star marker, respectively. It is clear that the critical point of PC-SAFT EoS is overestimated, as are the spinodal lines. The region between the saturation curve and the liquid spinodal defines stable vapor and metastable liquid. The region between

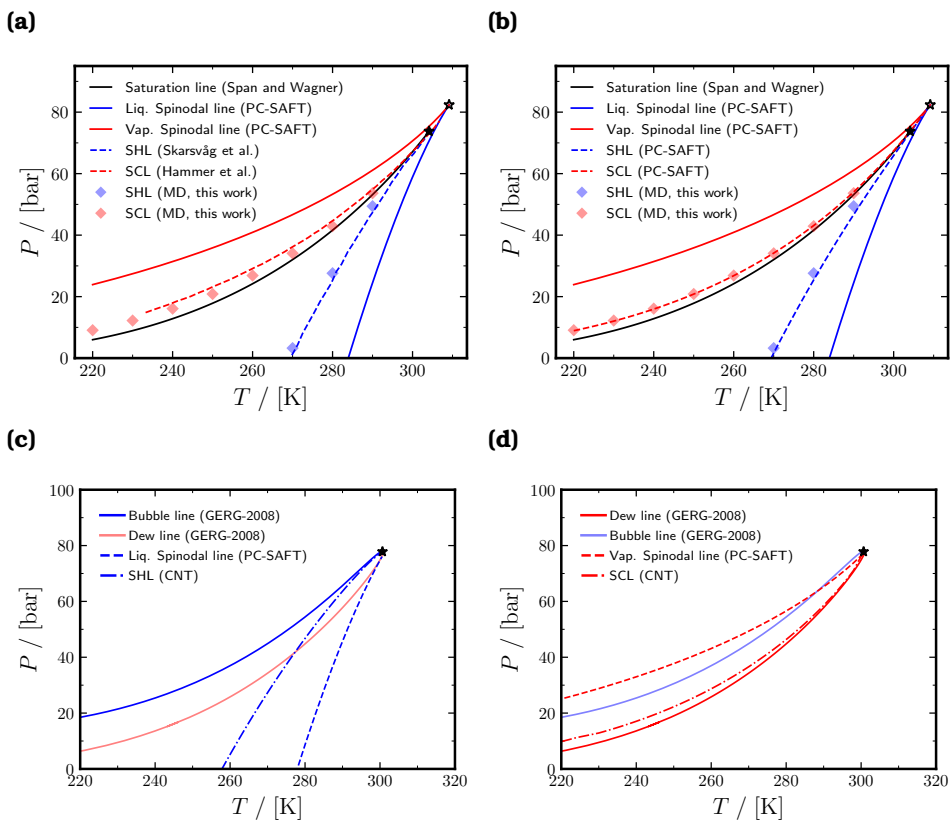


Figure 3.10. PT diagrams of pure CO_2 including the saturation line obtained from the Span and Wagner EoS [28], the liquid and vapor spinodal lines from the PC-SAFT EoS [68, 102], and the metastable limits (SHL and SCL) estimated using CNT [237] with IFTs computed from (a) MD simulations, together with the corresponding SHL and SCL from Skarsvåg *et al.* [235] and Hammer *et al.* [57]. (b) Comparison of the SHL and SCL obtained using IFTs from the PC-SAFT EoS with the limits obtained from MD computed IFTs. (c)–(j) PT diagrams for binary CO_2 mixtures containing 95 mol% CO_2 and 5 mol% impurity. The bubble and dew lines are calculated using the GERG-2008 EoS [58], and the spinodal lines are obtained from the PC-SAFT EoS [68, 102]. Metastable limits from the CNT [237] are determined using IFTs from the Parachor model for (c–d) CO_2 – CH_4 , (e–f) CO_2 – Ar , and (g–h) CO_2 – N_2 , and using IFTs from the WSD model [146] for (i–j) CO_2 – H_2 . (c), (e), (g), and (i) include the dew and bubble lines together with the liquid spinodal and SHL, whereas (d), (f), (h), and (j) include the dew and bubble lines together with the vapor spinodal and SCL. The area between the metastable and spinodal lines in (a–j) is the thermodynamically unstable area.

the vapor spinodal and the saturation curve defines stable liquid and metastable vapor. The area adjacent to the liquid spinodal corresponds to unstable liquid and stable vapor, whereas the region above the vapor spinodal corresponds to stable liquid and unstable vapor. In addition to the saturation and spinodal lines, the SHL and SCL of pure CO₂ estimated using the CNT with IFT values computed from MD simulations are shown in Fig. 3.10a. It is evident that the estimated SHL and SCL lines lie between the saturation curve and the spinodal limits, indicating that phase change proceeds through nucleation at these loci. The predicted SHL and SCL were validated with the results of Skarsvåg *et al.* [235] and Hammer *et al.* [57], respectively. In both cases, the SHL and SCL obtained in this chapter were found to be slightly lower than the corresponding limits reported by Skarsvåg *et al.* [235] and Hammer *et al.* [57]. In Fig. 3.10b we also show the saturation line, spinodal lines, SHL, and SCL (computed using IFTs obtained from MD simulations) in addition to the SHL, and SCL computed using IFTs obtained from the PC-SAFT EoS + cDFT. It is clear that the SHL predicted from the IFT values of the PC-SAFT EoS + cDFT is slightly overestimated when compared to SHL computed using IFTs obtained from MD simulation. This is due to the underestimation of IFTs computed from the PC-SAFT EoS + cDFT as shown in Figs. 3.3a and 3.3b. However, the SCL estimated using the IFTs computed from the PC-SAFT EoS + cDFT and MD simulations were in excellent agreement with each other. This implies that the SCL limit is not sensitive to the magnitude of IFT. The same argument can be made with the SHL and SCL limits estimated using IFTs computed from the Guggenheim-Katayama equation [190, 219, 246] with coefficients reported by Cachadiña, Mulero and Tian [246] (REFPROP v10).

Fig. 3.10 shows the *PT* phase diagrams of CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ binary mixtures containing 95 mol% CO₂ and 5 mol% impurities (CH₄, Ar, N₂, and H₂). The dew and bubble lines were computed using the GERG-2008 EoS, and the liquid and vapor spinodal lines were estimated using the PC-SAFT EoS. The critical points of binary mixtures were computed using the GERG-2008 EoS and shown as the closed markers in Figs. 3.10 and 3.11. For clarity, Fig. 3.10 excludes spinodal points above the critical point, which result from the PC-SAFT EoS overestimation of the critical point. To illustrate better, the SHL lines of CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ are shown in Figs. 3.10c, 3.10e, 3.10g, and 3.10i, respectively, with the bubble, dew, and liquid spinodal line. Similarly, the SCL lines of CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ are shown in Figs. 3.10d, 3.10f, 3.10h, and 3.10j, respectively, with the bubble, dew, and vapor spinodal line. The SHL lines adjacent to the bubble lines in Figs. 3.10c, 3.10e, 3.10g, and 3.10i mark the point at which the metastable liquid becomes thermodynamically unstable with respect to vapor formation, indicating the onset of homogeneous nucleation. Similarly, the SCL lines located above the dew lines in Figs. 3.10d, 3.10f, 3.10h, and 3.10j mark the

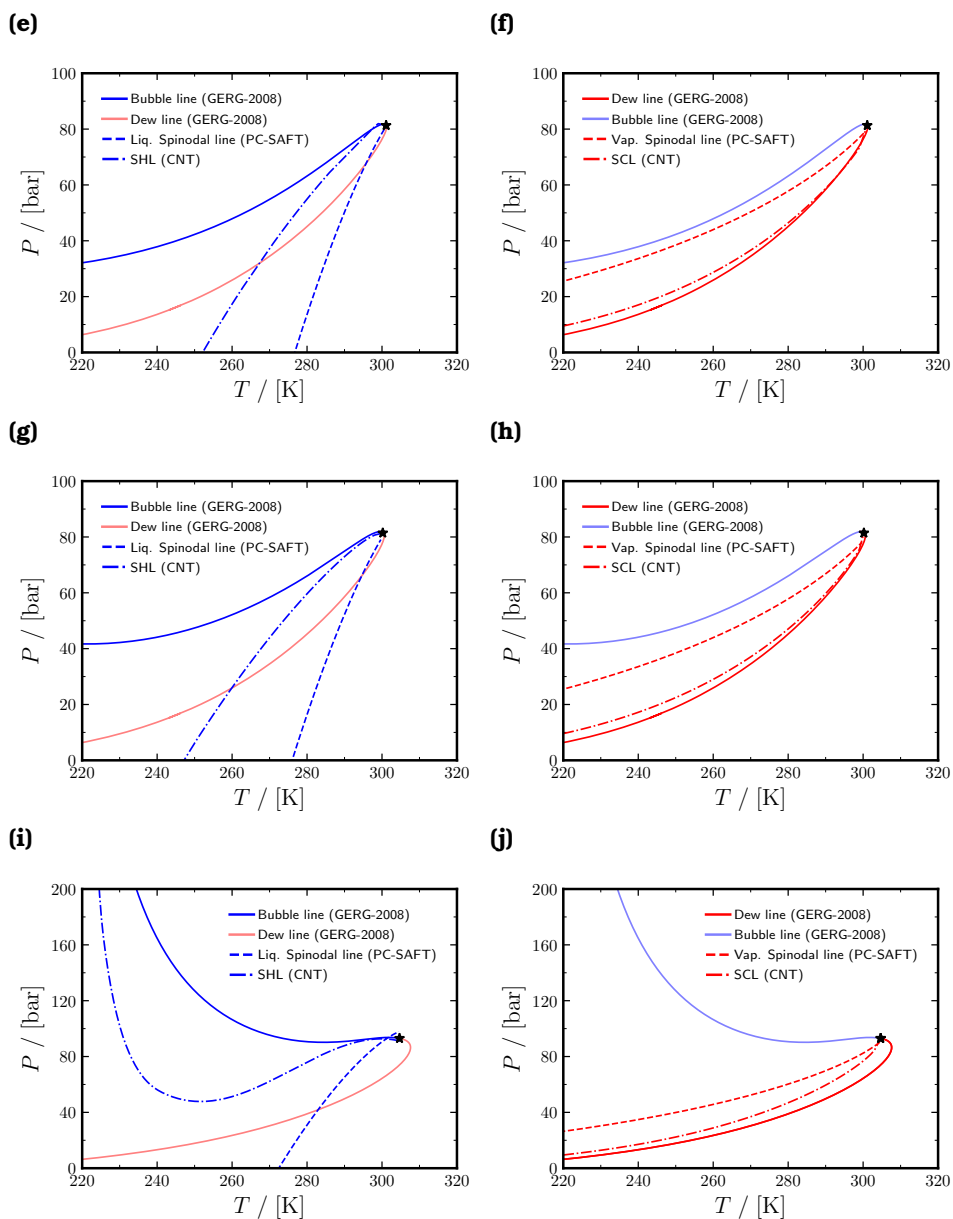


Figure 3.10. (continued)

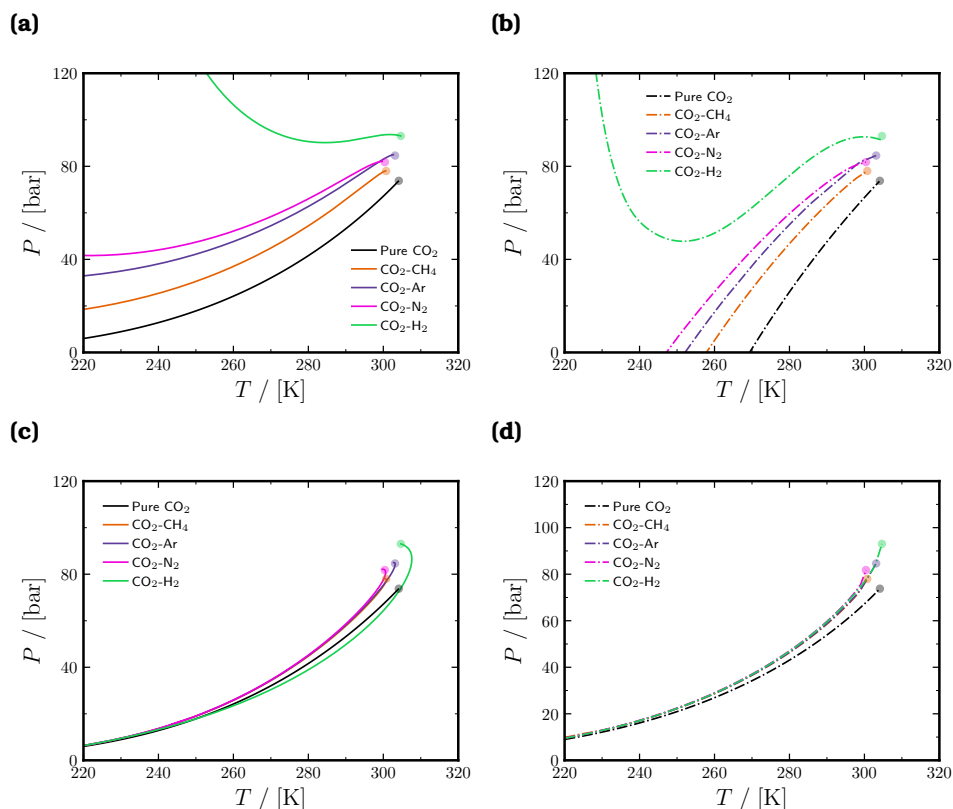


Figure 3.11. Comparison of (a) bubble and (c) dew lines of the CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ binary mixtures containing 95 mol% CO₂ and 5 mol% impurity, computed using the GERG-2008 EoS [58], with the saturation line of pure CO₂ computed using the Span and Wagner EoS [28]. Comparison of (b) SHL and (d) SCL lines of the same binary mixtures with the pure CO₂ SHL and SCL lines computed using CNT [237].

point at which the metastable vapor becomes unstable with respect to liquid formation. For example, in the PT diagram of the CO₂-CH₄ mixture at 270 K (Figs. 3.10c and 3.10d), the liquid becomes unstable to vapor nucleation at ca. 28 bar, significantly lower than the equilibrium bubble pressure of ca. 44 bar. Similarly, the vapor becomes unstable to liquid nucleation at ca. 38 bar, slightly above the equilibrium dew pressure of ca. 37 bar. The area between the metastable limits (SHL and SCL) and the corresponding spinodal lines shown in Fig. 3.10 is the thermodynamically unstable area, where infinitesimal density perturbations grow spontaneously and drive phase transition. In comparison with the SCL limits, the SHL limits show a more pronounced deviation from the corresponding saturation lines

(bubble and dew lines), particularly at temperatures far from the critical temperatures of binary mixtures. A direct comparison of the bubble lines of the CO_2 binary mixtures with the pure CO_2 saturation line is presented in Fig. 3.11a, while Fig. 3.11b compares the SHL lines of binary mixtures with the pure CO_2 SHL line. In both cases, the phase boundary shift follows the same trend, with the $\text{CO}_2\text{-H}_2$ mixture exhibiting the largest deviation from pure CO_2 , followed by $\text{CO}_2\text{-N}_2$, $\text{CO}_2\text{-Ar}$, and $\text{CO}_2\text{-CH}_4$. The dew lines of the CO_2 binary mixtures, shown in Fig. 3.11c, lie very close to one another and closely follow the pure CO_2 saturation line over the full two-phase temperature range, with noticeable deviations close to the critical region. The largest deviation compared to the pure CO_2 is observed for the $\text{CO}_2\text{-H}_2$. A similar behavior is observed for the SCL limits of binary mixtures shown in Fig. 3.11d, where all mixtures exhibit nearly identical curves at low and intermediate temperatures. Interestingly, impurities do not introduce the systematic shifts seen in the SHL limits, indicating that the vapor-phase metastable boundary is likely insensitive to impurities and thus the IFT. However, when compared with the pure CO_2 SCL line, the binary mixture SCL lines deviate substantially above ca. 270 K, where homogeneous nucleation dominates. This behavior indicates that a small 5 mol% impurity concentration can influence the instability of the vapor, despite the SCL lines of binary mixtures showing no significant difference. These findings indicate that the liquid-to-vapor transition through homogeneous nucleation is strongly influenced by both impurity type and concentration, whereas the vapor-to-liquid transition is primarily governed by concentration and exhibits minimal sensitivity to impurity type. The delay in phase transition, represented by the SHL and SCL lines estimated from the CNT, must be taken into account when modeling CO_2 flows undergoing decompression, depressurization, valve throttling, nozzle discharge, or flow through constrictions. The resulting PT diagrams, annotated with the saturation line, SHL, and SCL, provide essential insight into the thermodynamic behavior of two-phase CO_2 rich mixtures with impurities. The PT diagrams provide essential input for modeling two-phase CO_2 flows [57, 235] and for interpreting controlled laboratory studies, including vertical test section experiments that examine rapid phase transitions and nucleation behavior relevant to CCS injection systems [278]. Future work will focus on experiments that examine rapid phase transitions of CO_2 with impurities in the vertical test section.

3.4. Conclusions

In this chapter, we used MD simulations and the PC-SAFT EoS coupled with cDFT to compute the phase equilibria and interfacial properties of pure and $\text{CO}_2\text{-CH}_4$, $\text{CO}_2\text{-Ar}$, $\text{CO}_2\text{-N}_2$, and $\text{CO}_2\text{-H}_2$ binary CO_2 mixtures. The VLE and IFTs of pure CO_2 were validated with the Span and Wagner EoS [28] and the Guggenheim-Katayama [219] equation using the coefficients of

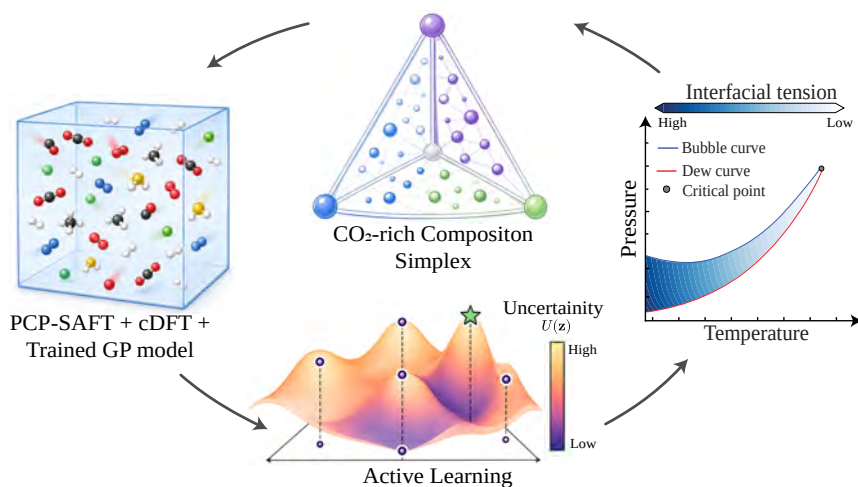
Cachadiña, Mulero and Tian [246], respectively. IFTs of pure CO₂ computed from MD simulations and the PC-SAFT EoS coupled with cDFT were within the maximum RPD of 11.7% reported by Cachadiña, Mulero and Tian [246], except at temperatures close to the critical temperature. The interfacial thickness of pure CO₂ was found to range from approximately 0.5 nm to 2 nm for the temperature range of 220 K to 290 K, based on both MD simulations and the PC-SAFT EoS combined with cDFT. The phase equilibria computed from MD simulations and the PC-SAFT EoS for CO₂-CH₄ mixtures at 230, 250, and 270 K and for CO₂-Ar mixtures at 223.1, 253.2, and 273.2 K showed good agreement with experimental data, except for the (well-known) overestimation of the critical pressures by the PC-SAFT EoS. For CO₂-N₂ mixtures at 220, 250, and 273.1 K, the PC-SAFT EoS with $k_{ij} \neq 0$ exhibited excellent agreement with experimental data, whereas MD simulations overpredicted the liquid mole fraction of CO₂ due to the absence of binary interaction parameters in the MD simulations. For CO₂-H₂ mixtures at 220, 250, and 273.1 K, the PC-SAFT EoS with $k_{ij} \neq 0$ produced phase envelopes that deviated significantly from the limited experimental data available, while MD simulations showed better agreement. At the same temperatures and pressures corresponding to the computed phase equilibria, the IFT values of CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ systems were computed. IFTs of binary mixtures computed from MD simulations showed good agreement with the PC-SAFT EoS with $k_{ij} \neq 0$, but IFTs of CO₂-N₂ and CO₂-H₂ differed significantly. The interfacial thickness of the CO₂-CH₄, CO₂-Ar, CO₂-N₂, and CO₂-H₂ binary mixtures was found to range from approximately 0.35 nm to 2 nm, excluding the vicinity of the critical pressure, where the interfacial thickness diverges. The density profiles from MD simulations and the PC-SAFT EoS + cDFT agreed well for the CO₂-CH₄, CO₂-Ar, and CO₂-H₂ systems. However, for the CO₂-N₂ system, the density profiles from MD simulations differed from those predicted by the PC-SAFT EoS combined with cDFT due to $k_{ij}^{\text{MD}} = 0$. The enrichment of the impurities (LBC) at the interface was identified for all binary mixtures based on the density profiles obtained from MD simulations and the PC-SAFT EoS + cDFT. The enrichment of LBC decreased with increasing temperatures and pressures. We showed that the enrichment at the interface is dependent on the LBC in the mixture. The influence of impurities on IFT was assessed by comparing IFTs predicted using the PC-SAFT EoS combined with cDFT framework. For binary CO₂ systems, the IFTs initially coincide with those of pure CO₂ at the onset of phase separation, since the liquid phase is almost entirely CO₂ at that point. As pressure increases and impurities dissolve into the CO₂ rich liquid phase, the IFT increases compared to pure CO₂. The magnitude of this increase follows impurity volatility, with H₂ having the largest effect, followed by Ar and N₂ (similar), and then CH₄. It was also observed that increasing temperature leads to lower IFTs for all binary mixtures. To determine which semi-empirical approach most

accurately predicts interfacial tensions of binary mixtures considered in this chapter, the Parachor and WSD models were evaluated with IFTs obtained from MD simulations and the PC-SAFT EoS + cDFT framework. Both semi-empirical models capture the overall IFT-pressure trend, but their quantitative accuracy varies depending on the mixture. The Parachor model performs best for the CO_2 -Ar and CO_2 - N_2 mixtures, while both models show strong agreement for the CO_2 - CH_4 mixture. For the CO_2 - H_2 mixtures, the WSD model provides the most reliable predictions compared to MD simulations. Based on the IFTs computed from the semi-empirical models and their agreement with results from the PC-SAFT EoS + cDFT and MD simulations, we applied the Parachor and WSD models to estimate the SHL and SCL lines in the PT phase diagram. These estimates were made for binary CO_2 mixtures containing 95 mol% CO_2 and 5 mol% impurity (CH_4 , Ar, N_2 , or H_2). To validate the approach for predicting the metastable limits (SHL and SCL), we computed SHL and SCL for pure CO_2 using IFTs computed from MD simulations, the PC-SAFT EoS + cDFT, and the Guggenheim-Katayama correlation. The computed SHL and SCL lines differed slightly from those reported in the literature [57, 235]. We compared the bubble, SHL, dew, and SCL lines of the binary mixtures with the saturation, SHL, and SCL lines of pure CO_2 . The bubble line exhibited the largest shift for the CO_2 - H_2 mixture, followed by CO_2 - N_2 , CO_2 -Ar, and CO_2 - CH_4 , relative to pure CO_2 . The SHL lines of the mixtures showed the same trend, as expected. In contrast, the SCL lines of the mixtures displayed negligible differences from one another and aligned closely with the pure CO_2 SCL line below 270 K. Near the critical region of each mixture, the SCL lines of all the binary mixtures exhibited noticeable deviations compared to the pure CO_2 SCL line. Our results demonstrate that the SHL depends on both the impurity type and its concentration, while the SCL is influenced primarily by the overall impurity concentration rather than the specific species present in the CO_2 stream. These findings emphasize that understanding IFT and its use in estimating metastable limits are crucial for modeling rapid phase transitions in CO_2 rich mixtures.

4

ML Predictions of CO₂ Phase Equilibria & Interfacial Properties

This chapter is based on the following publication: D. Raju, R. Skartlien, M. Ramdin and T. J. H. Vlugt. 'Machine Learning-Based Prediction of Phase Equilibria and Interfacial Properties of Multicomponent CO₂ Mixtures with Impurities: Application to CO₂ Transportation'. *Industrial & Engineering Chemistry Research* 65 (2026), pp. 18863–18890.



4.1. Introduction

Carbon dioxide CO₂ is one of the major greenhouse gases that drive global warming [148, 212]. Carbon capture has become an essential process to reduce CO₂ emissions and to prevent global warming [7, 9]. Techniques related to the CO₂ capture, transport, storage, and utilization of CO₂ are collectively referred to as Carbon Capture, Utilization, and Storage (CCUS). The transportation of CO₂ is essential for multiple CCUS applications, including direct air capture [11], enhanced oil recovery [27], methanol production [27], and dedicated reservoir storage (e.g., Northern Lights, Porthos) [17, 19]. CO₂ is transported efficiently from capture facilities to storage reservoirs by pipelines or ships [37]. CO₂ obtained from capture processes is typically not pure. The composition of the captured CO₂ depends on the capture technology and the source stream, such as flue gas in post-combustion capture, syngas in pre-combustion capture, and oxy-fuel flue gas in oxy-combustion capture [35]. The impurities found primarily in post-combustion are typically N₂, O₂, and Ar, whereas in oxy-fuel combustion, high concentrations of Ar, N₂, and O₂ are found [19, 27, 43, 46, 47]. In pre-combustion, the impurities are mostly H₂, CH₄, and CO [19, 27, 43, 46, 47]. Along with the noncondensable impurities such as Ar, N₂, H₂, CH₄, O₂, and CO [19, 27, 43, 46, 47], minor traces of corrosive impurities such as H₂O, H₂S, SO_x, NO_x, O₂, and hydrocarbons (C₂H₆ and C₃H₈) are also found in the captured source streams. These impurities affect the thermophysical behavior and phase stability of the CO₂ stream by broadening the two-phase envelope. For efficient transportation, CO₂ streams must be maintained in a dense single phase [49, 53, 157]. This reduces the volumetric flow rate and helps prevent phase separation, which can otherwise cause flow instability and pressure losses [53, 157]. This requirement is particularly challenging during transient operations, where pressure and temperature fluctuate [35, 213]. While steady-state operation targets a dense phase, systems often undergo startup, shutdown, depressurization, and planned maintenance, which can lead to two-phase flow in CO₂ transportation systems [55, 214, 278]. At such conditions, accurate knowledge of the thermodynamic and transport properties of impure CO₂ mixtures becomes indispensable for modeling their flow and phase behavior [35, 60]. Compared to other thermodynamic and transport properties such as density, heat capacity, speed of sound, and viscosity, the Interfacial Tension (IFT) is one of the important properties of the fluid mixture that governs the bubble and droplet formation and stability, and interphase heat and mass transfer [234, 235, 239]. During depressurization of the CO₂ stream through a valve or nozzle, the CO₂ flow may experience delayed phase change during that fast transient event [55]. The delayed phase change needs to be accounted for in Computational Fluid Dynamics (CFD) simulations to correctly predict mass flow rate and pressure, thereby enabling the accurate evaluation of pipeline forces and the associated risk of running ductile fracture [211, 235].

Experimental measurements and Molecular Dynamics (MD) simulations are widely used to determine phase equilibria and interfacial properties [127, 211, 220, 228, 239]. Experimental determination of IFTs for a wide range of mixtures and compositions encountered in CO₂ transportation is costly and time-consuming, resulting in limited and scarce experimental data for impure CO₂ mixtures [35, 36, 44, 215, 236]. MD simulations using classical force fields have proven to be a powerful tool to predict interfacial properties [127, 211, 236, 239]. Performing MD simulation for a wide range of compositions encompassing a broad range of two-phase conditions to compute interfacial properties can be extremely computationally intensive [211]. In addition to experimental measurements and MD based computations, Equation of State (EoS) based interfacial modeling approaches, including classical Density Functional Theory (cDFT), Density Gradient Theory (DGT), Square Gradient Theory (SGT), Linear Gradient Theory (LGT), and semi-empirical correlations, have been used to predict interfacial properties [211, 220, 231, 280]. The EoS-based cDFT, DGT, SGT, and LGT models closely reproduce experimental measurements [220, 281] and MD simulation results [109, 110, 211, 282]. To illustrate a few, Chow *et al.* [220] used the Statistical Associating Fluid Theory (SAFT)-Variable Range (VR) Mie EoS [69] coupled with SGT to predict the IFT of binary and ternary mixtures containing H₂O, CO₂, Ar, and N₂. The predictions closely reproduced the experimental measurements, and the SAFT-VR Mie EoS + SGT framework was recommended for modeling the IFTs of H₂O–CO₂ systems containing diluent gases. In Chapter 3, we computed the interfacial properties of binary CO₂ mixtures containing Ar, N₂, CH₄, and H₂. These properties included the IFT, interfacial thickness, and impurity enrichment at the interface. The computations were performed using MD simulations and the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) EoS + cDFT framework, and the results were compared with predictions from the semi-empirical Parachor [145] and Winterfeld–Scriven–Davis (WSD) [146] models. The PC-SAFT EoS + cDFT with temperature-dependent Binary Interaction Parameters (BIPs) showed a good agreement with MD simulations, and both the Parachor and the WSD semi-empirical models were found to have significant deviations. A wide range of empirical and semi-empirical models has been proposed in the literature to model the IFTs of mixtures, differing in their physical basis, flexibility, and number of adjustable or fitted coefficients. In a recent comparative study, Mulero *et al.* [247] assessed ten empirical and semi-empirical models for binary *n*-alkane mixtures and reported that the WSD model [146] yields the best overall performance for estimating IFT, provided that all components in the mixture are subcritical. When this condition is not satisfied, the Redlich–Kister model [283] was found to perform best. However, the WSD model requires only one fit coefficient, whereas the Redlich–Kister model requires three. Similar limitations in terms of requiring fitted coefficients and condition-dependent performance are also

observed in other semi-empirical models reported by Mulero *et al.* [247], particularly when one or a few components in the mixture are in a supercritical state [284]. Complementing these traditional modeling approaches, Machine Learning (ML) has emerged as a promising alternative by leveraging available thermophysical data [285]. Maginn and coworkers [285] used the CRC Handbook of Chemistry and Physics [286] as the primary experimental source to train the residuals obtained from the Joback–Reid (JR) Group Contribution (GC) method. The computationally efficient JR GC method often shows systematic biases (residuals). The systematic biases of normal boiling temperature (T_b), enthalpy of vaporization at T_b ($\Delta H_{\text{vap}, T_b}$), normal melting temperature (T_m), critical pressure (P_c), critical molar volume (V_c), and critical temperature (T_c) are compared to the experimental source [286] of 485 to 5640 molecules. The computed residuals were then trained on the Gaussian Process Regression (GPR) model to improve the accuracy of thermodynamic property predictions using the JR GC method and to quantify the associated uncertainties [285]. The hybrid model achieved predictive performance of a coefficient of determination (R^2) > 0.85 for T_b , $\Delta H_{\text{vap}, T_b}$, P_c , V_c , and T_c [285]. Focused on IFT, Mouallem *et al.* [287] developed ML models to estimate CO₂–brine IFT for CO₂ geo-storage using a large literature dataset of 2896 experimental data points covering wide ranges of pressure, temperature, salinity, salt type, and CH₄ and N₂ impurities. These authors compared Gradient Boosting (GB), eXtreme Gradient Boosting (XGB), Least Squares Boosting (LSB), GPR, Artificial Neural Network (ANN), and Genetic Programming, and found that GB performed the best with $R^2 = 0.964$. When a suitable database is unavailable or nonexistent, numerous studies have used physics-based models, such as molecular simulations and EoSs, to generate representative datasets for application-specific compositions and thermodynamic conditions [287–290]. Albà *et al.* [288] used the polar soft-SAFT EoS to generate saturation properties data for CO₂-based refrigerant mixtures and trained ANN to predict bubble and dew-point temperatures. The ANN model achieved $R^2 \approx 1$, demonstrating its suitability for screening sustainable refrigerant mixtures [288]. It is important to note that purely data-driven ML models do not always follow the expected physical behavior. This is especially important for tree-based models, which typically produce piecewise-constant predictions and have limited ability to extrapolate beyond the training range. As a result, these models may not fully capture the smooth variation of thermophysical properties, such as IFT, with temperature and pressure. This motivates the use of physics-informed hybrid models, where a physically based model first describes the main trend, and the ML model is then used to learn only the remaining residual error. Although several ML models have been developed to predict the IFT of CO₂–brine systems [287, 291–293] and other thermophysical properties of CO₂-based mixtures [294–296], including saturation properties [288], a comparative assessment of ML models for CO₂-rich mixtures containing

transportation-relevant impurities is still lacking. Such an assessment first requires a training dataset, and building that dataset dominates the cost because every point has to be generated by a physics-based computation, and the composition space of CO₂ transport is high-dimensional and thinly populated. The number of feed compositions required to reach a target accuracy, therefore, determines the computational cost of training ML models, and the most efficient sampling strategy for these compositions has not been established for CO₂-rich mixtures.

In this chapter, we use the Perturbed-Chain Polar Statistical Associating Fluid Theory (PCP-SAFT) EoS [68, 102, 103] combined with cDFT [245] to compute the IFT, interfacial thickness, bubble-point pressure, and dew-point pressure of multicomponent CO₂-rich mixtures containing transportation-relevant impurities. The CO₂ mole fraction ranges from 95% to 99.99%, while the total impurity concentration varies from 5% to 0.01%. This concentration range reflects practical CO₂ transportation conditions [19], where high CO₂ purity is desired to maintain dense-phase operation and reduce the risk of phase splitting and flow-assurance problems [19, 27, 35]. The impurities considered are Ar, N₂, CH₄, H₂, O₂, CO, and H₂S. Other impurities, such as H₂O, NO_x, and SO_x, are typically present at parts-per-million levels and are therefore not included [27, 35]. The inclusion of these impurities enables the model to cover the main classes of transportation-relevant contaminants, including light gases, noncondensable components, and acid-gas impurities [27, 35]. The systems considered range from binary to eight-component mixtures. The PCP-SAFT EoS predicts the phase equilibria and interfacial properties of CO₂-rich mixtures in close agreement with experimental data, yielding bubble- and dew-point pressures comparable to those obtained from the Groupe Européen de Recherches Gazières (GERG-2008) EoS [58]. This agreement supports its use for generating the phase-equilibrium and interfacial-property training data. Using the generated PCP-SAFT EoS+cDFT [70, 103] dataset, we develop data-driven ML models for four target properties: bubble-point pressure, dew-point pressure, IFT, and interfacial thickness. We compare several ML models, including Support Vector Machine (SVM), Random Forest (RF), XGB, and Tabular Prior-data Fitted Network (TabPFN). For all target properties, we use the same input features: temperature, pressure, and mixture composition. These features correspond to the input variables required when planning experiments or molecular simulations. We found that TabPFN with Hyperparameter Optimization (HPO), followed by XGB, achieved the lowest Cross-Validation (CV) Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for all target properties. The results also showed that dew-point pressure and interfacial thickness are less sensitive to impurity concentration. We further assessed the effect of mixture size and compared random, stratified, and GPR-based Active Learning (AL) sampling strategies. AL was the most data-efficient sampling strategy, requiring fewer feed compositions to

achieve a given accuracy. We further develop a hybrid ML model for IFT using the computationally efficient WSD model as a baseline. The ML model is trained on the deviation of the WSD model from the more rigorous PCP-SAFT EoS+cDFT predictions. The ML models used to develop the IFT hybrid models are Symbolic Regression (SR), GPR, and Sparse Variational Gaussian Process (SVGP). This hybrid strategy enables rapid estimation of IFTs for multicomponent CO₂-rich mixtures relevant to CO₂ transportation. After training, the surrogate models can predict interfacial and saturation properties much more rapidly than direct PCP-SAFT EoS+cDFT calculations. This reduction in computational cost is the central motivation of this chapter. Multiphase CFD simulations of CO₂ transport require repeated evaluations of interfacial and saturation properties over a wide range of compositions and thermodynamic states, which become prohibitively expensive when using direct PCP-SAFT EoS+cDFT or MD computations [55, 235]. In sharp contrast, once trained, the ML surrogates can provide these properties at negligible computational cost, making them suitable for repeated property evaluations in large-scale transport process simulations. Here, we combine the PCP-SAFT EoS+cDFT physics engine with two complementary ML surrogate strategies for the IFT of multicomponent CO₂-rich mixtures, while also developing fully data-driven ML models for phase equilibria. The two strategies are a fully data-driven model and a physics-informed WSD-residual model. Earlier ML studies are typically limited to a single property or to simple binary or CO₂-brine systems. In contrast, our framework spans binary to eight-component mixtures and includes seven CO₂ transportation-relevant impurities. It also provides uncertainty-aware predictions and active-learning-based composition sampling.

This chapter is organized as follows: we describe the methodology used to compute the saturation properties (bubble-point and dew-point pressures) and interfacial properties using the PCP-SAFT EoS + cDFT in Section 4.2. The WSD semi-empirical model and ML models used in this chapter are also explained briefly in Section 4.2. In Section 4.3, we present the predictive performance of the ML models for each target property. We also analyze the effects of composition space and mixture size and evaluate the performance of the hybrid residual ML model relative to the semi-empirical WSD model. Finally, Section 4.4 summarizes the main findings of this chapter. Further methodological details and results, as well as the experimental phase equilibrium data, the training datasets, and the trained ML models used in this chapter, are provided in the Supporting Information for this chapter [279] and are archived in an open Zenodo repository (DOI [10.5281/zenodo.21825307](https://doi.org/10.5281/zenodo.21825307)).

4.2. Methodology

4.2.1. Classical Density Functional Theory + PCP-SAFT

The phase equilibria (bubble and dew-point pressures) represented in the PT phase diagram, and the interfacial properties of multicomponent CO_2 mixtures with impurities are computed using the FeO_s (FeO_s) v0.9.5 framework [245]. In FeO_s , SAFT-based EoSs, such as PCP-SAFT and PC-SAFT [68, 102–104], are coupled with cDFT [70, 229, 230, 297] to model inhomogeneous systems, including vapor–liquid interfaces [70, 245]. In cDFT, the vapor–liquid interface is treated as a mesoscopic region in which the density varies smoothly from the bulk vapor phase to the bulk liquid phase. The inhomogeneous system is described using the grand potential functional [70],

$$\Omega(T, \mu, [\rho(\mathbf{r})]) = F(T, [\rho(\mathbf{r})]) - \sum_i \int \rho_i(\mathbf{r}) (\mu_i - V_i^{\text{ext}}(\mathbf{r})) d\mathbf{r}, \quad (4.1)$$

where $\mathbf{r}$ is the position vector in space, $\rho_i(\mathbf{r})$ is the number density profile of component i , $[\rho(\mathbf{r})]$ denotes the set of density profiles of all components in the mixture, $F(T, [\rho(\mathbf{r})])$ is the intrinsic Helmholtz free energy functional, μ_i is the chemical potential of component i , and $V_i^{\text{ext}}(\mathbf{r})$ is an external potential acting on all or some molecules [70]. The equilibrium density profiles are obtained by minimizing the grand potential with respect to the density profile at a constant temperature and chemical potential of each component,

$$\left. \frac{\delta \Omega}{\delta \rho_i(\mathbf{r})} \right|_{T, \mu_i} = 0. \quad (4.2)$$

Applying this equilibrium condition to the grand potential functional in Eq. 4.1 yields the Euler–Lagrange equation [245].

$$\left. \frac{\delta F[\rho(\mathbf{r})]}{\delta \rho_i(\mathbf{r})} \right|_T - \mu_i + V_i^{\text{ext}}(\mathbf{r}) = 0. \quad (4.3)$$

For a homogeneous bulk system, $V_i^{\text{ext}}(\mathbf{r}) = 0$, and the corresponding bulk relation is $F_{\rho_i}^{\text{b}} - \mu_i = 0$, where $F_{\rho_i}^{\text{b}}$ denotes the functional derivative of the Helmholtz free energy functional evaluated for the bulk system. Using this bulk relation, the Euler–Lagrange equation [245] in the absence of an external potential can be written relative to the bulk state as

$$\left. \frac{\delta F[\rho(\mathbf{r})]}{\delta \rho_i(\mathbf{r})} \right|_T - F_{\rho_i}^{\text{b}} = 0. \quad (4.4)$$

The intrinsic Helmholtz free energy functional in the Euler–Lagrange equation is commonly decomposed into ideal and residual contributions, $F[\rho(\mathbf{r})] = F^{\text{ideal}}[\rho(\mathbf{r})] + F^{\text{res}}[\rho(\mathbf{r})]$ where the ideal contribution is known exactly from statistical mechanics, while the residual contribution accounts for intermolecular interactions [70, 71, 130]. In PCP-SAFT-based

cDFT, the residual Helmholtz free energy functional can be decomposed into the same physical contributions as in the bulk PCP-SAFT EoS, namely hard-sphere, hard-chain, and dispersion terms, and, when relevant, association and polar terms [70, 71],

$$F^{\text{res}}[\rho(\mathbf{r})] = F^{\text{hs}}[\rho(\mathbf{r})] + F^{\text{chain}}[\rho(\mathbf{r})] + F^{\text{disp}}[\rho(\mathbf{r})] + F^{\text{assoc}}[\rho(\mathbf{r})] + F^{\text{polar}}[\rho(\mathbf{r})]. \quad (4.5)$$

For further details on the free energy functional forms and cDFT, the reader is referred to Refs. [70, 108, 229, 297]. In this chapter, the equilibrium density profiles are obtained by solving the Euler-Lagrange equation [229, 245] from Eq. 4.4, which combines the different contributions to the free energy functional. Once the equilibrium density profiles are obtained, the IFTs are computed from the excess grand potential per unit interfacial area A as

$$\gamma = \frac{\Omega^{\text{Eq.}} - \Omega^{\text{Bulk}}}{A}. \quad (4.6)$$

where $\Omega^{\text{Eq.}}$ and Ω^{Bulk} are the equilibrium and bulk grand potentials, respectively. The interfacial thickness (L_{10}^{90}), adsorption, and enrichment of components at the interface are computed from the density profiles [211, 271]. The Euler-Lagrange equations are solved numerically in one-dimensional Cartesian coordinates, and the required functional derivatives are evaluated by automatic differentiation [280]. The interfacial domain is discretized with adaptive grid width and number of grid points, with interface widths ranging from 100 to 1000 Å and grid sizes from 1024 to 2048 points, where finer discretization is applied near the critical region. The density profiles are computed using the default cDFT solver settings in FeO_s [245], which uses two sequential Anderson mixing steps [71, 298] with a relative density residual tolerance of 10^{-11} . Although the implementation is computationally efficient in FeO_s [245], the computation of interfacial properties for a full PT phase diagram can still be computationally demanding for multicomponent mixtures with temperature-dependent k_{ij} . This is because the equilibrium density profiles are obtained from coupled nonlinear Euler-Lagrange equations, which must be solved iteratively over the spatial domain for all components in the mixture at each state point [71, 72, 299, 300]. This cost becomes particularly important when repeated IFT computations are required, as in CFD and other process simulations. The pure-component parameters of the PCP-SAFT EoS [68, 102–104] used to describe the bulk phase behavior and interfacial properties of mixtures comprising CO₂, H₂, N₂, Ar, CH₄, O₂, CO, and H₂S are m_i , σ_{ii} , ε_{ii}/k_B , p_i , Q_i , n_A , n_B , κ^{AB} , and $\varepsilon^{\text{AB}}/k_B$. These parameters denote the number of spherical segments per molecule, segment size parameter, dispersion energy parameter, dipole moment, quadrupole moment, number of association sites of type A, number of association sites of type B, association volume, and association energy parameter, respectively. For the studied components, only H₂S was

modeled with an explicit association contribution using a 2A-2B association scheme, with $n_A = 2$ and $n_B = 2$. Hence, cross-association interactions between different molecular species were not considered. The indices ii and jj denote interactions between segments of the same component, whereas interactions between dissimilar atom pairs were modeled using the Lorentz–Berthelot (LB) mixing rules [79, 80] already given in Chapter 3, Eqs. 3.2 and 3.3, where the indices ij refer to cross interactions between atoms of different types and the parameters k_{ij} and l_{ij} are the binary interaction parameters for ε_{ij} and σ_{ij} , respectively, of components i and j . Here, l_{ij} is not considered ($l_{ij} = 0$), and only k_{ij} is considered. To accurately model the multicomponent CO₂-rich mixtures, temperature-dependent BIPs (k_{ij}) were determined for each dissimilar CO₂-impurity pair by fitting to the available experimental P_{xy} data at a temperature where experimental data were available. Here, P_{xy} refers to isothermal vapor–liquid equilibrium data in which the equilibrium pressure, P , is reported together with the corresponding liquid-phase composition, x , and vapor-phase composition, y . The fitted BIPs, together with the pure-component PCP-SAFT EoS [68, 102–104] parameters used in this chapter, are provided in Sections S1 and S2 of the Supporting Information for this chapter [279]. These parameters were used to compute the phase equilibria and interfacial properties of CO₂ mixtures with impurities. In addition, Section S2 of the Supporting Information of this chapter [279] describes the method used to determine the temperature-dependent BIPs and lists the experimental P_{xy} data used for fitting. For an N -component mixture, $N(N - 1)/2$ BIPs are required, so the eight-component mixture considered in this chapter requires 28 binary pairs. Experimental P_{xy} data were available for 23 pairs. For the remaining five pairs, H₂–O₂, CH₄–O₂, O₂–CO, Ar–H₂S, and O₂–H₂S, no experimental P_{xy} data were available in the literature. Hence, k_{ij} was set to zero for these pairs.

4.2.2. Winterfeld–Scriven–Davis Model

A theory for estimating the IFT of multicomponent fluid mixtures was formulated by Winterfeld, Scriven and Davis [146]. This theory extends the Fowler–Kirkwood–Buff framework developed [128, 262] by simplifying the statistical-mechanics expression for the calculation of IFT with the assumption of pairwise-additive intermolecular potentials [128, 146, 262]. The resulting WSD model provides a practical method for estimating IFTs of nonaqueous mixtures using the IFTs of the pure components present in the mixture and can be generalized to an N -component mixture as,

$$\begin{aligned} \gamma_{\text{mix}}(T, P, \mathbf{z}) = & \sum_{\alpha=1}^N \sum_{\beta=1}^N \left[\left(\frac{\rho_{\alpha}^L - \rho_{\alpha}^V}{\rho_{\alpha}^{0,L} - \rho_{\alpha}^{0,V}} \right)^2 \gamma_{\alpha}^0 \delta_{\alpha\beta} \right. \\ & \left. + \Phi_{\alpha\beta} \left(\frac{\rho_{\alpha}^L - \rho_{\alpha}^V}{\rho_{\alpha}^{0,L} - \rho_{\alpha}^{0,V}} \right) \left(\frac{\rho_{\beta}^L - \rho_{\beta}^V}{\rho_{\beta}^{0,L} - \rho_{\beta}^{0,V}} \right) \sqrt{\gamma_{\alpha}^0 \gamma_{\beta}^0} \right], \end{aligned} \quad (4.7)$$

where γ_{mix} denotes the mixture interfacial tension, which depends on temperature T , pressure P , and mole fraction vector $\mathbf{z} = (z_1, z_2, \dots, z_N)$. Here, N is the number of components in the mixture, and the indices α and β denote the components in the system. The double summation in Eq. 4.7 runs over all component pairs, with $\alpha, \beta = 1, 2, \dots, N$. The quantities γ_α^0 and γ_β^0 represent the pure-component IFTs of components α and β , respectively. Similarly, $\rho_\alpha^{0,L}$, $\rho_\alpha^{0,V}$, $\rho_\beta^{0,L}$, and $\rho_\beta^{0,V}$ correspond to the equilibrium liquid and vapor molar densities, in mol/cm³, of the pure components α and β , respectively. The parameters ρ_α^L , ρ_α^V , ρ_β^L , and ρ_β^V denote the liquid and vapor-phase molar densities of components α and β in the mixture [146]. The term $\Phi_{\alpha\beta}$ is the cross-interaction, or mixing, parameter that accounts for nonideal interactions between dissimilar molecules. For nonaqueous and weakly polar fluids, $\Phi_{\alpha\beta}$ is commonly approximated as 1 ($\Phi_{\alpha\beta} \simeq 1$) [146]. The Kronecker delta $\delta_{\alpha\beta}$ equals one when $\alpha = \beta$, corresponding to the pure-component contribution, and zero otherwise [146]. For an N -component mixture, Eq. 4.7 can also be written compactly in matrix form as

$$\gamma_{\text{mix}}(T, P, \mathbf{z}) = \mathbf{a}^\top \Phi \mathbf{a}, \quad (4.8)$$

where the vector $\mathbf{a}$ contains the density-scaled pure-component IFT terms,

$$a_\alpha = \left(\frac{\rho_\alpha^L - \rho_\alpha^V}{\rho_\alpha^{0,L} - \rho_\alpha^{0,V}} \right) \sqrt{\gamma_\alpha^0}, \quad (4.9)$$

and Φ is the symmetric matrix of cross-interaction parameters with diagonal elements $\Phi_{\alpha\alpha} = 1$. The WSD model has been shown to provide reliable IFT predictions for nonaqueous mixtures and therefore serves as a useful semi-empirical framework when molecular-correlation data are unavailable [146, 211, 247]. The formulation becomes less reliable when one or more components in the mixture are supercritical at T . For a three-component mixture, assuming $\Phi_{\alpha\beta} = 1$ for the cross-interaction term and treating component 3 as supercritical, Eq. 4.7 simplifies to

$$\begin{aligned} \gamma_{\text{mix}}(T, P, z_1, z_2, z_3) = & \left(\frac{\rho_1^L - \rho_1^V}{\rho_1^{0,L} - \rho_1^{0,V}} \right)^2 \gamma_1^0 + \left(\frac{\rho_2^L - \rho_2^V}{\rho_2^{0,L} - \rho_2^{0,V}} \right)^2 \gamma_2^0 \\ & + \left(\frac{\rho_1^L - \rho_1^V}{\rho_1^{0,L} - \rho_1^{0,V}} \right) \left(\frac{\rho_2^L - \rho_2^V}{\rho_2^{0,L} - \rho_2^{0,V}} \right) \sqrt{\gamma_1^0 \gamma_2^0}. \end{aligned} \quad (4.10)$$

In this form, component 3 does not contribute directly to the IFT because its pure-component interfacial tension is not defined at supercritical conditions. As a consequence, the WSD model neglects the direct contribution of the supercritical component, which can be physically limiting because component 3 can still influence the mixture phase behavior, density difference, and interfacial properties. In practice, component 3 is expected to lower the IFT of the subcritical mixture by weakening the cohesive forces

between unlike molecules, similar to the action of a surfactant. For the CO₂-rich mixtures with impurities considered in this chapter, all impurity components are supercritical at the investigated conditions, except H₂S. In future work, the Widom line [29, 31] could be explored as a thermodynamic criterion for distinguishing liquid-like and gas-like supercritical states and for estimating the effective contribution of supercritical components to the mixture IFT. To represent this behavior, in Chapter 3 we introduced a correction term [211] for binary mixtures based on the total liquid-phase mole fraction of the supercritical components. Here, we will use the baseline WSD model Eq. 4.7 to compute IFT and then to train the residual-based hybrid ML model with respect to the PCP-SAFT EoS + cDFT [103, 104, 245].

4.2.3. Machine Learning Models

The capabilities of ML models in predicting thermophysical properties of fluid mixtures considered in this chapter have been widely reported in the literature [285, 301–307]. Hence, for brevity, we do not provide a detailed description of each model examined. Instead, we focus on the rationale for selecting the ML models, kernels, and hyperparameters used in this chapter. For the target properties, namely, dew-point pressure, bubble-point pressure, IFT, and interfacial thickness, we use the same set of input features to avoid overfitting and enable a consistent comparison. The selected features are physically well motivated because the feature set contains only the thermodynamic state variables T , P , and $\mathbf{z}$, which are the same variables used to define phase equilibrium and IFT measurements. The supervised ML models considered in this chapter include Support Vector Machine (SVM) [308, 309], Random Forest (RF) [310], eXtreme Gradient Boosting (XGB) [311, 312], and Tabular Prior-data Fitted Network (TabPFN) [305, 306, 313]. SVM uses kernel functions to construct linear or nonlinear models in the feature space. Although it is commonly used for classification, it can also be applied to regression through Support Vector Regression (SVR), which is suitable for continuous thermophysical property prediction [309]. In SVR, the model fits a regression function like linear regression while allowing errors within a tolerance region called the ϵ -insensitive tube. Data points on or outside this tube are the support vectors and mainly determine the fitted model. In this chapter, SVR with a Radial Basis Function (RBF) kernel was used for all target properties. The main hyperparameters are the regularization parameter C , the tube width ϵ , and the kernel coefficient γ_{RBF} [309]. C controls the penalty for errors outside the tube, ϵ defines the error tolerance, and γ_{RBF} controls the length scale of the RBF kernel [309]. RF is an ensemble ML method composed of multiple decision trees [310]. Each tree is trained on a bootstrapped sample of the training dataset, where data points are selected randomly with replacement. At each node, the optimal split is selected from a random subset of input features, reducing correlation over individual trees and improving model

robustness. For regression, the final prediction is obtained by averaging the outputs of all trees. The key hyperparameters of the RF model include the total number of trees, $n_{\text{estimators}}$, the maximum allowable depth of each tree, d_{max} , the minimum sample count required to divide an internal node, n_{split} , the minimum number of samples retained in a terminal leaf, n_{leaf} , and the number of input features randomly evaluated when determining each split, n_{features} [310]. XGB is a gradient-boosted decision-tree algorithm in which trees are added sequentially to correct the errors of the previous model [311, 312]. At each iteration, a new tree is fitted to the negative gradient of the loss function, and its contribution is scaled by a learning rate. The final prediction is obtained by summing the contributions from all trees. Compared with Adaptive Boosting (AdaBoost) [314], XGB uses a more general gradient-based optimization framework and includes regularization terms that help reduce overfitting. Compared with Categorical Boosting (CatBoost) [315], XGB is commonly used for numerical tabular data, whereas CatBoost is particularly designed for datasets with categorical features. The main hyperparameters of XGB besides $n_{\text{estimators}}$ and d_{max} are sample fraction, feature fraction, and regularization parameters [311, 312]. In addition to SVR and decision-tree-based models, this chapter uses the state-of-the-art Tabular Prior-data Fitted Network (TabPFN) deep-learning model [305]. TabPFN is a transformer-based tabular foundation model designed for heterogeneous tabular data with different feature scales, missing values, and mixed data types, where conventional Neural Networks (NNs) often struggle, and tree-based models usually perform strongly [305]. This makes TabPFN suitable for thermodynamic property prediction, where the inputs are structured state variables such as T , P , and $\mathbf{z}$, and the targets are continuous phase-equilibrium and interfacial properties. TabPFN is pre-trained on millions of synthetic tabular datasets and uses In-Context Learning (ICL), the same mechanism used in large language models, to make predictions on new datasets with little or no dataset-specific training [305]. As such, extensive hyperparameter tuning is not strictly required, although HPO using Bayesian optimization can further adapt the model to a specific dataset [305, 313]. In this chapter, the TabPFN HPO was performed using $N_{\text{trials}} = 100$, where each trial sampled a different hyperparameter configuration, fitted the TabPFN model, and evaluated its performance using CV. Additionally, an RF-TabPFN hybrid model is also considered to combine the robustness of tree-based learning with the strong small-data prediction ability of TabPFN [313]. The hyperparameter tuning procedures for SVM, RF, XGB, and TabPFN, together with the corresponding optimal hyperparameter values, are provided in Section S4 of the Supporting Information for this chapter [279]. The SVM and RF models were implemented using the scikit-learn [316] ML library in Python, XGB was implemented using the XGBoost Python package [312], and TabPFN was implemented using the TabPFN-v2.6 Python package [305, 306]. The HPO of TabPFN and RF-TabPFN models were implemented using

the TabPFN-v0.4 extensions package [313]. All the ML models were trained on CPUs, except for the TabPFN-based models, which were run on an NVIDIA H100 GPU.

In addition to data-driven ML models, a residual-learning strategy was used for the IFT. The computationally inexpensive WSD model [146] was taken as the baseline, and ML models were trained to predict the residual between the IFT obtained from the PCP-SAFT EoS + cDFT [245] computation and the WSD prediction [146, 211]. This residual represents the missing correction to the semi-empirical WSD model. Three ML models were considered for this task: Symbolic Regression (SR) [317], Gaussian Process Regression (GPR) [318], and Sparse Variational Gaussian Process (SVGP) [307, 319, 320]. SR [317] was used to examine whether the residual can be represented by a compact and physically interpretable analytical expression, especially for mixtures containing supercritical components where the WSD model may lack T , P , and composition-dependent effects. Following the SR framework of Cranmer [317], SR searches over mathematical expressions generated from selected features, operators, and numerical constants, and identifies expressions that balance predictive accuracy and complexity. The exact GPR was chosen because it is widely used to learn smooth, unknown residual functions while also providing predictive uncertainty [285]. In GPR, the residual function is described by a Gaussian Process (GP) with a mean function $m(\mathbf{x})$ and a covariance kernel $k(\mathbf{x}, \mathbf{x}')$ [285, 318]. Here, $\mathbf{x}$ and $\mathbf{x}'$ denote two input feature vectors, such as thermodynamic states defined by T , P , and composition $\mathbf{z}$. In this chapter, the Matérn-5/2 kernel was used, with the main hyperparameters being the kernel variance σ_f^2 , the length scale ℓ , and the noise variance σ_n^2 [285, 318]. These parameters control the magnitude of the residual variation, the smoothness of the residual function, and the unresolved numerical or modeling noise, respectively, and were optimized by maximizing the Log Marginal Likelihood (LML) [285, 318]. Since the standard or exact GPR requires building and factorizing on the full $N \times N$ kernel matrix, its training time scales approximately as $\mathcal{O}(N^3)$ and its memory requirement as $\mathcal{O}(N^2)$, where N is the number of training points [318]. To improve scalability, Sparse Variational Gaussian Process (SVGP) [307, 319, 320] was also used as an approximate GP model based on M inducing points, where $M \ll N$. The SVGP model used a constant mean function and a Matérn-5/2 kernel with Automatic Relevance Determination (ARD), allowing each input feature to have an individual length scale and thereby providing a measure of feature relevance. A Cholesky variational distribution [319] was used to approximate the posterior distribution over the inducing function values, and the inducing locations were optimized during training. The model was trained using the Evidence Lower Bound (ELBO) [321, 322], which balances a data-fit term with a regularization term based on the Kullback-Leibler (KL) divergence [323] between the variational posterior and the prior [307, 320]. To account for the highly skewed residual distribution, 2000 inducing

points were initialized through stratified sampling. This sampling covered the *PT* region where most residuals are concentrated, as well as the mid-tail and extreme-tail residual regions. The target residuals were transformed using a quantile transformation followed by standardization, and the ELBO was maximized using mini-batch Adam optimization with a batch size of 1024, tail-weighted sampling, over 500 epochs, with the learning rate cosine-annealed from 10^{-2} to 10^{-4} . The SVGP model was implemented using the GPyTorch package [324], while the exact GPR model was implemented using the scikit-learn library in Python [316]. For SR, the symbolic-regression framework-v1.5 of Cranmer [317] was used in Julia.

4

4.2.4. Composition Space Sampling

To compute the saturation pressures, namely, bubble and dew pressures, IFT, and interfacial thickness for compositions relevant to CO₂ transportation, a pool of mixture compositions was generated. Since transported CO₂ streams typically contain high CO₂ mole fractions, the composition space was mainly focused on mixtures with z_{CO_2} in the range of 95–99.99%, while the remaining fraction was distributed between the impurity components. To distribute the impurity concentration, we used Dirichlet sampling [325], where the CO₂ mole fraction was fixed initially, and the remaining fraction was treated as the impurity budget, $s = 1 - z_{\text{CO}_2}$. The Dirichlet distribution is defined on a simplex, which is the geometric space of nonnegative composition vectors whose components sum to unity. As a result, for k impurity components, the sampled normalized impurity vector lies on a $(k - 1)$ -dimensional simplex [325]. In this chapter, with seven impurity components considered, the Dirichlet distribution samples points on a six-dimensional impurity simplex. A normalized impurity composition vector was then sampled from a Dirichlet distribution [325],

$$(\tilde{z}_1, \dots, \tilde{z}_k) \sim \text{Dir}(\alpha_1, \dots, \alpha_k), \quad \tilde{z}_i \geq 0, \quad \sum_{i=1}^k \tilde{z}_i = 1, \quad (4.11)$$

where k is the number of impurity components. The impurity mole fractions were obtained by scaling this normalized vector with the impurity budget,

$$z_i = s\tilde{z}_i = (1 - z_{\text{CO}_2})\tilde{z}_i, \quad i = 1, \dots, k. \quad (4.12)$$

Thus, the final composition of a mixture satisfies

$$z_{\text{CO}_2} + \sum_{i=1}^k z_i = z_{\text{CO}_2} + (1 - z_{\text{CO}_2}) \sum_{i=1}^k \tilde{z}_i = z_{\text{CO}_2} + (1 - z_{\text{CO}_2}) = 1. \quad (4.13)$$

This procedure ensures that all impurity mole fractions are nonnegative, the impurity fractions sum exactly to the prescribed remaining fraction, and the CO₂ mole fraction remains fixed at the selected level. Apart

from Dirichlet sampling [325], systematic grid scaling was used to ensure coverage of highly impure CO₂ conditions, corresponding to $z_{\text{CO}_2} = 95\text{--}96\%$ with an impurity combination (binary mixtures). The final composition pool was generated using 70% Dirichlet sampling, 20% systematic grid scaling, and 10% compositions from industrial specifications and reported CO₂ transport streams [27]. The industrial compositions were taken from the review article of Simonsen, Hansen and Pedersen [27], including quality specifications from transportation projects such as Northern Lights [17] and Porthos [19], capture technologies such as post-combustion, pre-combustion, oxy-fuel combustion, and operational CO₂ pipelines in the USA and Canada [27]. Trace impurities reported in parts per million, such as SO_x, NO_x, and NH₃, were neglected, and their fractions were assigned to CO₂. Since seven impurities were considered in this chapter, the possible combinations of impurity types and their concentrations are numerous. To address this, three sampling strategies, namely, random sampling, stratified sampling, and Active Learning (AL), were used to circumvent the computational cost of evaluating every mixture in the composition space and to identify the minimum number of mixtures required for reliable ML model performance. For random sampling, a target set of $\mathcal{S}$ mixture compositions was selected from the full composition pool of size $\mathcal{P}$ using a fixed random seed. The pool $\mathcal{P}$ contained compositions generated from Dirichlet sampling, systematic grid scaling, and industrially reported feed specifications. For stratified sampling, the CO₂ mole fraction range was divided into five bins. Within each CO₂ bin, the impurity compositions were further partitioned into five clusters using k -means clustering based on Euclidean distance [326]. The number of impurity clusters was selected based on the elbow method and the silhouette score [326–328]. This procedure resulted in 25 strata, combining five CO₂ composition bins with five impurity composition clusters. The stratified sampling algorithm and its detailed implementation are provided in Section S3 of the Supporting Information for this chapter [279]. In addition to random and stratified sampling, a kernel-based AL scheme was adopted to reduce the computational cost associated with evaluating a large number of candidate mixtures in the composition pool $\mathcal{P}$. The AL approach used in this chapter is shown in Algorithm 1. This approach was adapted from the work of Xiang *et al.* [329], where it was originally applied to chemical compound-space exploration of alkanes. Here, the algorithm was modified for composition space sampling. A GP regressor with an ARD Matérn-5/2 kernel, signal variance σ_f^2 , and additive noise variance σ_n^2 was first fitted to the pre-existing P_{bubble} labeled set $\mathcal{T}_{\text{lab}}$ using standardized composition vectors $\tilde{\mathbf{x}}$. The vector $\tilde{\mathbf{x}}$ denotes the scaled form of the previously generated composition vector, obtained by mean-centering each composition variable and normalizing it by its standard deviation before fitting the GPR model [318]. Xiang *et al.* [329] similarly used the previously trained GPR model [330] on MD simulations and the National Institute of Standards and Technology

Algorithm 1 CO₂-transportation composition space Active Learning (AL) based on Automatic Relevance Determination (ARD)-kernel uncertainty. A Gaussian Process Regression (GPR) model with an ARD Matérn-5/2 kernel is fitted to the labeled bubble-point pressure set, $\mathcal{T}_{\text{lab}}$, and the optimal hyperparameters, θ^* , are obtained by maximizing the Log Marginal Likelihood (LML).

Require: $\mathcal{P}$: candidate composition pool; $\mathcal{T}_{\text{lab}}$: P_{bubble} labels; N : target size.

Ensure: $\mathcal{S}$: selected composition set of target size N .

$\mathcal{S} \leftarrow \{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}\}, \quad \mathcal{P}' \leftarrow \mathcal{P} \setminus \mathcal{S}$

Fit GPR using $k(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = \sigma_f^2 M_{5/2}(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j; \ell) + \sigma_n^2 \delta_{ij}$

$\theta^* \leftarrow \arg \max_{\theta} \log p(\mathbf{y} | \tilde{\mathbf{X}}, \theta), \quad \theta^* \leftarrow \{\ell^*, \sigma_f^{2*}, \sigma_n^{2*}\}$

$k_s(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) \leftarrow \sigma_f^{2*} M_{5/2}(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j; \ell^*)$

while $|\mathcal{S}| < N$ and $\mathcal{P}' \neq \emptyset$ **do**

$U(\mathbf{x}_i) \leftarrow k_s(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i) - \mathbf{k}_{\mathcal{S}}(\tilde{\mathbf{x}}_i)^T \mathbf{K}_{\mathcal{S}}^{-1} \mathbf{k}_{\mathcal{S}}(\tilde{\mathbf{x}}_i), \quad \forall \mathbf{x}_i \in \mathcal{P}'$

$\mathbf{x}^* \leftarrow \arg \max_{\mathbf{x}_i \in \mathcal{P}'} U(\mathbf{x}_i)$

$\mathcal{S} \leftarrow \mathcal{S} \cup \{\mathbf{x}^*\}, \quad \mathcal{P}' \leftarrow \mathcal{P}' \setminus \{\mathbf{x}^*\}$

$\mathcal{P}' \leftarrow \mathcal{P}' \setminus \{\mathbf{x}_i \in \mathcal{P}' : U(\mathbf{x}_i) < u_t\}$

end while

return $\mathcal{S}$

(NIST) Standard Reference Database [331] with the marginalized graph kernel. However, in this chapter, we generated a composition pool of size 100 and computed the P_{bubble} curve using the PCP-SAFT EoS [103, 104] to train the GPR [318] model and determine its hyperparameters. The kernel hyperparameters were obtained by maximizing the LML, giving $\theta^* = \{\ell^*, \sigma_f^{2*}, \sigma_n^{2*}\}$. The optimized component-wise length scales ℓ_k^* provide a measure of the relevance of each species to P_{bubble} . The fitted noise term was then discarded, retaining only the signal kernel $k_s(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = \sigma_f^{2*} M_{5/2}(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j; \ell^*)$, which defines an ARD-weighted metric in the composition space. Starting from two randomly selected compositions, the selected set $\mathcal{S}$ was grown by adding the candidate with the largest posterior variance (uncertainty measure), defined as $U(\mathbf{x}_i) = k_s(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_i) - \mathbf{k}_{\mathcal{S}}(\tilde{\mathbf{x}}_i)^T \mathbf{K}_{\mathcal{S}}^{-1} \mathbf{k}_{\mathcal{S}}(\tilde{\mathbf{x}}_i)$. This uncertainty depends on the position of each candidate mixture relative to the compositions already selected in $\mathcal{S}$ and does not require additional PCP-SAFT EoS [103, 104] evaluations during the selection step. For the Matérn-5/2 kernel, this procedure corresponds to a sequential maximum-variance design that selects new points from regions of composition space least covered by the current selected set. After each acquisition, candidates with $U < u_t$ were removed from the remaining pool $\mathcal{P}'$ because they were

Table 4.1. Temperature, pressure, and composition ranges of the dataset used to train the ML models, generated using the PCP-SAFT EoS + cDFT framework. The dataset comprises 19361 state points (ca. 1.9×10^4) evaluated for 100 training feeds sampled from the composition pool with sampling fractions $(f_{\text{Dirichlet}}, f_{\text{systematic}}, f_{\text{industrial}}) = (0.70, 0.20, 0.10)$. For each feed, state points were generated at temperatures from 200 to 300 K in steps of 5 K, with ten pressures spaced equally between the dew-point and bubble-point pressures at each temperature. The impurity mole fractions have a minimum of zero because not every impurity is present in every mixture. Predictions outside the reported ranges constitute extrapolation, for which the reported error metrics do not apply.

Variable	Minimum	Maximum
T / [K]	200	300
P / [bar]	2.62	162.88
z_{CO_2} / [mol mol ⁻¹]	0.95	0.99
z_{H_2} / [mol mol ⁻¹]	0	0.0313
z_{Ar} / [mol mol ⁻¹]	0	0.05
z_{N_2} / [mol mol ⁻¹]	0	0.04
z_{CH_4} / [mol mol ⁻¹]	0	0.05
z_{O_2} / [mol mol ⁻¹]	0	0.04
z_{CO} / [mol mol ⁻¹]	0	0.03
$z_{\text{H}_2\text{S}}$ / [mol mol ⁻¹]	0	0.03

considered informationally redundant. In contrast to Xiang *et al.* [329], who used a fixed uncertainty threshold u_t as the stopping criterion, u_t was adjusted in this chapter according to the target sample size N , so that the AL loop terminates when the selected set reaches $|\mathcal{S}| = N$. The candidate pool $\mathcal{P}$ consisted of 1000 multicomponent CO₂ mixtures. An initial subset, denoted as $\mathcal{P}_{\text{init}}$, containing 100 multicomponent CO₂ mixtures was generated from this pool and labeled by computing P_{bubble} using PCP-SAFT EoS [103, 104]. These labeled compositions were then used to pre-train the GPR model used in the acquisition step. The same candidate pool $\mathcal{P}$ was used for AL, random sampling, and stratified sampling. Each sampling strategy was repeated for 20 independent trials. The reported metrics correspond to the trial-averaged values, and the uncertainty is quantified using the 95% confidence interval for the independent trials. The temperature, pressure, and composition ranges spanned by the resulting training dataset are summarized in Table 4.1. Predictions outside these ranges constitute extrapolation, for which the reported error metrics do not apply.

4.2.5. Model Evaluation

To evaluate the quantitative performance of the regression ML models, we used the following metrics: the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE). These metrics are defined as

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (4.14)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (4.15)$$

and

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|. \quad (4.16)$$

Here, N is the total number of data points, i indexes each data point, y_i is the reference value of the i -th data point, $\hat{y}_i$ is the corresponding predicted value from the ML model, and $\bar{y}$ is the mean of the reference values. The coefficient of determination (R^2) measures how well the predicted values from an ML model reproduce the variation in the reference data. Values of R^2 closer to unity indicate better predictive performance. The RMSE measures the average prediction error in the same physical units as the target property. Because the errors are squared before averaging, larger deviations are penalized more strongly. The MAE measures the average absolute prediction error and is less affected by large individual deviations than the RMSE. To compare prediction errors of target properties with different physical units and numerical ranges, the Normalized Root Mean Squared Error (NRMSE) and Normalized Mean Absolute Error (NMAE) were also evaluated. These metrics are defined as

$$\text{NRMSE} = \frac{\text{RMSE}}{y_{\max} - y_{\min}}, \quad (4.17)$$

$$\text{NMAE} = \frac{\text{MAE}}{y_{\max} - y_{\min}}. \quad (4.18)$$

where $y_{\max}$ and $y_{\min}$ denote the maximum and minimum reference values of the corresponding target property, respectively. To ensure consistency, the same composition sets were used to compute P_{bubble} , P_{dew} , IFT, interfacial thickness, and the WSD residual using PCP-SAFT EoS + cDFT. The datasets were split into training, testing, and validation sets with a 0.70:0.15:0.15 ratio, and the same split was used for all ML models for each target property. To evaluate model robustness, a five-fold Cross-Validation (CV) was performed on the training set. In five-fold CV, the training set was divided into five subsets. For each CV run, four subsets were used to train the model, while the remaining subset was used for validation. This process

was repeated five times so that each subset served once as the validation set. The resulting fold-wise metrics were then used to assess model robustness and identify possible overfitting or underfitting of the ML model.

4.3. Results and Discussion

This section evaluates the performance of Machine Learning (ML) models in predicting the *PT* phase envelope, including bubble- and dew-point pressures, and the interfacial properties (γ and L_{10}^{90}) of multicomponent CO₂-rich mixtures. The reference phase envelope and interfacial properties were computed using the Perturbed-Chain Polar Statistical Associating Fluid Theory (PCP-SAFT) Equation of State (EoS) + classical Density Functional Theory (cDFT) [68, 103, 245] of the CO₂-rich mixtures. Section 4.3.1 describes the composition pool of CO₂-rich mixtures considered in this chapter. The effect of impurities on P_{bubble} , P_{dew} , γ , and L_{10}^{90} are presented and discussed in Section 4.3.2. Section 4.3.3 evaluates the relationships between the target properties, P_{bubble} , P_{dew} , γ , and L_{10}^{90} and the input features, T , P , and $\mathbf{z}$ using the Spearman rank correlation [332] and the Kolmogorov–Smirnov (KS) [333] statistical metrics. Two supervised ML-based modeling strategies are considered in this chapter. For the full data-driven ML approach, Random Forest (RF) [310], Support Vector Regression (SVR) [308] with a Radial Basis Function (RBF) kernel, eXtreme Gradient Boosting (XGB) [311, 312], and Tabular Prior-data Fitted Network (TabPFN) [306] were used. The performance of these models for P_{bubble} , P_{dew} , γ , and L_{10}^{90} is presented in Section 4.3.4. In the full data-driven models, the thermodynamic state variables T and P , together with the feed composition vector $\mathbf{z}$, were used as input features. The effects of the number of compositions and the sampling strategy on the predictive accuracy of the aforementioned ML models are evaluated in Section 4.3.5, using random, stratified, and Active Learning (AL) sampling methods. In addition to the full data-driven models, Section 4.3.6 discusses hybrid residual ML models for improving the Interfacial Tension (IFT) predicted by the semi-empirical Winterfeld–Scriven–Davis (WSD) correlation [146]. In this residual-learning approach, Symbolic Regression (SR) [317], Gaussian Process Regression (GPR) [318], and Sparse Variational Gaussian Process (SVGP) [307, 320, 324] were trained on the difference between the WSD prediction and the PCP-SAFT EoS + cDFT [68, 103, 245] reference IFT.

4.3.1. Composition Pool

In this chapter, multicomponent CO₂-rich mixtures were generated with the constraint $z_{\text{CO}_2} \geq 95\%$ to represent compositions relevant to CO₂ transportation. The remaining impurity fraction, $z_{\text{impurity}} \leq 5\%$, was distributed between the impurity components using two sampling approaches: the Dirichlet distribution and systematic scaling. For each generated mixture, z_{CO_2} was then determined from the total impurity fraction to satisfy the

composition closure condition. In addition, industrial feed compositions reported by Simonsen, Hansen and Pedersen [27] were included to introduce realistic CO₂ transport mixtures into the composition pool. We considered a pool of 1000 CO₂-rich compositions. Fig. 4.1a shows the CO₂ mole fraction distribution for the three composition sources: Dirichlet distribution, systematic scaling, and industrial feeds. The individual z_{CO_2} compositions are plotted as markers, and the outliers are not shown separately. The middle 50% of the data, represented by the Interquartile Range (IQR), lies between approximately $z_{\text{CO}_2} = 0.96$ and 0.98 for the Dirichlet distribution, between $z_{\text{CO}_2} = 0.95$ and 0.97 for systematic scaling, and between $z_{\text{CO}_2} = 0.96$ and 1.00 for the industrial feeds. The CO₂ mole fractions from the Dirichlet distribution and systematic scaling appear at fixed values, such as 0.96 and 0.97 , because they were generated with two-decimal precision. This discretization was intentionally introduced so that the ML models could learn and interpolate thermodynamic properties for different feed composition levels. In contrast, the industrial feed compositions have more continuous z_{CO_2} values because they come from reported feed data. Fig. 4.1b shows the z_{impurity} distribution in a box plot with outliers (markers) for each impurity component. The majority of z_{impurity} is < 0.03 , since the other impurity component takes the remaining fraction, which is more likely in multicomponent systems. The distribution of z_{impurity} with respect to the composition source is presented in Fig. 4.1c. The Dirichlet-sampling compositions mainly have their IQR in the low-impurity region, whereas the systematic-scaling compositions have their IQR at higher impurity fractions, mostly for $z_{\text{impurity}} > 0.01$. For the industrial feeds, the major impurities, such as H₂, Ar, N₂, and CH₄, have a wide IQR range compared with minor impurities such as O₂, CO, and H₂S. This combined strategy ensures that the impurity composition space is sampled efficiently while keeping the mixtures representative of CO₂-rich transport streams. For the generated composition pool of size $|\mathcal{P}| = 1000$, Fig. 4.1d presents the empirical distribution of mixture sizes, where the mixture size denotes the number of active components in each composition. The bars represent the percent-normalized relative frequency, defined as $\hat{f}_i^N = (n_i/N) \times 100\%$, with $\sum_i \hat{f}_i^N = 100\%$, where n_i is the number of compositions containing exactly i components and $N = |\mathcal{P}|$. The solid red line represents the Kernel Density Estimation (KDE) of the same sample, plotted on the same percent-normalized axis, $\hat{f}^N$. Since the mixture size is integer-valued with unit bin width, the bar heights and KDE [334, 335] values are directly comparable. The distribution is approximately bell-shaped and centered between four and five components. Eight-component mixtures originate mainly from the industrial feed composition source, whereas the Dirichlet sampling, due to the impurity-weighting parameter $\mathbf{a}$, favors lower-cardinality mixtures and rarely activates all seven impurities at the same time. Consequently, the shape of $\hat{f}_i^N$ can be controlled through the sampling fractions $(f_{\text{Dirichlet}}, f_{\text{systematic}}, f_{\text{industrial}}) = (0.70, 0.20, 0.10)$,

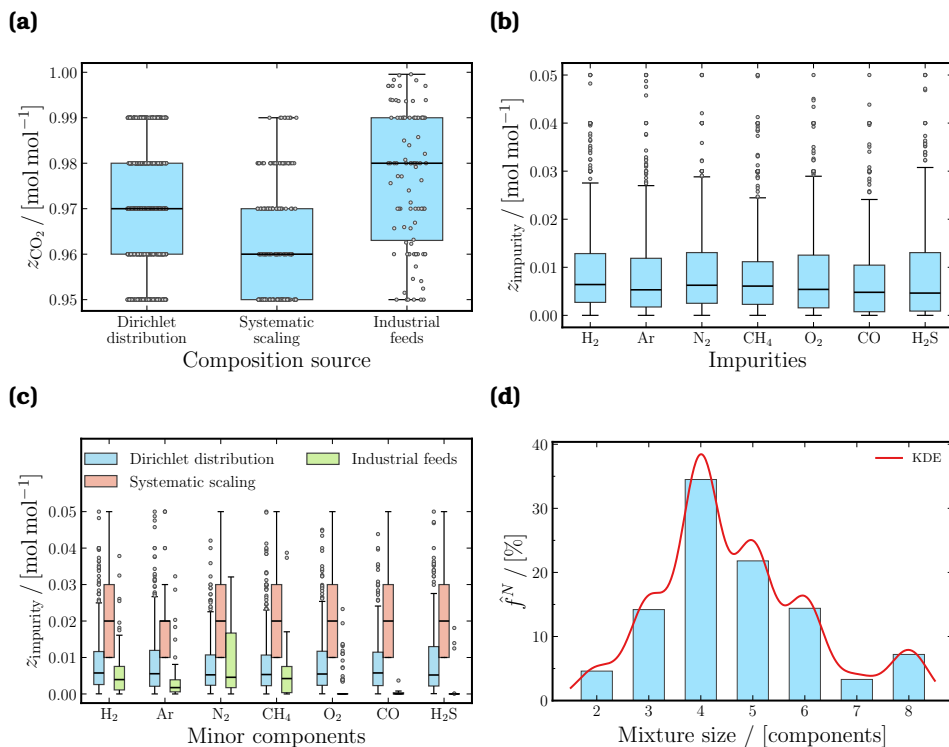


Figure 4.1. Composition-pool characteristics for the CO₂-rich mixtures considered in this chapter for computing the phase envelope and interfacial properties. The composition pool $\mathcal{P}$ comprises $N = 1000$ compositions. (a) z_{CO_2} distribution grouped by each composition source. The markers represent the individual z_{CO_2} compositions from each feed obtained using the Dirichlet distribution [325], systematic scaling, and industrial feeds [27]. (b) Box plots of the impurity mole fractions, z_{impurity} , distribution for H₂, Ar, N₂, CH₄, O₂, CO, and H₂S in the composition pool. (c) Comparison of z_{impurity} distributions separated by composition source, illustrating the different z_{impurity} ranges generated by the Dirichlet distribution [325], systematic scaling, and industrial feeds [27]. (d) Normalized histogram of the mixture-size distribution. Bars give the percent-normalized relative frequency $\hat{f}_i^N$, and the solid red line shows the Kernel Density Estimation (KDE). The pool spans binary to eight-component mixtures, with the upper tail contributed by the industrial feed sampling fraction.

allowing the composition pool to be biased toward low, medium, or high cardinality mixture sets as required. The temperature, pressure, and composition ranges of the training dataset generated from this pool are summarized in Table 4.1.

4.3.2. Effect of Impurities

Fig. 4.2a-c present the phase envelopes, Interfacial Tension (IFT), and interfacial thickness of selected CO₂-rich mixtures computed using the PCP-SAFT EoS [102, 104, 245]. The corresponding feed compositions are listed in Table 4.2. These five feeds were selected from the composition pool, $\mathcal{P}$, to include different numbers of active components present in CO₂-rich mixtures and to capture distinct P_{bubble} behavior. The phase envelopes in Fig. 4.2a are compared with the pure-CO₂ saturation line. The dew-point curves of the CO₂-rich mixtures are nearly identical. Compared to the pure CO₂ saturation line, noticeable differences near the critical region were identified. Compared to dew-point curves, bubble-point curves are much more sensitive to impurity type and its composition. For the investigated composition range, most impurities broaden the phase envelope of CO₂-rich mixtures, while H₂S tends to narrow it. For CO₂ transportation, this broadening increases the range of temperatures and pressures over which vapor-liquid phase splitting can occur, so higher operating pressures may be required to maintain single-phase dense CO₂ transport. This behavior is mainly related to the impurity critical temperature and volatility relative to CO₂. Low critical temperature impurities such as H₂, N₂, Ar, and CH₄ are more volatile than CO₂ and increase the separation between the bubble- and dew-point curves. In contrast, H₂S has a higher critical temperature and stronger condensability than CO₂, and therefore behaves more like a heavy component. The broadening effect is especially clear for mixtures containing H₂. For example, feed 1, a binary mixture without H₂, has a narrow phase envelope, whereas feed 2, a quaternary mixture containing H₂, has a much wider phase envelope. For feeds 2 and 4, P_{bubble} is higher at lower temperatures than at higher temperatures, which can be attributed to the negative Joule-Thomson behavior of H₂ [336]. Fig. 4.2b presents IFTs of the same selected mixtures computed at the P_{bubble} and P_{dew} curves and compared with pure CO₂. The IFT values along the dew curves remain nearly identical to the pure-CO₂ Interfacial Tension (IFT), almost independent of impurity type and its composition. IFT values at bubble curves show a stronger dependence on the feed composition and impurity. Compared with pure CO₂, all selected CO₂-rich mixtures have lower IFT on the bubble branch, with Mean Relative Deviation (MRD) values ranging from ca. -14% to ca. -26%. Feed 3 shows the lowest average deviation of ca. -14%, whereas feed 4 exhibits the largest average decrease of ca. -26%. The largest deviations are noticed at the high temperatures, around $T = 298\text{--}300$ K. Fig. 4.2c presents the corresponding interfacial thickness, L_{10}^{90} , computed at the P_{bubble} and P_{dew} curves. Similar to the IFT, the L_{10}^{90} values of the CO₂-rich mixtures along the dew branch are almost identical to those of pure CO₂, indicating that the impurities have only a minor effect on the dew branch interfacial properties. In sharp contrast, the mixtures have larger L_{10}^{90} values than pure CO₂ along the bubble branch, with MRD values ranging from approximately 10% to 20%. Feed 3 shows

Table 4.2. Mixture compositions z_i in mole fractions used for the sensitivity analysis shown in Fig. 4.2. The feeds considered for the sensitivity analysis were selected from a composition pool of size 1000. The selected feeds span mixture sizes from 2 to 8 components.

Feed	N_c	z_{CO_2}	z_{H_2}	z_{Ar}	z_{N_2}	z_{CH_4}	z_{O_2}	z_{CO}	z_{H_2S}
1	2	0.9700	-	-	-	-	-	0.0300	-
2	4	0.9500	0.0313	-	-	0.0039	-	-	0.0148
3	6	0.9800	0.0058	0.0091	0.0011	0.0029	0.0011	-	-
4	7	0.9600	0.0200	-	0.0080	0.0040	0.0040	0.0020	0.0020
5	8	0.9700	0.0075	0.0040	0.0079	0.0099	3.79×10^{-5}	6.99×10^{-4}	3.90×10^{-7}

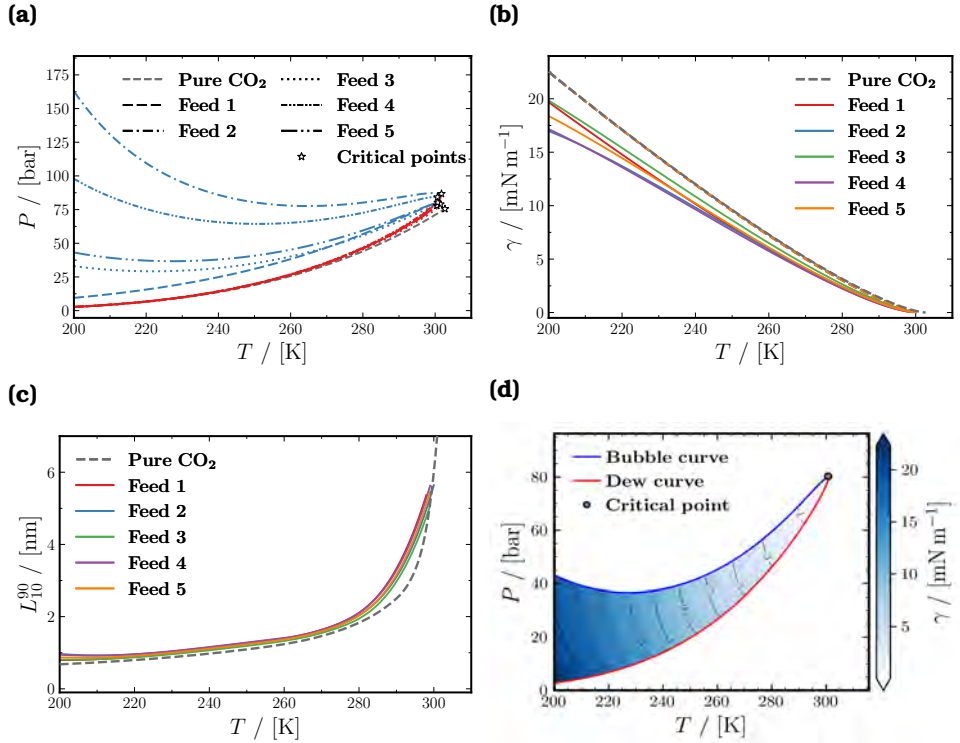


Figure 4.2. Sensitivity analysis of CO₂-rich mixtures for feeds 1–5, whose compositions are listed in Table 4.2. (a) Bubble-point (blue) and dew-point (red) pressure curves of each feed, distinguished by line style, compared with the pure-CO₂ saturation line, with critical points. (b) Interfacial tension, γ , along the bubble curve (solid lines) and the dew curve (dashed lines), compared with pure CO₂. (c) Interfacial thickness, L_{10}^{90} , along the bubble curve, compared with pure CO₂. (d) Two-phase envelope of feed 4, including the bubble curve, dew curve, critical point, and the contour distribution of γ within the two-phase region. The phase envelope and interfacial properties were computed using the PCP-SAFT EoS coupled with cDFT using FeO_s [68, 102, 103, 245].

the smallest deviation, with an MRD of ca. 10%, whereas feed 4 exhibits the largest deviation, with an MRD of ca. 20%. However, the differences in L_{10}^{90} between the CO₂-rich mixtures listed in Table 4.2 are small compared with their deviations from pure CO₂. Overall, impurities lower the IFT and increase the interfacial thickness, indicating a more diffuse interface than in pure CO₂. From a transportation perspective, these changes can affect two-phase flow behavior once phase splitting occurs, because lower IFT facilitates phase dispersion and increases the interfacial surface

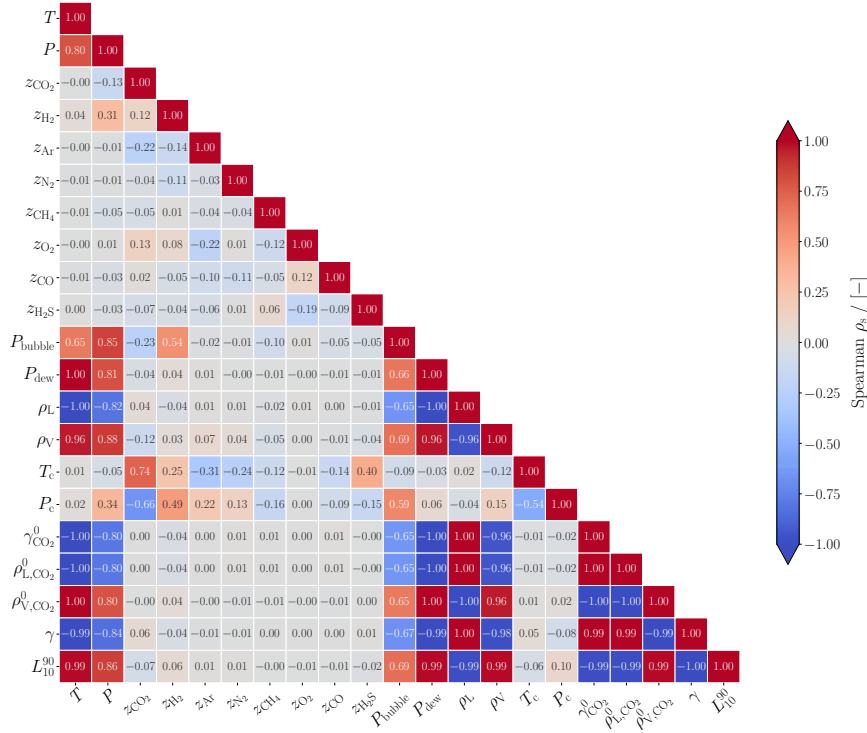
area available for heat and mass transfer. The phase envelopes and interfacial properties were computed for several CO₂-rich mixtures from the composition pool, $\mathcal{P}$, similar to the five representative feeds listed in Table 4.2. The phase envelopes and interfacial properties data computed using the PCP-SAFT EoS+cDFT were then used for training ML models. For each feed composition, a two-phase contour map was generated to check the consistency of the PCP-SAFT EoS+cDFT computations, as illustrated in Fig. 4.2d for feed 4. The contour map indicates that the IFT is strongly a function of T . At constant T , the IFT changes almost linearly with pressure, indicating that its pressure dependence along an isotherm can be approximated using a simple linear relation.

4.3.3. Correlation Analysis of Features

To guide the input feature selection for the ML models, the feature–target relationships were evaluated using the Spearman correlation matrix [332] and the Kolmogorov–Smirnov (KS) statistic [333], as shown in Fig. 4.3. These analyses were performed on a composition pool with $|\mathcal{P}| = 100$, resulting in 19361 thermodynamic state points generated using the PCP-SAFT EoS + cDFT framework [68, 103, 245]. The Spearman correlation matrix in Fig. 4.3a was used to identify monotonic nonlinear feature–target relationships and was preferred over Pearson correlation because the investigated phase-equilibrium and interfacial properties show nonlinear behavior, as illustrated in Fig. 4.2a–c. However, a weak Spearman correlation does not necessarily imply feature irrelevance because nonmonotonic relationships can still give a ρ_s value close to zero. For this reason, the KS statistic in Fig. 4.3b was used as a complementary direction-free relevance measure by comparing the target property distributions between low- and high-feature-value groups. A high KS value with low ρ_s therefore indicates an important but potentially nonmonotonic feature–target relationship. This combination is useful here because all target properties show nonlinear behavior. For example, P_{bubble} can show a strongly nonlinear dependence on T for mixtures with higher H₂ content, where the curve shape is affected by Joule–Thomson inversion behavior [336]. The P_{dew} curves, which closely follow the pure-CO₂ saturation line, can be related to an Antoine-type relation, $\ln P \sim -A/T + B$ [270]. Similarly, the IFT decreases nonlinearly with temperature, consistent with semi-empirical correlations of the form $\gamma = \gamma_0(1 - T/T_c)^n$ [219], where $n \neq 1$. The interfacial thickness also shows critical behavior and can be represented qualitatively as $L \sim (1 - T/T_c)^{-\nu}$ [232, 263], where the positive exponent ν causes L to increase sharply as T approaches T_c . Overall, using both Spearman correlation and the KS statistic provides a more useful feature-relevance analysis than relying only on a linear correlation measure.

The KS analysis confirms the importance of the thermodynamic state variables, particularly T and P . Pressure shows relatively high KS values of ca. 0.7 for P_{bubble} , P_{dew} , γ , and L_{10}^{90} , while temperature reaches values

(a)



(b)

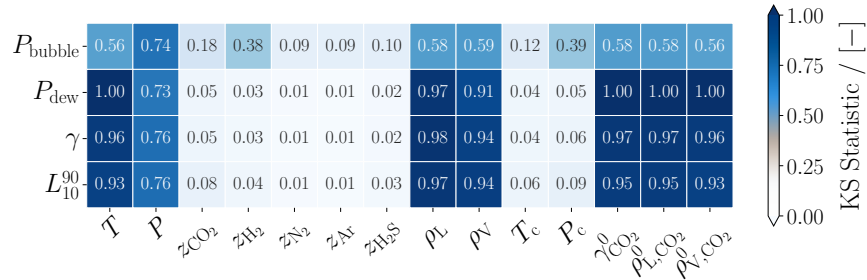


Figure 4.3. Feature relevance analysis for the CO₂-rich composition pool $\mathcal{P} = 100$. (a) Spearman rank-correlation matrix [332] between the input features (T , P , $\mathbf{z}$, ρ_L , ρ_V , T_c , P_c , $\gamma_{\text{CO}_2}^0$, ρ_{L,CO_2}^0 , and ρ_{V,CO_2}^0) and the target properties P_{bubble} , P_{dew} , γ , and L_{10}^{90} . (b) Kolmogorov-Smirnov (KS) statistic [333] between each input feature and target property. The Spearman coefficient, ρ_s , identifies monotonic feature-target relationships, while the KS statistic quantifies distributional differences and provides an additional measure of feature sensitivity.

close to one for P_{dew} , γ , and L_{10}^{90} . The phase densities, ρ^L and ρ^V , also show high KS values of ca. 0.9 for P_{dew} , γ , and L_{10}^{90} . The Spearman matrix supports the same physical interpretation, where P_{dew} , γ , and L_{10}^{90} are strongly related to T , ρ^L , and ρ^V , whereas P_{bubble} is more sensitive to the feed composition, particularly z_{H_2} and z_{CO_2} . Based on this feature-relevance analysis, the full data-driven models were trained using T , P , and the feed composition vector $\mathbf{z}$ as the final input features, defined as $\mathcal{X} = \{T, P, \mathbf{z}\}$. Since the feed mole fractions satisfy $\sum_i z_i = 1$, only $N_c - 1$ composition variables are independent, giving a scalar feature set cardinality of $|\mathcal{X}_{\text{scalar}}| = 2 + (N_c - 1)$. The composition vector was kept explicit to ensure that the models remain direct functions of mixture composition and to allow the effects of individual impurities to be evaluated. The phase mole fraction vectors, $\mathbf{x}$ and $\mathbf{y}$, pure- CO_2 properties, and mixture critical properties were excluded because they largely repeat information already contained in T , P , and $\mathbf{z}$. This avoids multicollinearity and unnecessary input dimensions while keeping the model physically interpretable and composition-dependent.

4.3.4. Machine Learning Model Performance and Outlier Analysis

The dataset for the Machine Learning (ML) regression models was generated using the PCP-SAFT EoS + cDFT framework [68, 104, 245]. To ensure broad coverage of the composition space, a pool of 100 feed compositions was constructed, resulting in approximately 1.9×10^4 data points. The feed pools combined Dirichlet sampling, systematic scaling, and industrial feed compositions, with relative contributions of 0.7, 0.2, and 0.1, respectively. The predictive performance of the full data-driven models was evaluated using parity plots and prediction-error plots. The parity plots are presented in Fig. S11 of the Supporting Information for this chapter [279] for clarity, whereas the prediction-error plots are presented in Figs. 4.4-4.7. These plots compare the predicted and reference values for P_{bubble} , P_{dew} , γ , and L_{10}^{90} using Random Forest (RF), Support Vector Regression (SVR), eXtreme Gradient Boosting (XGB), and Tabular Prior-data Fitted Network (TabPFN). The parity plots report the R^2 values, while Figs. 4.4-4.7 report the corresponding test and validation RMSE values using consistent axis ranges for each target property. The five-fold Cross-Validation (CV) metrics, including the mean and uncertainty of R^2 , RMSE, and Mean Absolute Error (MAE), are reported in Table 4.3. Although all models achieved high R^2 values in the single train-test-validation split, the error magnitudes, outlier behavior, and CV robustness varied widely. Outliers were identified using a three-standard-deviation criterion (3σ), where a data point was flagged when its absolute residual exceeded three times the residual standard deviation. Since the 3σ threshold was computed separately for each model,

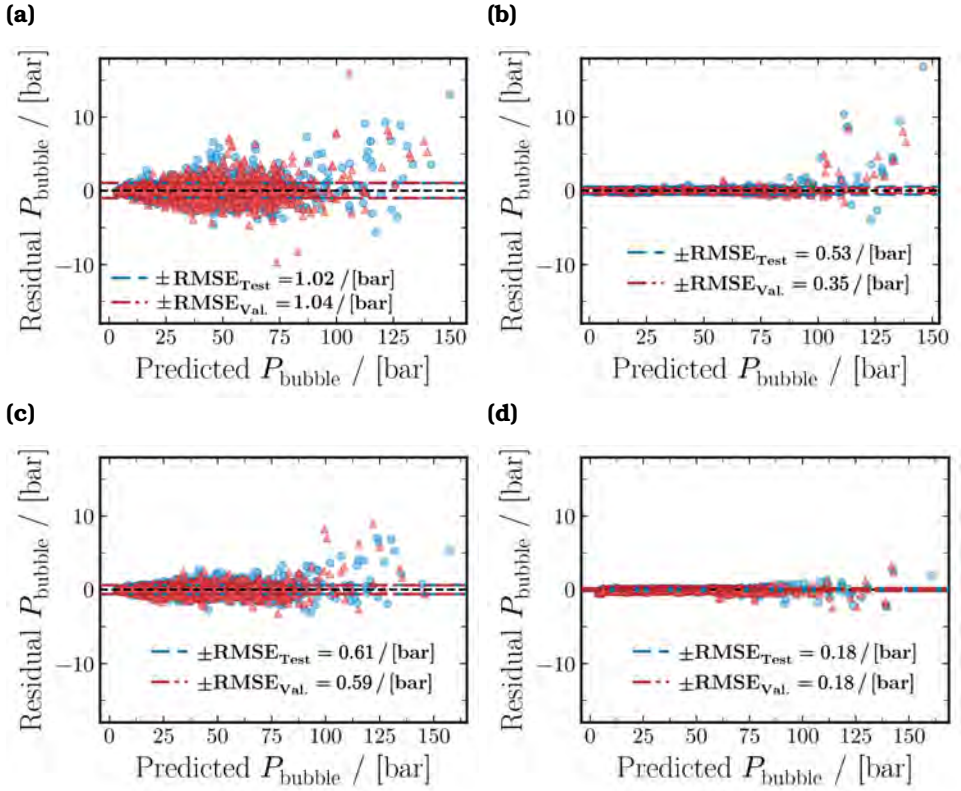


Figure 4.4. Prediction-error (residual) plots for the bubble-point pressure P_{bubble} using (a) Random Forest (RF), (b) Support Vector Regression (SVR), (c) eXtreme Gradient Boosting (XGB), and (d) Tabular Prior-data Fitted Network (TabPFN). Residuals are plotted against the predicted values for the test and validation sets, and the corresponding Root Mean Squared Error (RMSE) is reported in each panel to identify possible model bias, heteroscedasticity, and systematic deviations throughout the prediction range. The test set is shown as blue circles with dashed $\pm$ RMSE bands, and the validation set as red triangles with dash-dotted $\pm$ RMSE bands. [305, 309–312]

a more accurate model can have a smaller threshold and still contain flagged outliers while maintaining better overall predictive performance.

Bubble-point pressure (P_{bubble})

For P_{bubble} , the prediction-error plots are shown in Fig. 4.4. The outlier analysis shows that TabPFN and SVR have fewer large deviations than RF and XGB, with RF showing the most outliers. The maximum residual is

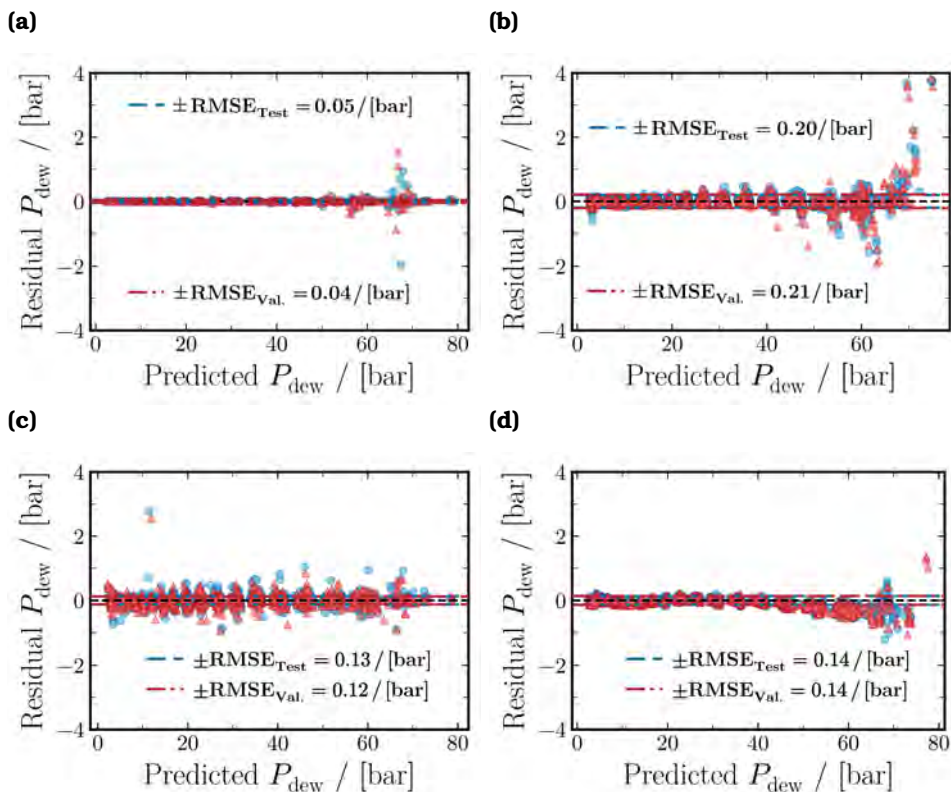


Figure 4.5. Prediction-error (residual) plots for the dew-point pressure P_{dew} using (a) RF, (b) SVR, (c) XGB, and (d) TabPFN. Residuals are plotted against the predicted values for the test and validation sets, and the corresponding RMSE is reported in each panel to identify possible model bias, heteroscedasticity, and systematic deviations throughout the prediction range. The test set is shown as blue circles with dashed $\pm\text{RMSE}$ bands, and the validation set as red triangles with dash-dotted $\pm\text{RMSE}$ bands. [305, 309–312]

ca. 3 bar for TabPFN, ca. 9 bar for XGB, and ca. 16 bar for both RF and SVR. These large residuals occur mainly at low T , especially in the 200–210 K range, suggesting that this region of the thermodynamic space is more difficult to learn. This behavior is associated with mixtures containing higher H_2 concentrations, where P_{bubble} changes strongly with composition. The single-split test and validation RMSE values show that TabPFN achieves the lowest error, ca. 0.18 bar, followed by SVR, XGB, and RF, with values of ca. 0.53 bar, 0.60 bar, and 1.0 bar, respectively. However, the five-fold CV metrics in Table 4.3 indicate that XGB generalizes better than

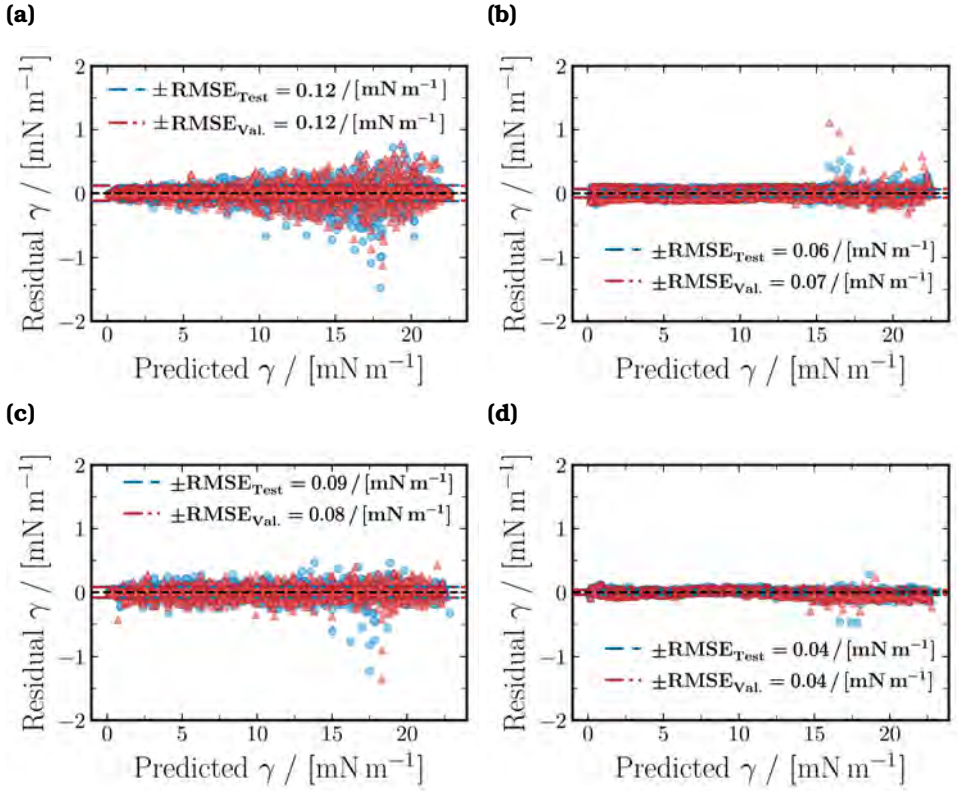


Figure 4.6. Prediction-error (residual) plots for the interfacial tension γ using (a) RF, (b) SVR, (c) XGB, and (d) TabPFN. Residuals are plotted against the predicted values for the test and validation sets, and the corresponding RMSE is reported in each panel to identify possible model bias, heteroscedasticity, and systematic deviations throughout the prediction range. The test set is shown as blue circles with dashed $\pm$ RMSE bands, and the validation set as red triangles with dash-dotted $\pm$ RMSE bands. [305, 309–312]

SVR. Overall, TabPFN provides the best balance between accuracy and robustness, followed by XGB, SVR, and RF.

Dew-point pressure (P_{dew})

For P_{dew} , the prediction-error plots are shown in Fig. 4.5. Unlike P_{bubble} , P_{dew} is less affected by impurities and remains close to the pure CO₂ saturation behavior. Hence, all models resulted in low test and validation RMSE values. For the single train-test-validation split, RF yielded the lowest RMSE of ca. 0.05 bar, followed by TabPFN and XGB with ca. 0.1 bar

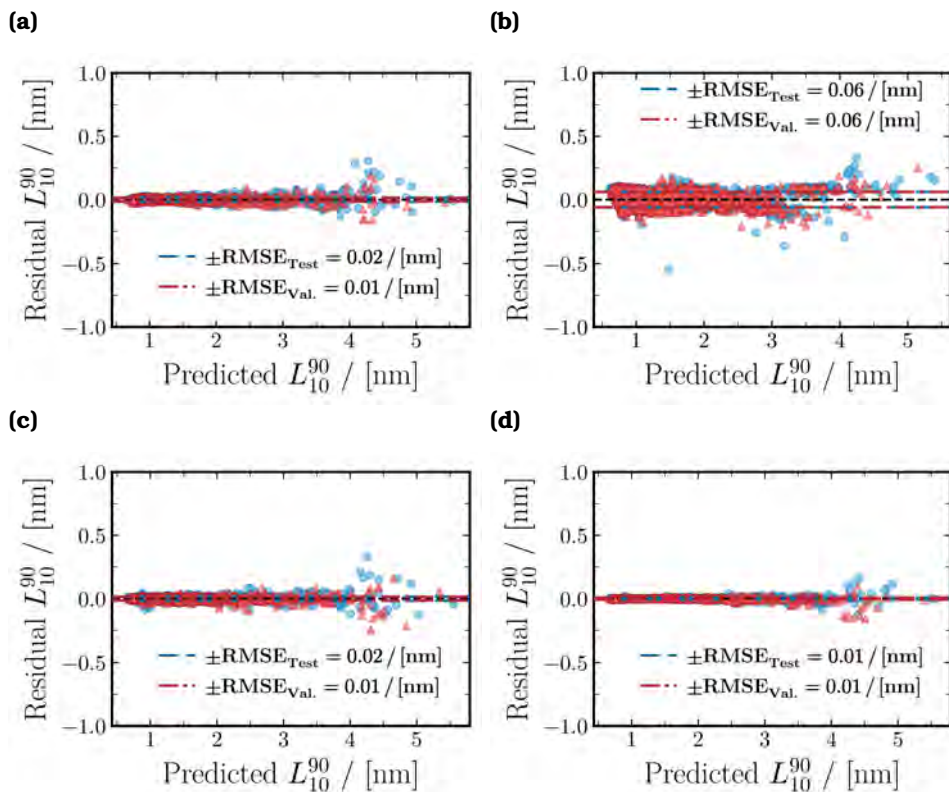


Figure 4.7. Prediction-error (residual) plots for the interfacial thickness L_{10}^{90} using (a) RF, (b) SVR, (c) XGB, and (d) TabPFN. Residuals are plotted against the predicted values for the test and validation sets, and the corresponding RMSE is reported in each panel to identify possible model bias, heteroscedasticity, and systematic deviations throughout the prediction range. The test set is shown as blue circles with dashed $\pm$ RMSE bands, and the validation set as red triangles with dash-dotted $\pm$ RMSE bands. [305, 309–312]

and SVR with ca. 0.2 bar. However, TabPFN showed the best five-fold RMSE, indicating more consistent fold-to-fold generalization. The maximum residuals were ca. 1.4 bar for TabPFN, ca. 2 bar for RF, ca. 3 bar for XGB, and ca. 4 bar for SVR. Most outliers were found in the near-critical region, mainly at $T > 280$ K. Overall, RF, TabPFN, and XGB perform similarly well for P_{dew} , while SVR is the weakest model.

Interfacial tension (γ)

For γ , the prediction-error plots are shown in Fig. 4.6. Most outliers occur in the low- T region, particularly in the range of 200–210 K, similar to P_{bubble} . The maximum residual is ca. 0.5 mN m⁻¹ for TabPFN, ca. 1.1 mN m⁻¹ for SVR, ca. 1.3 mN m⁻¹ for XGB, and ca. 1.5 mN m⁻¹ for RF. The single-split RMSE values also show that TabPFN performs best, followed by SVR, XGB, and RF. This suggests that the tree-based models struggle more in the low- T region compared to the transformer-based and SVM models. The five-fold CV RMSE values in Table 4.3 show that the tree-based models generalize better than SVR, while TabPFN outperforms all models. Overall, TabPFN provides the best γ prediction, combining the lowest RMSE, the lowest worst-case residuals, and the best CV performance, while XGB serves as the closest conventional baseline.

Interfacial thickness (L_{10}^{90})

For L_{10}^{90} , the prediction-error plots are shown in Fig. 4.7. Similar to P_{dew} , L_{10}^{90} is less sensitive to impurities, leading to low RMSE values for all investigated models. In contrast, SVR yielded the highest RMSE and the lowest R^2 , with a five-fold CV R^2 of ca. 0.77, indicating weak generalization. This may be attributed to the difficulty of the ϵ -insensitive loss in capturing the strongly nonlinear increase of L_{10}^{90} near the critical point. TabPFN, XGB, and RF performed similarly on the single train-test-validation split, with RMSE values below ca. 0.02 nm, while TabPFN achieved the lowest test RMSE and the best five-fold CV RMSE. Based on the outlier analysis, the maximum residuals for TabPFN, XGB, and RF were below ca. 0.3 nm, whereas for SVR the maximum residual was about 0.5 nm. Overall, TabPFN is the best model for interfacial thickness prediction, with XGB and RF as close competitors.

Selection and evaluation of TabPFN-based models

Based on the prediction-error, outlier, and five-fold CV analyses, TabPFN was selected as the most reliable full data-driven model for predicting the bubble- and dew-point pressures and interfacial properties of CO₂-rich mixtures. For the other models, XGB showed the most consistent overall performance, RF performed well for impurity-insensitive properties, and SVR showed weaker fold-to-fold robustness during CV. Based on this result, different transformer-based TabPFN configurations [313] were further investigated for P_{bubble} , P_{dew} , γ , and L_{10}^{90} . These configurations include the standard TabPFN model without HPO, TabPFN with HPO, and a hybrid RF + TabPFN model. In the hybrid model, RF partitions the input feature space, while TabPFN improves local predictions within the resulting partitions, thereby combining the structured partitioning ability of tree-based models with the predictive capability of the transformer-based TabPFN model [313]. The same dataset used for the RF, SVR, and XGB

Table 4.3. Five-fold Cross-Validation (CV) performance of the Random Forest (RF), Support Vector Regression (SVR), eXtreme Gradient Boosting (XGB), Tabular Prior-data Fitted Network (TabPFN), and Random Forest (RF) + Tabular Prior-data Fitted Network (TabPFN) Machine Learning (ML) models for all target properties. The reported values correspond to the mean $\pm$ standard deviation for the five validation folds. The arrows ($\uparrow$ and $\downarrow$) indicate the desired direction of model performance. Higher values are better for R^2 ($\uparrow$), whereas lower values are better for Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) ($\downarrow$). The RMSE and MAE are reported in bar for P_{bubble} and P_{dew} , mN m $^{-1}$ for γ , and nm for the interfacial thickness (L_{10}^{90}). Superscripts indicate the TabPFN configuration, where a denotes no Hyperparameter Optimization (HPO) and b denotes HPO.

Target property	ML model	R^2 ($\uparrow$)	RMSE ($\downarrow$)	MAE ($\downarrow$)
P_{bubble}	RF	0.9691 ± 0.0114	3.657 ± 0.849	2.456 ± 0.529
	SVR	0.9402 ± 0.0438	4.658 ± 2.368	2.436 ± 1.041
	XGB	0.9726 ± 0.0161	3.406 ± 1.181	2.300 ± 0.717
	TabPFN ^a	0.9828 ± 0.0062	2.732 ± 0.587	1.841 ± 0.254
	TabPFN ^b	0.9960 ± 0.0016	1.313 ± 0.302	0.766 ± 0.168
	RF + TabPFN ^a	0.9721 ± 0.0176	3.390 ± 1.308	2.034 ± 0.513
P_{dew}	RF	0.9995 ± 0.0004	0.392 ± 0.139	0.120 ± 0.032
	SVR	0.9730 ± 0.0213	2.708 ± 1.314	1.336 ± 0.500
	XGB	0.9993 ± 0.0006	0.419 ± 0.219	0.170 ± 0.060
	TabPFN ^a	0.9997 ± 0.0002	0.298 ± 0.077	0.167 ± 0.010
	TabPFN ^b	0.9998 ± 0.0004	0.201 ± 0.185	0.041 ± 0.012
	RF + TabPFN ^a	0.9997 ± 0.0003	0.296 ± 0.138	0.124 ± 0.027
γ	RF	0.9991 ± 0.0003	0.186 ± 0.030	0.101 ± 0.011
	SVR	0.9765 ± 0.0184	0.856 ± 0.391	0.449 ± 0.168
	XGB	0.9992 ± 0.0004	0.174 ± 0.042	0.109 ± 0.024
	TabPFN ^a	0.9993 ± 0.0002	0.163 ± 0.017	0.122 ± 0.010
	TabPFN ^b	1.0000 ± 0.0000	0.040 ± 0.009	0.020 ± 0.004
	RF + TabPFN ^a	0.9993 ± 0.0004	0.157 ± 0.043	0.091 ± 0.019
L_{10}^{90}	RF	0.9964 ± 0.0046	0.0353 ± 0.0219	0.0116 ± 0.0019
	SVR	0.7786 ± 0.1583	0.2981 ± 0.1240	0.1726 ± 0.0526
	XGB	0.9965 ± 0.0048	0.0339 ± 0.0230	0.0118 ± 0.0015
	TabPFN ^a	0.9980 ± 0.0021	0.0272 ± 0.0143	0.0111 ± 0.0017
	TabPFN ^b	0.9982 ± 0.0010	0.0277 ± 0.0081	0.0074 ± 0.0026
	RF + TabPFN ^a	0.9970 ± 0.0038	0.0320 ± 0.0205	0.0096 ± 0.0022

^a TabPFN without Hyperparameter Optimization (HPO)

^b TabPFN with Hyperparameter Optimization (HPO)

models was used for this comparison. Fig. 4.8a-c compare the NRMSE, NMAE, and R^2 values, respectively, for all target properties. To enable direct comparison between properties with different units and magnitudes, the error metrics were normalized with respect to each target property, as shown in Fig. 4.8a,b. All three TabPFN-based models resulted in low normalized errors for the four target properties. Overall, the tuned TabPFN model with HPO showed the best performance, giving the lowest NRMSE

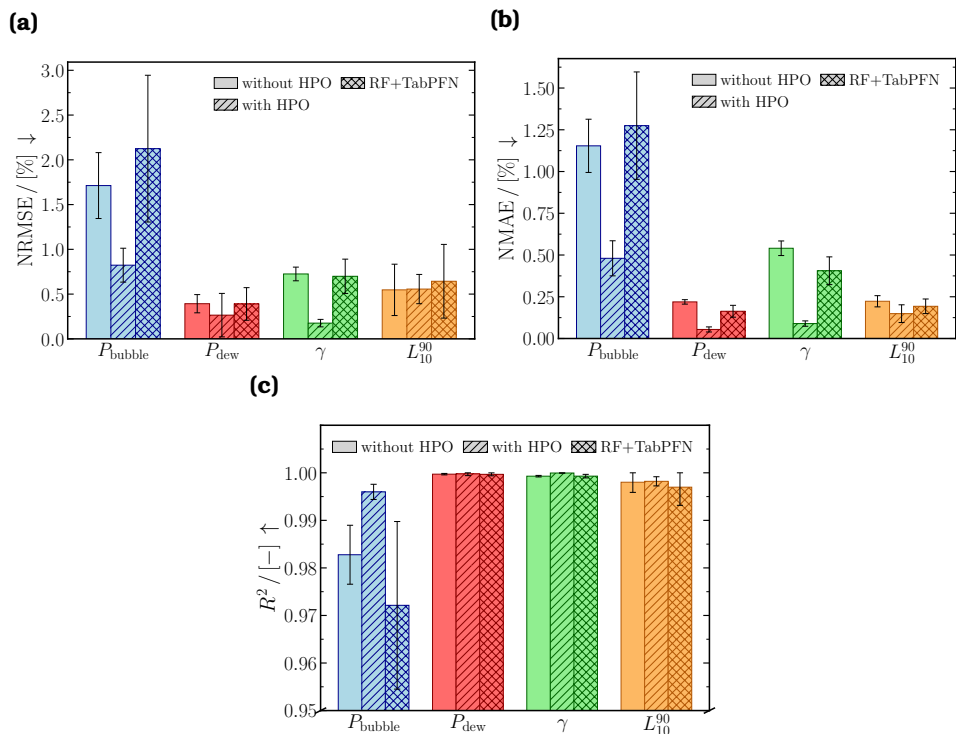


Figure 4.8. Five-fold Cross-Validation (CV) performance comparison of the Tabular Prior-data Fitted Network (TabPFN)-based models for the four target properties. (a–c) Comparison of the Normalized Root Mean Squared Error (NRMSE), Normalized Mean Absolute Error (NMAE), and R^2 , respectively, for P_{bubble} , P_{dew} , γ , and L_{10}^{90} . The three model configurations correspond to TabPFN without Hyperparameter Optimization (HPO), TabPFN with HPO, and the hybrid Random Forest (RF) + TabPFN model. The bar heights represent the mean values from the five CV folds, while the error bars indicate the corresponding standard deviations. The downward arrows in panels (a) and (b) indicate that lower error values are preferred, whereas the upward arrow in panel (c) indicates that higher R^2 values are preferred. [305, 313]

and NMAE for P_{bubble} , P_{dew} , and γ , as well as the highest R^2 values for all four properties. The improvement from tuning was most pronounced for γ , where the NRMSE decreased from ca. 0.7% for the default model to ca. 0.2%, corresponding to nearly a four-fold reduction. For P_{bubble} , the NMAE also decreased from ca. 1.2% to ca. 0.5% after tuning. The default TabPFN and RF-TabPFN models showed similar performance, although both were slightly less accurate than the tuned model. The RF-TabPFN model yielded the largest errors for P_{bubble} , with an NRMSE of ca. 2.1% and an NMAE of ca. 1.3%, but performed comparably to the other models for P_{dew} and

γ . For L_{10}^{90} , the three models showed very similar results: the default TabPFN had the lowest NRMSE of ca. 0.6%, whereas the tuned model had the lowest NMAE of ca. 0.2% and the highest R^2 of 0.9982. These results show that HPO improved TabPFN for most target properties, with the strongest effect observed for γ . In contrast, the gain was limited for L_{10}^{90} , where all three models already performed similarly well. For CO₂ transportation, these accuracies show that the tuned TabPFN model can rapidly screen the effects of impurities on phase boundaries and interfacial properties. This can help identify operating conditions near the two-phase region and estimate the compression pressure required to maintain single-phase dense CO₂ transport without repeatedly performing expensive PCP-SAFT EoS+cDFT computations. The trained TabPFN surrogate for P_{bubble} and P_{dew} can therefore replace the rigorous flash calculations required in CFD and process simulations. The same holds for γ and L_{10}^{90} in two-phase flow conditions. A surrogate is applicable only where it can be trusted. For a given temperature, pressure, and feed composition, TabPFN outputs a distribution of possible values for the target property rather than a single predicted value. Each prediction therefore carries a Prediction Interval (PI) at no additional cost [305]. For example, Section S4.3 of the Supporting Information for this chapter [279] reports the PI for the HPO-tuned P_{bubble} model of feed 2 in Table 4.2, together with the PCP-SAFT EoS reference values.

SHAP analysis

To interpret the feature contributions for best-performing ML models, SHAP analysis was performed for P_{bubble} and γ [313, 337]. These properties were selected because they are more sensitive to impurity type and concentration than P_{dew} and L_{10}^{90} , and are therefore more critical for CO₂ transportation. The SHAP distributions for P_{bubble} predicted using XGB and TabPFN with HPO are shown in Fig. 4.9a and b, respectively, and the corresponding distributions for γ are shown in Fig. 4.9c and d. In each beeswarm plot, features are ranked by mean absolute SHAP value, and the sign indicates whether a feature increases or decreases the prediction relative to the model base value. For P_{bubble} , both models identify z_{H_2} as the dominant feature, with high z_{H_2} values giving positive SHAP values and therefore increasing the predicted bubble-point pressure. This agrees with the expected behavior of H₂ as a light impurity, which increases the pressure required to maintain the liquid phase. T and P are the next most important variables, while high z_{CO_2} values shift the prediction toward lower P_{bubble} , consistent with more CO₂-rich mixtures having lower bubble-point pressures. The remaining impurities have smaller contributions, although TabPFN shows some bidirectional trace-impurity effects, suggesting composition- and state-dependent interactions. For IFT, the dominant features shift from composition to thermodynamic state variables, with T ranked highest in both models, followed by P , while the composition

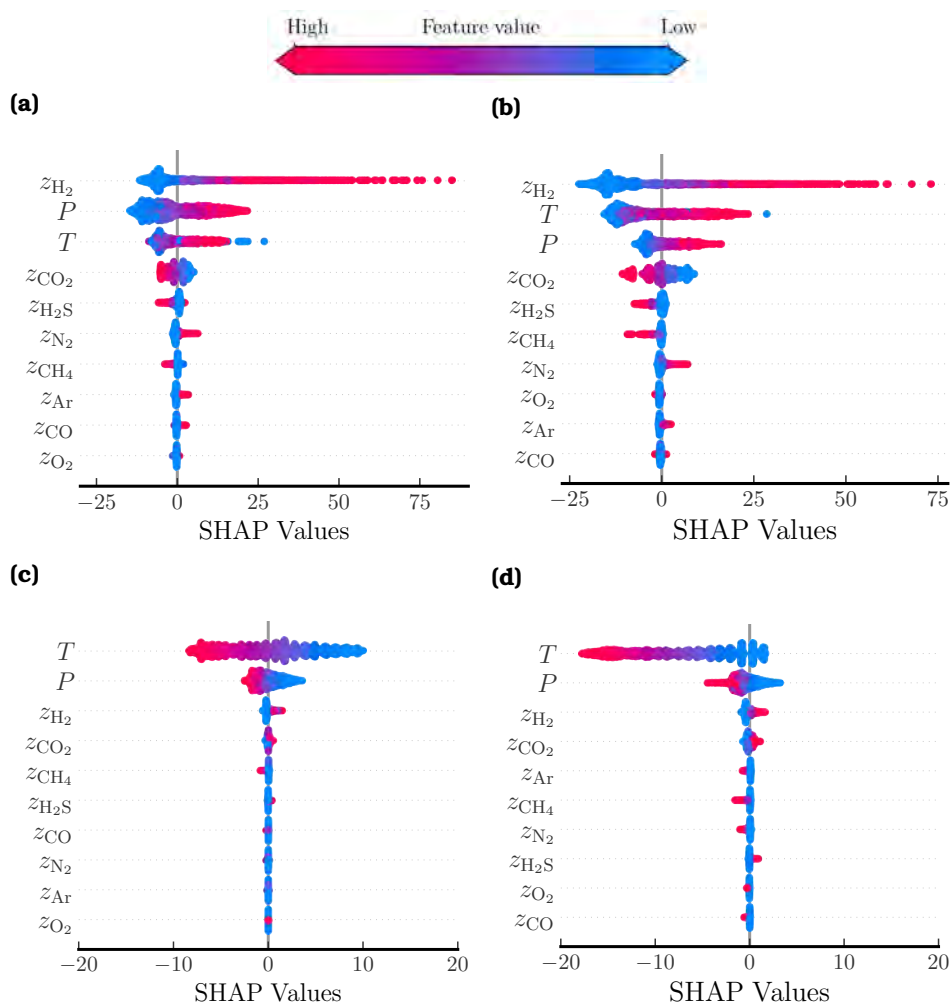


Figure 4.9. SHapley Additive exPlanations (SHAP) value distribution for the eXtreme Gradient Boosting (XGB) and Tabular Prior-data Fitted Network (TabPFN) models with Hyperparameter Optimization (HPO). For P_{bubble} , panel (a) corresponds to eXtreme Gradient Boosting (XGB) and panel (b) corresponds to Tabular Prior-data Fitted Network (TabPFN). For γ , panel (c) corresponds to XGB and panel (d) corresponds to TabPFN. The features are ordered by overall importance, and each point represents the SHapley Additive exPlanations (SHAP) value of a feature for a single sample. The color scale indicates the feature value, with high values in red and low values in blue. [305, 311, 312]

features remain close to zero. High-temperature points mainly appear on the negative SHAP side, confirming that increasing temperature lowers the predicted interfacial tension. Overall, the agreement between XGB and TabPFN on the dominant features and their direction of influence supports the physical consistency of the learned trends, while the broader impurity response in TabPFN suggests that the transformer-based model captures more composition-dependent interactions than XGB.

To summarize, the model performance and interpretability analyses show that TabPFN provides the most reliable full data-driven model for predicting phase-equilibrium and interfacial properties of CO₂-rich mixtures. Compared with the tested models, the HPO-tuned TabPFN model achieves the best overall performance, especially for P_{bubble} and γ . These two properties are particularly important because they are more sensitive to impurity type and concentration and therefore cannot be easily generalized for different CO₂-rich mixture compositions. The SHAP analysis further confirms that P_{bubble} predictions are governed mainly by the z_{H_2} concentration, whereas γ predictions are governed primarily by T and P , supporting the physical consistency of the learned trends. Since this chapter aims not only to screen suitable ML models but also to evaluate the number of feed compositions required to achieve a target prediction accuracy, a composition space sampling analysis was performed using the two best-performing models, TabPFN and XGB, for P_{bubble} and γ , as described in the next section.

4.3.5. Composition Space Sampling

To evaluate the number of feed compositions required to train full data-driven ML models to a given accuracy, a composition space sampling analysis was performed. The feed compositions were sampled from the composition pool $\mathcal{P}$, which represents CO₂ transportation-relevant mixtures. The pool contained 1000 feed compositions, and the pool characteristics are shown in Fig. 4.1. The three sampling methods investigated in this chapter were random sampling, stratified sampling, and AL sampling with four feed composition sizes (N_{train}) of 25, 50, 75, and 100. The minimum size was set to 25 because the stratified sampling method used 25 strata. For each selected feed composition, the target properties were generated using the PCP-SAFT EoS + cDFT framework [68, 104, 245]. For each sampling method and N_{train} size, the evaluation was performed over 20 independent trials. In each trial, a different random seed was used for both N_{train} composition selection and five-fold CV partitioning so that the reported metrics reflect the effect of the selected mixture compositions rather than a particular split. For TabPFN, the default model without HPO was used, because performing HPO for 20 independent trials and 4 N_{train} sizes was computationally demanding. Fig. 4.10a and b show the five-fold CV RMSE of P_{bubble} , averaged over the 20 trials for the TabPFN and XGB models, respectively. For both models, the RMSE decreases as the number of feed compositions increases, with

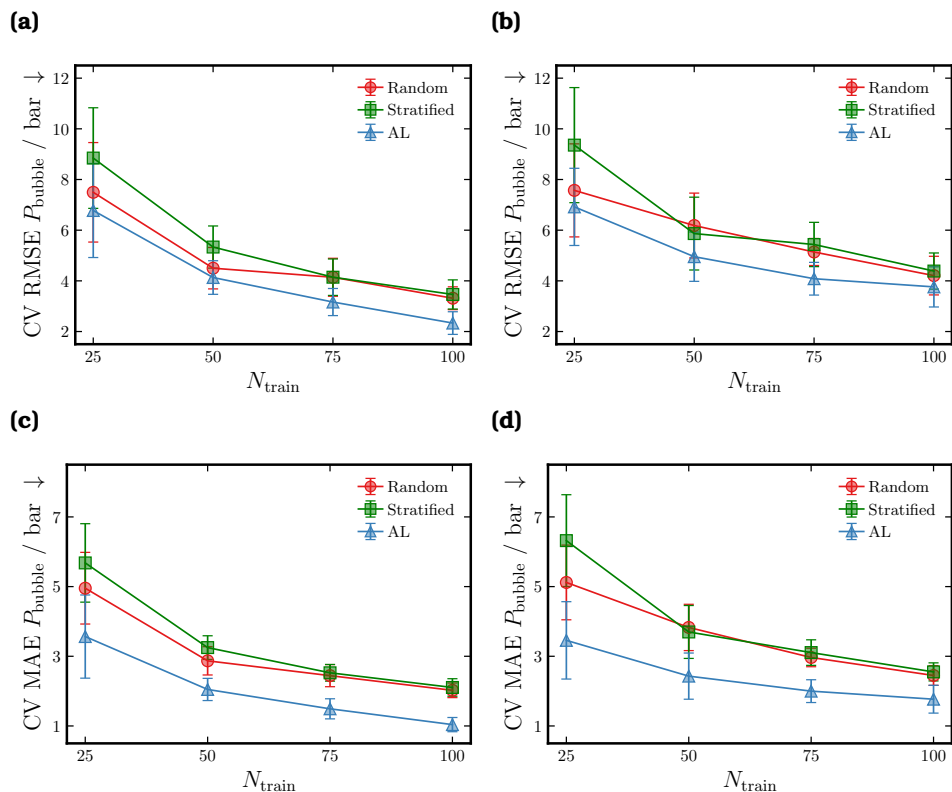


Figure 4.10. Comparison of random, stratified, and Active Learning (AL) sampling strategies for different training composition sizes, N_{train} . For P_{bubble} , panels (a) and (b) correspond to Root Mean Squared Error (RMSE) for Tabular Prior-data Fitted Network (TabPFN) and eXtreme Gradient Boosting (XGB), respectively, while panels (c) and (d) correspond to Mean Absolute Error (MAE) for TabPFN and XGB, respectively. The corresponding γ panels (e–h) are shown on the continued figure. The RMSE and MAE values were obtained from five-fold Cross-Validation (CV). The markers represent the mean values over 20 independent trials, and the error bars represent the corresponding 95% confidence intervals. Low RMSE and MAE values indicate better predictive performance, as indicated by the arrow $\downarrow$. [305, 311, 312, 329]

the largest improvement observed when increasing the training size from 25 to 50 compositions. For AL sampling, the TabPFN RMSE decreases from 6.8 to 4.1 bar, while the MAE decreases from 3.6 to 2.0 bar as shown in Fig. 4.10c, corresponding to an error reduction of about 40%. For XGB, the RMSE decreases from 6.9 to 5.0 bar, while the MAE decreases from 3.5 to 2.4 bar as shown in Fig. 4.10d, corresponding to an error reduction

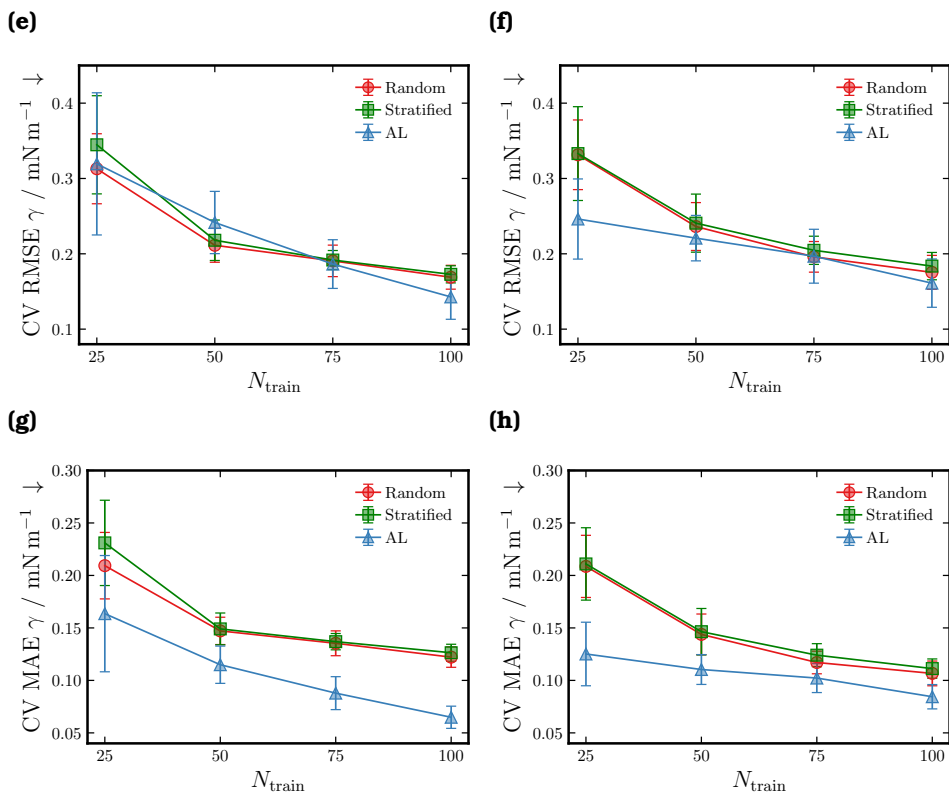


Figure 4.10. (Continued) For γ , panels (e) and (f) correspond to Root Mean Squared Error (RMSE) and panels (g) and (h) correspond to Mean Absolute Error (MAE) for Tabular Prior-data Fitted Network (TabPFN) and eXtreme Gradient Boosting (XGB), respectively.

of about 28%. For N_{train} larger than 50, the error continues to decrease, but the improvement becomes smaller, reaching RMSE values of 2.3 bar for TabPFN and 3.8 bar for XGB at 100 compositions. This indicates that the models are approaching a performance plateau, where further training compositions yield diminishing improvements. The same trend is observed in the CV MAE results shown in Fig. 4.10c and d for TabPFN and XGB, respectively. Compared with random and stratified sampling, AL achieves the lowest RMSE and MAE for all N_{train} sizes, indicating that adaptive selection of informative feed compositions is more data-efficient than nonadaptive sampling. Random and stratified sampling show similar performance at larger training sizes, namely, $N_{\text{train}} = 50, 75$, and 100, but at the smallest size, $N_{\text{train}} = 25$, random sampling performs better than stratified sampling, with RMSE values of 7.5 and 8.8 bar, respectively, for TabPFN. For this reason, random sampling is preferred as the nonadaptive

baseline because it matches stratified sampling at larger training sizes while outperforming it in the low-data regime. Overall, these results indicate that approximately 50 feed compositions provide a favorable balance between accuracy and computational cost for predicting P_{bubble} . Moreover, AL provides the most data-efficient sampling strategy by reducing the number of feed compositions required to reach a given prediction accuracy.

Fig. 4.10e and f show the five-fold CV RMSE of γ for the TabPFN and XGB models, respectively. The values are averaged over the 20 independent trials and are shown as a function of N_{train} . The corresponding five-fold CV MAE values are shown in Fig. 4.10g and h. For TabPFN, all sampling strategies show similar RMSE values for the investigated N_{train} sizes, indicating that the model is less sensitive to the sampling strategy for γ than for P_{bubble} . Similar to P_{bubble} , the largest improvement is observed when N_{train} is increased from 25 to 50. For AL sampling, the TabPFN RMSE decreases from 0.32 to 0.24 mNm⁻¹, while the MAE decreases from 0.16 to 0.12 mNm⁻¹. For XGB, the RMSE decreases from 0.25 to 0.22 mNm⁻¹, while the MAE decreases from 0.13 to 0.11 mNm⁻¹. When the N_{train} is increased to 100, the RMSE values of 0.14 mNm⁻¹ for TabPFN and 0.16 mNm⁻¹ for XGB are noticed, showing that the improvement becomes smaller beyond 50 compositions. Overall, the γ predictions are highly accurate, with errors below 0.5 mNm⁻¹ for all sampling methods and N_{train} . For XGB, AL sampling produces the lowest RMSE for all N_{train} , whereas random and stratified sampling perform similarly. The MAE results show the benefit of AL more clearly, as AL produces the lowest mean error for both TabPFN and XGB for the investigated feed composition sizes. This indicates that AL selects more informative feed compositions for predicting γ , although for TabPFN, the RMSE values of AL and random sampling remain comparable at $N_{\text{train}} < 100$. Random and stratified sampling again show similar performance, with random sampling slightly better at $N_{\text{train}} = 25$, making it the preferred nonadaptive baseline. Together with the P_{bubble} results, these findings indicate that approximately 50 feed compositions provide a good balance between accuracy and computational cost for training full data-driven ML models, while AL offers the most data-efficient sampling strategy, especially when the target is to reduce the number of training compositions required to reach a given accuracy. Based on these results, we recommend GPR-based AL as the sampling strategy for extending the present dataset, for example, to additional impurity types.

4.3.6. Hybrid Machine Learning Model for Interfacial Tension

We further evaluated γ using a hybrid residual-learning approach, in which the simple semi-empirical WSD model [146] was combined with an ML model trained to learn the systematic residuals of the WSD predictions. This approach tests whether incorporating physical prior knowledge improves the predictive accuracy of IFT. The residual, $\Delta\gamma$, was defined as the difference between the IFT computed using the PCP-SAFT EoS + cDFT

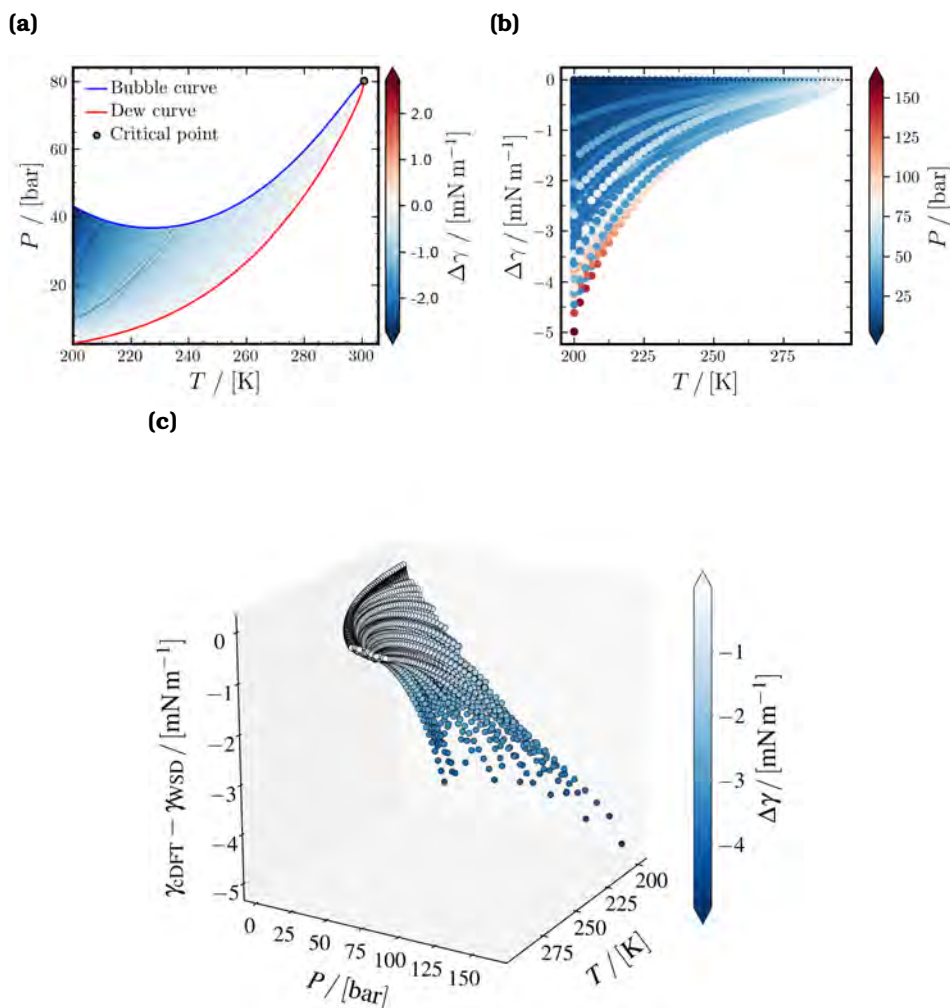


Figure 4.11. Residual analysis of the Winterfeld–Scriven–Davis (WSD) model for Interfacial Tension (IFT), using the Perturbed-Chain Polar Statistical Associating Fluid Theory (PCP-SAFT) Equation of State (EoS) + classical Density Functional Theory (cDFT) results as reference. (a) PT two-phase envelope with the bubble curve, dew curve, critical point, and contour distribution of the IFT residual, $\Delta\gamma$. (b) Variation of $\Delta\gamma$ with temperature inside the two-phase region, with the color scale indicating pressure. (c) Three-dimensional residual surface of $\Delta\gamma$ as a function of temperature and pressure. The residual maps show the deviation of the WSD model from the PCP-SAFT + cDFT reference calculations throughout the two-phase region. [146]

framework [68, 103, 245] and that predicted by the WSD correlation, i.e., $\Delta\gamma = \gamma_{\text{cdft}} - \gamma_{\text{wsd}}$. The residuals were computed for the same 100 feed compositions constructed in Section 4.3.4. Fig. 4.11a shows the contour map of $\Delta\gamma$ within the two-phase region, bounded by the dew- and bubble-point curves, for feed 4 listed in Table 4.2. The residuals are mostly negative, indicating that the WSD correlation overpredicts IFT compared to the PCP-SAFT EoS + cDFT framework [68, 103, 245]. This overprediction is more pronounced at lower temperatures and higher pressures, where the magnitude of $\Delta\gamma$ increases. This behavior can be attributed to the larger density difference between the coexisting phases at these conditions. At lower temperatures, the liquid phase becomes more cohesive and denser, whereas volatile impurities preferentially partition into the vapor phase. As a result, the WSD model tends to overestimate the cohesive contribution of the subcritical mixture and does not fully capture the impurity-induced reduction in IFT. The residual distribution also shows a smooth, structured pattern within the two-phase region rather than random scatter. In particular, $\Delta\gamma$ remains close to zero near the critical point and becomes increasingly negative toward the low-temperature and high-pressure region. This indicates that the error of the WSD correlation has a systematic dependence on the thermodynamic state. Similar contour maps were generated for all 100 feed compositions, and the residuals were found to range from 0 to ca. -5 mNm^{-1} , as shown in Fig. 4.11b. The pressure-colored scatter plot shows that the largest negative residuals are mainly associated with high-pressure states, particularly above ca. 75 bar. This trend is also evident in Fig. 4.11c, where $\Delta\gamma$ is plotted as a function of both T and P , confirming that the largest deviations are concentrated in the low-temperature and high-pressure region of the two-phase envelope. The structured residual field shown in Fig. 4.11c supports the residual-learning strategy because the ML model only needs to learn a smooth state-dependent (T and P) correction to the semi-empirical WSD prediction rather than learning the complete IFT response. For this purpose, we used Symbolic Regression (SR), Gaussian Process Regression (GPR), and Sparse Variational Gaussian Process (SVGP).

Physically interpretable Symbolic Regression (SR) correction for the Winterfeld–Scriven–Davis (WSD) semi-empirical correlation

Fig. 4.12a and b show the parity plot and prediction-error analysis for the Symbolic Regression (SR)-based hybrid model for IFT, in which the baseline WSD semi-empirical correlation was combined with a learned correction for the relative residual. Instead of training the SR model directly on the dimensional residual $\Delta\gamma$ (mNm^{-1}), the target was defined as the dimensionless relative residual $\epsilon_\gamma = (\gamma_{\text{cdft}} - \gamma_{\text{base}})/\gamma_{\text{base}}$, where γ_{cdft} is the PCP-SAFT EoS + cDFT reference value and γ_{base} is the baseline WSD prediction. The predicted residual was then mapped back to the IFT using $\hat{\gamma} = \gamma_{\text{base}}(1 + \epsilon_\gamma)$. Hence, the parity plot and prediction-error analysis for

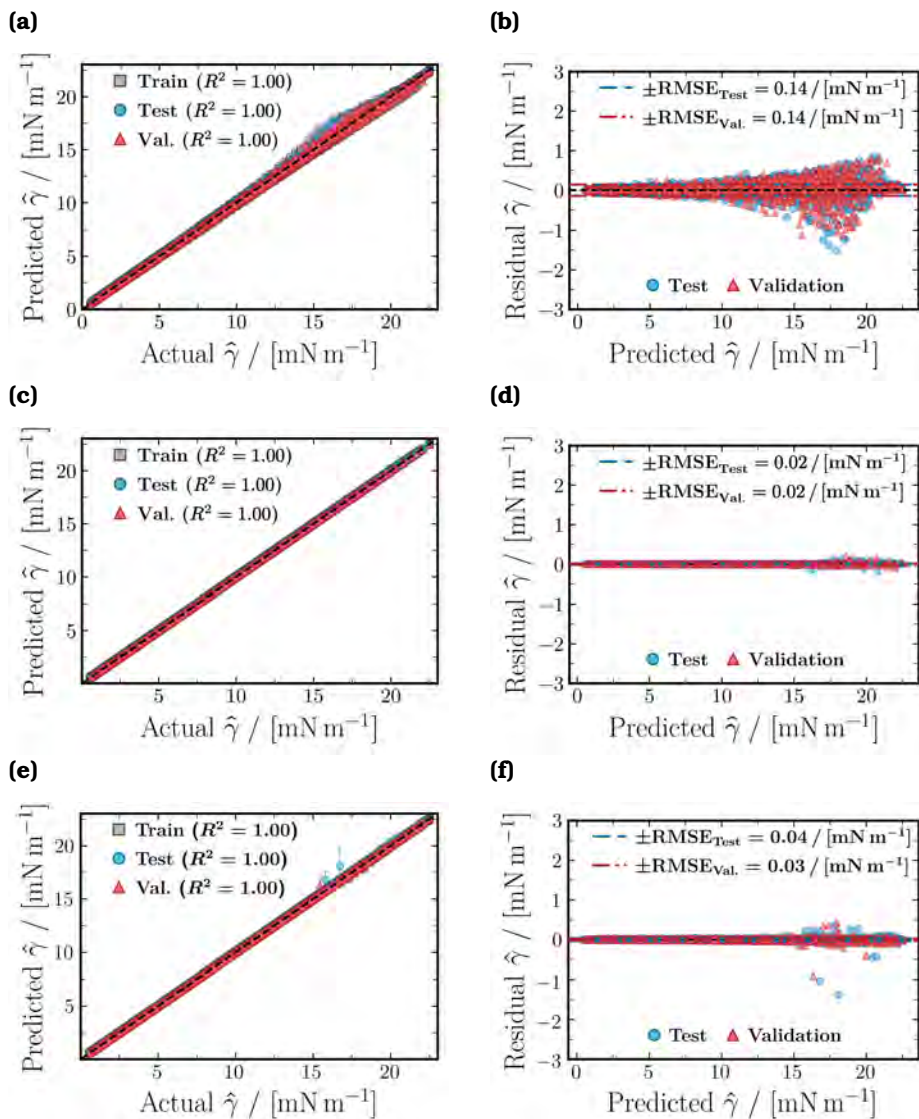


Figure 4.12. Parity and prediction-error analysis for the hybrid Interfacial Tension (IFT) residual models applied to γ . (a,c,e) Comparison of the predicted values with the reference values, with the corresponding R^2 values reported. The right panels (b,d,f) present the residuals as a function of the predicted values with the corresponding Root Mean Squared Error (RMSE) values for the test and validation datasets. (a,b), (c,d), and (e,f) correspond to Symbolic Regression (SR), Gaussian Process Regression (GPR), and Sparse Variational Gaussian Process (SVGP), respectively. For GPR and SVGP, the parity plots include the model-predicted uncertainties. [307, 317–320]

the reconstructed $\hat{\gamma}_{\text{cDFT}}$ are shown in Fig. 4.12a and b, respectively. The input feature set $\mathcal{F}$ for training the ϵ_{γ} was constructed from two groups of physically motivated variables. The first group contained reduced state and interfacial quantities, including $T_{\text{r}}^* = T/T_{\text{c}}$, $1 - T_{\text{r}}^*$, $P_{\text{r}}^* = P/P_{\text{c}}$, $\Pi_{\text{r}}^* = P/P_{\text{sat}}^0$, $\Delta\rho = \rho^{\text{L}} - \rho^{\text{V}}$, $\Delta\rho^* = (\rho^{\text{L}} - \rho^{\text{V}})/(\rho^{0,\text{L}} - \rho^{0,\text{V}})$, and $\gamma_{\text{r}}^* = \gamma_{\text{base}}/\gamma_{\text{CO}_2}^0$. The dimensionless input features used are represented by *. The T_{c} and P_{c} of pure CO₂ are used in T_{r}^* and P_{r}^* . The second group contained composition-based impurity aggregates such as the liquid-, vapor-, and feed-phase impurity sums, which are defined as $x_{\text{imp}} = \sum_{i \neq \text{CO}_2} x_i$, $y_{\text{imp}} = \sum_{i \neq \text{CO}_2} y_i$, and $z_{\text{imp}} = \sum_{i \neq \text{CO}_2} z_i$, respectively. In addition, the second group was also combined with the phase-partitioning difference $\Delta_{\text{imp}} = x_{\text{imp}} - y_{\text{imp}}$ and the impurity partition coefficient $\ell_K = \ln[(y_{\text{imp}} + \xi)/(x_{\text{imp}} + \xi)]$, with $\xi = 10^{-12}$ to avoid division by zero. Combining both groups resulted in a 12-feature set, i.e., $|\mathcal{F}| = 12$, which provides the SR model with physically informed variables to correct the limitations of the baseline WSD model for CO₂-rich mixtures with impurities. Because the baseline WSD model already contains the dominant temperature- and pressure-dependent contribution to IFT, the SR model was also tested using only the composition-based impurity aggregates $\{z_{\text{imp}}, x_{\text{imp}}, y_{\text{imp}}, \Delta_{\text{imp}}, \ell_K\}$ to examine whether the missing correction could be represented only through impurity partitioning features. For this purpose, the SR search was performed over 200 iterations using simple unary and binary operators, yielding several candidate equations with varying complexities and corresponding RMSE values. This reduced feature set SR search recovered the previously suggested correction $\hat{\epsilon}_{\gamma} \approx -2.56x_{\text{imp}}$ [211], where the prefactor 2.56 is an additional empirical constant obtained compared to the previous correction factor suggested by us [211]. This shows that the correction is approximately proportional to the liquid-phase impurity fraction, although it remains less accurate. From the full 12-feature set, the selected correction from the Pareto frontier was

$$\hat{\epsilon}_{\gamma} \approx \ln(\Delta\rho^*)/[\sqrt{P_{\text{r}}^*} + (\gamma_{\text{r}}^* - \Delta\rho^*)], \quad (4.19)$$

which achieved a test R^2 of 0.94 in residual space. After reconstruction to IFT, this correction yielded a $R^2 \approx 1$ (as shown in Fig. 4.12a), with an RMSE of 0.14 mN m⁻¹ and an MAE of 0.074 mN m⁻¹. In comparison, the correction ($\hat{\epsilon}_{\gamma} \approx -2.56x_{\text{imp}}$) obtained using only the composition-based feature set resulted in a residual-space test R^2 of 0.618 and a reconstructed-IFT RMSE of 0.29 mN m⁻¹. The robustness of both the full 12-feature and composition-only corrections was further confirmed by five-fold CV. The corresponding scores are summarized in Table 4.4.

Although both corrections were obtained from the SR search, the selected input features are physically motivated and allow a clear interpretation. The Pareto fronts obtained from the physics-informed and composition-only feature sets are shown in Tables S9 and S10, respectively. The Pareto-frontier equations shown in Table S9 are expressed in terms of the

Table 4.4. Five-fold Cross-Validation (CV) performance of the hybrid ML residual model, evaluated for the reconstructed IFT ($\hat{\gamma}$): Symbolic Regression (SR), Gaussian Process Regression (GPR), and Sparse Variational Gaussian Process (SVGP). Reported R^2 , Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) values are the mean $\pm$ standard deviation over the five folds. The arrows indicate the desired direction of model performance. The SR model was trained on the dimensionless relative residual ϵ_γ , whereas the GPR and SVGP models were trained on the dimensional residual $\Delta\gamma$. For the GPR and SVGP, the reconstructed IFT RMSE and MAE are therefore identical to the residual-space values. The two SR entries correspond to the composition feature set $\mathcal{F}_{\text{Comp.}} = \{z_{\text{imp}}, x_{\text{imp}}, y_{\text{imp}}, \Delta_{\text{imp}}, \ell_K\}$ and the full physics-informed feature set $\mathcal{F} = \mathcal{F}_{\text{Comp.}} + \{T_r^*, 1 - T_r^*, P_r^*, \Pi_r^*, \Delta\rho, \Delta\rho^*, \gamma_r^*\}$.

Model	R^2 ($\uparrow$)	RMSE / [mN m ⁻¹] ($\downarrow$)	MAE / [mN m ⁻¹] ($\downarrow$)
SR ($\mathcal{F}_{\text{Comp.}}$)	0.9983 ± 0.0002	0.2475 ± 0.0110	0.1434 ± 0.0073
SR ($\mathcal{F}$)	0.9996 ± 0.0000	0.1130 ± 0.0073	0.0603 ± 0.0031
GPR	0.9999 ± 0.0001^a	0.0379 ± 0.0262	0.0051 ± 0.0011
SVGP	1.0000 ± 0.0000^a	0.0233 ± 0.0030	0.0119 ± 0.0014

^a The reconstructed-IFT R^2 is obtained per fold as $1 - \text{RMSE}^2 / \text{Var}(\gamma_{\text{cDFT}})$.

density difference ($\Delta\rho$) between the liquid and vapor phases. Similarly, the equations listed in Table S10 are expressed in terms of the total liquid mole fraction of impurities (x_{imp}). The x_{imp} -based equations have a direct physical analogy to surfactant behavior because the impurities decrease the IFT of the CO₂-rich mixture much like a surfactant, consistent with our previous findings [211]. The negative prefactor (-2.56) reflects this monotonic reduction in IFT with increasing liquid-phase impurity content. However, the WSD residual is not governed by composition alone. It also varies with temperature and pressure, with the largest deviations observed at low temperature and high pressure, as shown in Fig. 4.11. The physics-informed Pareto-frontier equations account for these dependencies, even though the baseline WSD model already contains strong T - and P -dependent terms. For the chosen correction, Eq. 4.19, the numerator $\ln(\Delta\rho^*)$ reflects the sharpness of the interface. A large density difference corresponds to a steep density gradient and a highly cohesive, sharp interface. In sharp contrast, a smaller density difference corresponds to a more diffuse interface, with $\Delta\rho^*$ vanishing smoothly toward the critical point. The impurity effect enters implicitly through the bulk densities because impurities lower both the liquid and vapor densities relative to pure CO₂. The denominator, $\sqrt{P_r^*} + (\gamma_r^* - \Delta\rho^*)$, combines a pressure-dependent term with a mismatch between the reduced baseline IFT and the reduced density difference. The term $\sqrt{P_r^*}$ increases with pressure and therefore damps the correction at higher pressures. The difference

$(\gamma_r^* - \Delta\rho^*)$ accounts for deviations between the interfacial response and the bulk density contrast, which can arise from impurity partitioning and interfacial enrichment. Together, these terms make the denominator a state-dependent damping factor. The equations, feature-engineering details, and SR configuration are provided in Section S5.2 and Tables S8–S12 of the Supporting Information for this chapter [279].

GP-based models

Fig. 4.12c and d show the parity plot and prediction-error analysis for the GPR hybrid model, respectively. GPR is well suited for learning the residual function shown in Fig. 4.11c because it defines a posterior distribution over functions consistent with the training data. This allows the GPR to provide both a posterior mean prediction and a calibrated predictive uncertainty for each test and validation data point. However, the main limitation of GPR is its poor scalability with dataset size, as the training cost of exact GPR grows as $\mathcal{O}(n^3)$ with the number of training points. To keep the training tractable, a subset of ca. 5000 points was selected from the full $\sim 1.9 \times 10^4$ -point dataset by stratified sampling over T and P using 10 bins for each variable, ensuring adequate coverage of the state-space variables. With a 70/15/15 split, this left ca. 3500 data points for training. The GPR model was trained on the dimensional residual $\Delta\gamma = \gamma_{\text{cDFT}} - \gamma_{\text{base}}$ using the feature set $\mathcal{X} = \{T, P, \mathbf{z}\}$, which is the same feature set used for the full data-driven ML models. The model reproduced the residual field with a test R^2 of 0.99, an RMSE of 0.016 mN m^{-1} for both the test and validation datasets, and an MAE of 0.006 mN m^{-1} . After reconstructing the IFT using $\hat{\gamma} = \gamma_{\text{base}} + \Delta\hat{\gamma}$, the model achieved $R^2 \approx 1$ for both the test and validation sets, with a reconstructed IFT RMSE of 0.02 mN m^{-1} for both datasets, as shown in Fig. 4.12c and d. Because γ_{base} is a fixed quantity, the reconstruction leaves the absolute errors unchanged, so the reconstructed-IFT RMSE and MAE equal their residual-space values, whereas the reconstructed-IFT R^2 is rescaled to the larger variance of γ_{cDFT} as $1 - \text{RMSE}^2 / \text{Var}(\gamma_{\text{cDFT}})$. Five-fold CV was performed to further confirm the robustness of the model. The corresponding reconstructed-IFT CV scores are summarized in Table 4.4. From Table 4.4, it is evident that the GPR correction outperforms the SR correction in terms of absolute accuracy while additionally providing a per-point uncertainty estimate, although this improvement comes with a strict limitation on the trainable dataset size. To overcome this scalability limit and use the full dataset, a Sparse Variational Gaussian Process (SVGP) model was used. The SVGP model uses inducing points and a variational approximation to retain the probabilistic, uncertainty-aware character of GPR, while reducing the training cost so that it scales linearly with the number of data points. Fig. 4.12e and f show the parity plot and prediction-error analysis for the SVGP hybrid model, respectively. In contrast to the exact GPR, the SVGP model approximates the full posterior using 2000 inducing points. This strategy allowed the model to be trained on the

complete $\sim 1.9 \times 10^4$ dataset while retaining per-point predictive uncertainty. Using the same residual target, $\Delta\gamma$, the SVGP model reproduced the residual field with a test R^2 of 0.996, an RMSE of 0.037 mN m^{-1} , and an MAE of 0.010 mN m^{-1} . For the validation set, the model produced an $R^2 = 0.998$ and an RMSE of 0.026 mN m^{-1} . After reconstructing the IFT, the model achieved $R^2 \approx 1$ on both the test and validation sets, as shown in Fig. 4.12e. The reconstructed IFT RMSE values shown in Fig. 4.12f were 0.04 and 0.03 mN m^{-1} for the test and validation sets, respectively, while the corresponding MAE values were 0.01 mN m^{-1} for both datasets. Five-fold CV also confirmed the robustness of the SVGP model (Table 4.4), with a very small spread for all the folds, indicating robust performance.

Fig. 4.13a and b show the posterior mean $\langle \Delta\gamma \rangle$ and its predictive uncertainty $\sigma_{\Delta\gamma}$ within the two-phase PT envelope considered in this chapter for the exact GPR model. The markers in Fig. 4.13a represent the training data overlaid, sharing the same color map as the posterior mean. The posterior mean is negative throughout the envelope, ranging from ca. -0.02 to -2.84 mN m^{-1} . This is consistent with the actual training residuals spanning a range of 0 to ca. -5 mN m^{-1} (Fig. 4.11b). In Fig. 4.13a, the underestimation of IFT is most pronounced at low T and high P , where the overlaid training markers show the largest negative residuals, reaching ca. -5 mN m^{-1} at $T \approx 200 \text{ K}$ and $P \approx 160 \text{ bar}$. In this region, the posterior mean surface reaches only ca. -2.8 mN m^{-1} , indicating sparse training data. The exact GPR nevertheless reproduces these markers accurately (test $R^2 = 0.99$, RMSE = 0.016 mN m^{-1}), so the apparent outlier reflects that more training data at high pressures need to be included for training to improve the posterior mean predictions. From Fig. 4.13b, the predictive uncertainty rises sharply in this same high-pressure region, where $\sigma_{\Delta\gamma}$ exceeds 2.4 mN m^{-1} above 100 bar , also reflecting the sparse training coverage in the high-pressure zone. This also reflects that more $\text{CO}_2\text{-H}_2$ based mixtures need to be added to generate the training dataset for GPR. Fig. 4.13c and d show the $\langle \Delta\gamma \rangle$ and $\sigma_{\Delta\gamma}$ of the SVGP model. Compared to the exact GPR, the overlaid training data are well reproduced by the SVGP, especially in the low-temperature and high-pressure zones, where the posterior mean surface reaches ca. -4.6 mN m^{-1} and closely matches the dark markers, in contrast to the exact GPR surface that flattened to ca. -2.8 mN m^{-1} . As observed in Fig. 4.13c, $\langle \Delta\gamma \rangle$ exceeds 3.8 mN m^{-1} in magnitude for pressures larger than ca. 100 bar , showing that training the SVGP on the full dataset yields better predictions of $\Delta\gamma$. This improved performance is also reflected in the $\sigma_{\Delta\gamma}$ contour shown in Fig. 4.13d, which reveals that $\sigma_{\Delta\gamma}$ is well below ca. 1.2 mN m^{-1} . This is roughly half the peak uncertainty of the exact GPR (ca. 2.4 mN m^{-1}), with the largest values again confined to the high-pressure zones. Although the residual-space accuracy of the SVGP model (Table 4.4) is marginally lower than that of the exact GPR trained on the 3500-point subset, the SVGP attains better performance while training on the full dataset and scaling linearly with the number

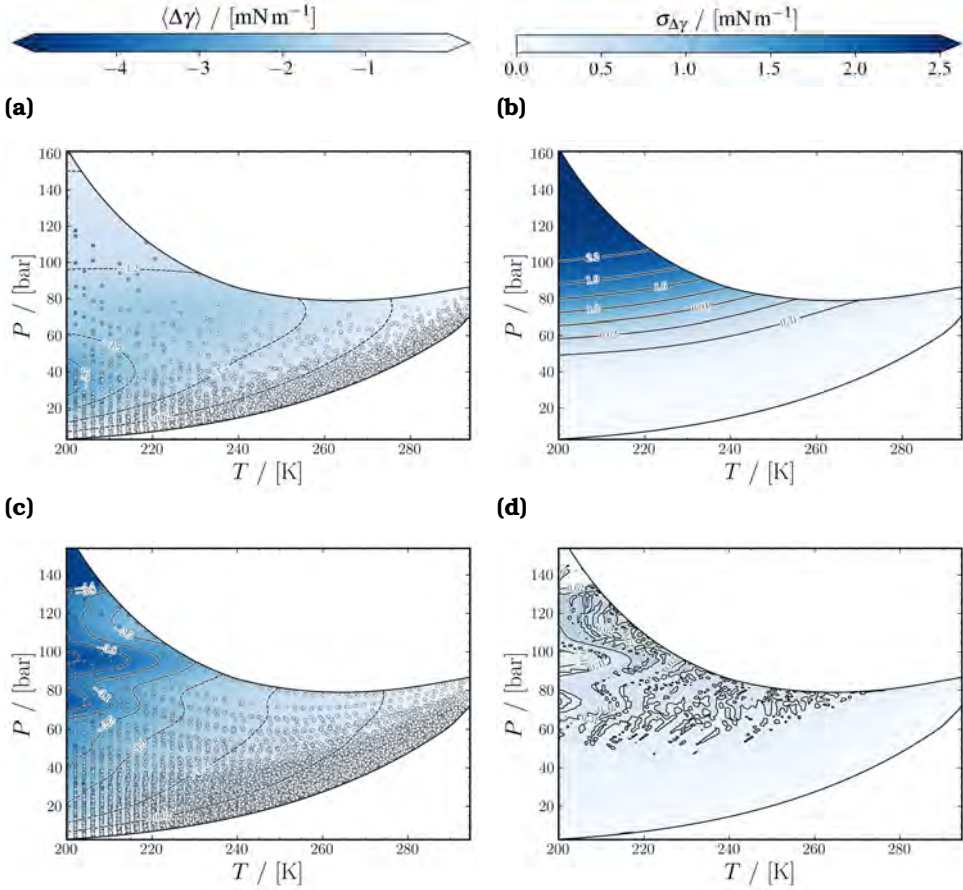


Figure 4.13. Two-dimensional response surfaces of the hybrid residual Machine Learning (ML) models over the PT phase envelope for (a,b) Gaussian Process Regression (GPR) and (c,d) Sparse Variational Gaussian Process (SVGP). (a,c) Posterior mean Interfacial Tension (IFT) residual $\langle \Delta \gamma \rangle$, with the training data overlaid as points colored on the same scale. (b,d) Predictive standard deviation $\sigma_{\Delta \gamma}$. Each column shares the colorbar at the top of the column. The black curves denote the bubble and dew boundaries. [307, 318–320]

of data points. This makes the SVGP a more practical uncertainty-aware correction for large datasets.

In summary, the SVGP learns the residual function in Fig. 4.11c more accurately than the exact GPR, whereas the corrected WSD expression obtained from SR reproduces the same residual only approximately, with a reconstructed-IFT CV RMSE of ca. 0.1 mN m^{-1} for the full physics-informed

feature set (Table 4.4). The SR correction is nonetheless the most computationally efficient route to compute IFT, since its WSD expression carries essentially no inference cost, illustrating the trade-off between accuracy and computational efficiency. To compare the two best ML surrogate strategies investigated in this chapter, namely, the full data-driven model (TabPFN with HPO) and the residual-correction model (SVGP), we examine their CV RMSE and MAE in Table 4.3 and Table 4.4. Both models predict IFT accurately, with CV RMSE values of 0.040 mNm^{-1} (TabPFN) and 0.023 mNm^{-1} (SVGP). Likewise, the CV MAE of SVGP (0.012 mNm^{-1}) is about 1.7 times lower than that of TabPFN (0.020 mNm^{-1}). We further compared the per-state-point inference time, TabPFN with HPO infers in ca. 1.7 ms on an NVIDIA H100 GPU, whereas the SVGP requires ca. 5.2 ms on an Intel Xeon Silver 4214R CPU. On the corresponding hardware, the SVGP is about three times slower, reflecting the same accuracy versus cost trade-off. Regardless of the surrogate, both TabPFN and SVGP reproduce the IFT of the PCP-SAFT EoS + cDFT reference with high accuracy. To quantify the cost, we benchmarked the wall time to compute the IFT map (similar to Fig. 4.2d) of the multicomponent CO_2 mixtures listed in Table 4.2 with temperature-dependent k_{ij} on the AMD EPYC 9754 node and reduced it to a per-state-point cost. This cost ranges from ca. 184 ms for the binary system to ca. 1791 ms for the 8-component system. The cDFT wall time generally increases with the number of components, roughly by one order of magnitude from the 2-component to the 8-component feeds. However, the trend is not strictly monotonic because the cDFT cost also depends on the convergence behavior of the coupled density profiles, which varies between mixtures. Consequently, for each state point, the cDFT reference computation is approximately two to three orders of magnitude slower than TabPFN and one to two orders of magnitude slower than SVGP. This confirms that the ML surrogates reduce the physics-engine cost by two to three orders of magnitude with negligible loss of accuracy.

4.4. Conclusions

In this chapter, we used the PCP-SAFT EoS + cDFT framework to compute the phase equilibria and interfacial properties of multicomponent CO_2 -rich mixtures. The phase-equilibrium properties include the bubble- and dew-point pressures in the PT phase envelope, while the interfacial properties include the IFT and interfacial thickness. The mixtures contain seven CO_2 transportation-relevant impurities, namely, CH_4 , H_2 , N_2 , Ar, O_2 , CO, and H_2S . To obtain reliable mixture predictions, the values of k_{ij} were first optimized for the 23 binary CO_2 -containing pairs using available experimental P_{xy} data. For binary systems with sufficient data at different temperatures, temperature-dependent k_{ij} correlations were fitted to capture the variation in phase behavior with temperature. Using the optimized BIP, we generated a dataset of ca. 1.9×10^4 points for 100 multicomponent CO_2 -rich

mixtures. The mixture compositions were sampled by combining Dirichlet sampling, systematic scaling, and CO₂ feeds from standards and operating pipelines in the ratio 0.7/0.2/0.1. A sensitivity analysis was performed for five multicomponent CO₂-rich mixtures with varying compositions and numbers of components from the 100 multicomponent CO₂-rich mixtures. The results showed that the dew-point pressure and interfacial thickness are less sensitive to impurity concentration than the bubble-point pressure and the IFT at the bubble pressure. The bubble-point pressure and bubble-point IFT both respond strongly to the impurity content. Screening of the full data-driven ML models was performed using the same input variables used to plan an experiment or molecular simulation, namely temperature (T), pressure (P), and mixture composition ($\mathbf{z}$). Compared to the tested models, TabPFN with Hyperparameter Optimization (HPO) provided the best balance between accuracy and robustness for all four properties, especially for the impurity-sensitive P_{bubble} and γ , for which HPO was most beneficial, reaching five-fold CV RMSE values of 1.31 bar for P_{bubble} and 0.040 mN m⁻¹ for γ . The next-best models were XGB, RF, and SVR, except for P_{dew} , where RF and XGB showed comparable performance. The SHAP analysis further confirmed that the prediction of P_{bubble} is influenced primarily by the H₂ concentration, followed by T and P , whereas γ is governed mainly by T and P . It is also shown that TabPFN captures more composition-dependent impurity interactions than XGB. We also evaluated the number of feed compositions required to train the full data-driven ML models to a given accuracy by comparing random, stratified, and GPR-based active-learning sampling strategies. The results showed that active learning is the most data-efficient strategy and that ~50 feed compositions already provide a favorable accuracy-cost balance for the CO₂-rich multicomponent mixtures. Active learning is therefore the strategy we recommend for extending the dataset to new impurity types. In addition, we developed a physics-informed hybrid model for IFT, in which the semi-empirical WSD correlation provides the baseline γ_{base} , and an ML model learns its residual obtained from the PCP-SAFT EoS + cDFT. The IFT is then reconstructed as $\hat{\gamma} = \gamma_{\text{base}}(1 + \hat{\epsilon}_{\gamma})$. Using SR, we obtained the interpretable correction $\hat{\epsilon}_{\gamma} = \ln(\Delta\rho^*)/[\sqrt{P_r^*} + (\gamma_r^* - \Delta\rho^*)]$, with a CV RMSE of 0.11 mN m⁻¹. The SVGP, trained on the full dataset, achieved the lowest CV RMSE of 0.023 mN m⁻¹, outperforming the exact GPR, which was limited to a ~3500-point subset and yielded 0.038 mN m⁻¹. Notably, the best residual-correction model (SVGP, 0.023 mN m⁻¹) outperformed the best full data-driven model for IFT (TabPFN with HPO, 0.040 mN m⁻¹), confirming that embedding the WSD physical baseline improves the prediction of IFT. Finally, benchmarking the computational cost showed that a single cDFT evaluation of γ ranges from ca. 184 to ca. 1791 ms as the system size increases from the binary to the 8-component system. In sharp contrast, the trained surrogates predict a state point in ca. 2 ms for TabPFN and ca. 5 ms for SVGP, resulting in substantial speed-ups with negligible loss of accuracy. This rapid inference

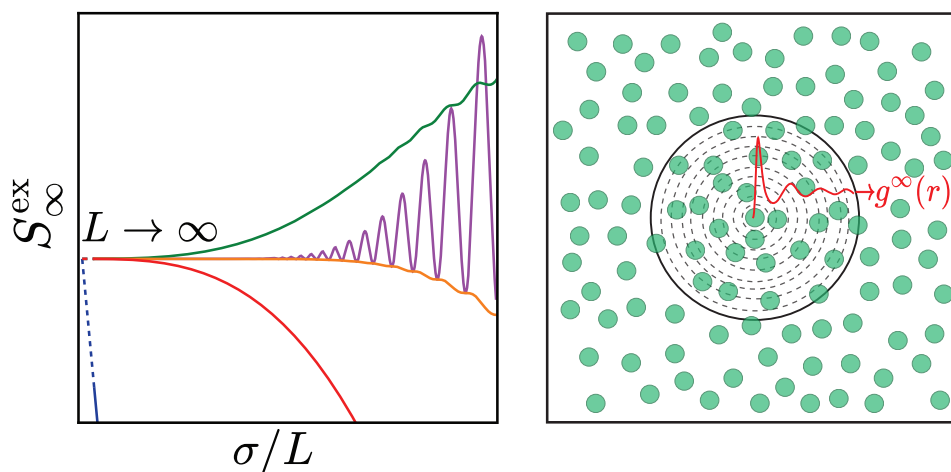
is particularly relevant for dynamic CO₂ transportation simulations, including CFD-based models and transient multiphase flow simulators such as Olga™ [338] and LedaFlow® [339, 340], where efficient pipeline-scale computations require fast property evaluations or precomputed lookup tables. In such applications, the surrogate models act as a composition-aware replacement for precomputed lookup tables, taking T , P , and $\mathbf{z}$ as direct inputs so that no new table is required per feed. To the best of our knowledge, this is the first systematic comparison of the physics engine, PCP-SAFT + cDFT, with two complementary ML surrogate strategies for the interfacial and phase-equilibrium properties of multicomponent CO₂-rich mixtures relevant to CO₂ transportation. These strategies include a full data-driven model and a physics-informed residual-correction model. This framework enables rapid and uncertainty-aware estimation of IFT and saturation properties for multiphase process simulations and inverse design. This chapter supports the design of experiments and molecular simulations by pre-screening the most informative feed compositions for measurement or simulation and by identifying the most suitable ML model for each target property.

5

Finite-size effects of the excess entropy

This chapter is based on the following publication: D. Raju, M. Ramdin, J.-M. Simon, P. Krüger and T. J. H. Vlugt. 'Finite-size effects of the excess entropy computed from integrating the radial distribution function'. *Molecular Physics* 123 (2025), e2456115.

$$\frac{S^{\text{ex}}}{k_{\text{B}}N} = -\frac{\rho}{2} \int_0^L w(r, L) [g(r) \ln(g(r)) - g(r) + 1] dr$$



5.1. Introduction

The excess entropy S^{ex} of a system of N interacting molecules is defined as the difference between the entropy S of the system and the entropy S^{ig} of an ideal gas at the same temperature T and number density $\rho = N/V$, so $S^{\text{ex}}(T, \rho) = S(T, \rho) - S^{\text{ig}}(T, \rho)$ [342]. The excess entropy plays a crucial role in recent theories for predicting transport properties of fluids such as diffusion coefficients, viscosities, and thermal conductivities [74, 75, 142–144, 343, 344]. Hence, there is considerable interest in computing the excess entropy of systems of interacting molecules from molecular simulation. For example, this can be done by performing a free energy calculation (i.e. computing the excess free energy A^{ex} [79]) and using the definition $A^{\text{ex}} = U^{\text{ex}} - TS^{\text{ex}}$ in which T is temperature and $U^{\text{ex}}(T, \rho) = U(T, \rho) - U^{\text{ig}}(T, \rho)$ is the excess potential energy of the system. A more convenient (and computationally less expensive) way is to approximate the excess entropy by a second-order density expansion of the entropy [342, 345, 346]. For an infinitely large system, one can derive the following approximation for the excess entropy [342, 345]

$$\frac{S^{\text{ex}}}{k_{\text{B}}N} \approx -2\pi\rho \int_0^\infty [g(r)\ln(g(r)) - g(r) + 1] r^2 dr \quad (5.1)$$

in which k_{B} is the Boltzmann constant and $g(r)$ is the Radial Distribution Function (RDF), which describes the local density at distance r around a central molecule. As $g(r)$ is computed directly for monoatomic molecules and based on the center of mass for polyatomic molecules by most molecular simulation software, Eq. 5.1 provides a straightforward way to access the excess entropy of a system of N interacting molecules. Eq. 5.1 is also used to compute excess entropies of mixtures by calculating the weighted average of the excess entropies of individual components [343, 347, 348]. As a result, Eq. 5.1 is used in screening studies [343, 349] and crystallization studies [350, 351] to compute S^{ex} . Including higher-order terms in the density expansion of the entropy requires 3-molecule correlation functions [342, 346] which are not often computed due to their complexity [342, 346]. In the context of liquids, the literature often highlights that the second-order density expansion of the entropy accounts for approximately 90% of the excess entropy [342, 346, 350, 352, 353]. Recently, Huang and Widom [354] computed entropies from the third-order density expansion of the entropy using 3-molecule correlation functions following the Kirkwood and Boggs superposition approximation [355]. These authors show that the third-order density expansion of entropy marginally enhances the accuracies in estimating S^{ex} compared to the second-order expansion.

In this chapter, we investigate in detail the underlying approximations of Eq. 5.1 to compute the excess entropy. (1) Eq. 5.1 is a low-density approximation [346] so at high densities one would expect deviations from the exact value of S^{ex} . (2) RDFs computed by molecular simulations show

finite-size effects [356–358], e.g. $g(r)$ approaches 1 at large distances r only if very large systems are considered. This may influence the computed value of S^{ex} . (3) Similar to Kirkwood–Buff (KB) integrals [356, 359], Eq. 5.1 is valid only for infinite systems and it is not a priori clear if it is allowed to truncate the integration of Eq. 5.1 at finite distances. In section 5.2, we investigate the truncation of Eq. 5.1 using an analytic model function for $g(r)$, and we will show that this truncation is possible provided that the range of $g(r)$ is not too long. In the next sections, we systematically investigate the other two assumptions by comparing them with molecular simulations. Simulation details are provided in section 5.3, and a detailed analysis of finite-size effects is provided in section 5.4. Our main findings are summarized in section 5.5.

5.2. Truncation of the Integral for S^{ex}

It is important to note that Eq. 5.1 is strictly speaking only valid for infinite systems. For a finite system with a volume V , to obtain S^{ex} one has to integrate the function $[g(r) \ln(g(r)) - g(r) + 1]$ over the positions of two particles $\mathbf{r}_1$ and $\mathbf{r}_2$ inside this volume V [359]. Only when the volume V is infinitely large, one can replace the integral over the positions $\mathbf{r}_1$ and $\mathbf{r}_2$ by an integral over their distance $r = |\mathbf{r}_1 - \mathbf{r}_2|$ between $\mathbf{r}_1$ and $\mathbf{r}_2$. For S^{ex} , we have for a spherical volume V with diameter L [359]

$$\begin{aligned} \frac{S^{\text{ex}}}{k_{\text{B}}N} &\approx -\frac{\rho}{2} \int_V \int_V [g(r) \ln(g(r)) - g(r) + 1] d\mathbf{r}_1 d\mathbf{r}_2 \\ &= -\frac{\rho}{2} \int_0^L w(r, L) [g(r) \ln(g(r)) - g(r) + 1] dr \end{aligned} \quad (5.2)$$

in which

$$w(r, L) = 4\pi r^2 \left[1 - \frac{3}{2} \left(\frac{r}{L} \right) + \frac{1}{2} \left(\frac{r}{L} \right)^3 \right] \quad (5.3)$$

is the geometric weight function for a sphere with diameter L [356]. Clearly, in the limit, $L \rightarrow \infty$, Eq. 5.2 reduces to Eq. 5.1. For finite-size systems, this is not valid, so one must strictly use Eq. 5.2 instead of Eq. 5.1. This finite-size effect was derived first in the context of KB integrals [356, 359] where one has to integrate $[g(r) - 1]$ over the positions of two particles $\mathbf{r}_1$ and $\mathbf{r}_2$ inside volume V , rather than integrating the function $[g(r) \ln(g(r)) - g(r) + 1]$. For simplicity, let us define the integral over positions $\mathbf{r}_1$ and $\mathbf{r}_2$ in volume V

$$X(L) = \int_V \int_V q(r) d\mathbf{r}_1 d\mathbf{r}_2 = 4\pi \int_0^L q(r) \left[1 - \frac{3}{2} \left(\frac{r}{L} \right) + \frac{1}{2} \left(\frac{r}{L} \right)^3 \right] r^2 dr. \quad (5.4)$$

We also define

$$X^*(L) = 4\pi \int_0^L q(r) r^2 dr, \quad (5.5)$$

which is commonly referred to as the running integral. Only in the limit $L \rightarrow \infty$, $X(L)$ can be replaced by $X^*(L)$. We are interested in an estimation of the value X in the thermodynamic limit (i.e. $L \rightarrow \infty$) which we will denote by X_∞ , obtained by extrapolating from a system at finite volume V . In case of KB coefficients, we have $q_{\text{KB}}(r) = [g(r) - 1]$ and for the excess entropy we have $q_{\text{S}}(r) = [g(r) \ln(g(r)) - g(r) + 1]$. It is important to note that for large distances r , the scaling behavior of these properties is different. As for large distances r , $g(r)$ is close to 1, we can write $g(r) = 1 + \delta$, where $|\delta| \ll 1$, leading to $q_{\text{KB}} \approx \delta$ and $q_{\text{S}} \approx \delta^2$. This indicates that the convergence of KB integrals will generally be much more difficult than the integrals for computing the excess entropy.

In Ref. [359] it was shown that for a finite-correlation length of $q(r)$, the value of X_∞ can be approximated by a Taylor expansion in $1/L$ and only the first-order derivative was considered. We can write the approximation up to the third-order as

$$X_\infty = X(L) - \frac{1}{L} \frac{dX(L)}{d(1/L)} + \frac{1}{2L^2} \frac{d^2X(L)}{d(1/L)^2} - \frac{1}{6L^3} \frac{d^3X(L)}{d(1/L)^3} + \mathcal{O}(1/L^4). \quad (5.6)$$

Using the Leibniz rule [360], we find for the derivatives

$$\frac{dX(L)}{d(1/L)} = 4\pi \int_0^L \left[-\frac{3r}{2} + \frac{3r^3}{2L^2} \right] q(r) r^2 dr, \quad (5.7)$$

$$\frac{d^2X(L)}{d(1/L)^2} = 4\pi \int_0^L \frac{3r^3}{L} q(r) r^2 dr, \quad (5.8)$$

$$\frac{d^3X(L)}{d(1/L)^3} = -12\pi L^6 q(L) + 4\pi \int_0^L 3r^3 q(r) r^2 dr. \quad (5.9)$$

By substitution of these expressions into Eq. 5.6, we obtain approximations for X_∞ of different order

$$\begin{aligned} X_\infty^1 &= X(L) - \frac{1}{L} \frac{dX(L)}{d(1/L)} \\ &= 4\pi \int_0^L \left[1 - \frac{r^3}{L^3} \right] q(r) r^2 dr, \end{aligned} \quad (5.10)$$

$$\begin{aligned} X_\infty^2 &= X(L) - \frac{1}{L} \frac{dX(L)}{d(1/L)} + \frac{1}{2L^2} \frac{d^2X(L)}{d(1/L)^2} \\ &= 4\pi \int_0^L \left[1 + \frac{r^3}{2L^3} \right] q(r) r^2 dr, \end{aligned} \quad (5.11)$$

$$X_\infty^3 = X(L) - \frac{1}{L} \frac{dX(L)}{d(1/L)} + \frac{1}{2L^2} \frac{d^2X(L)}{d(1/L)^2} - \frac{1}{6L^3} \frac{d^3X(L)}{d(1/L)^3}$$

$$= 4\pi \int_0^L q(r)r^2 dr + 2\pi L^3 q(L) = X^*(L) + 2\pi L^3 q(L). \quad (5.12)$$

We are not considering even higher-order derivatives, as these would involve derivatives of $q(L)$ with respect to L . The third-order approximation X_∞^3 includes the running integral shown in Eq. 5.5, which is often used at large L , as an alternative to the KB and excess entropy integrals in the thermodynamic limit [361]. For KB integrals, it was previously found that the first-order approximation X_∞^1 provides accurate results and that higher-order derivatives can be neglected [359], clearly showing that $X(L)$ scales nearly linearly with $1/L$.

To test the various estimates for X_∞ (for both KB ($q_{\text{KB}}(r)$) and excess entropy ($q_{\text{S}}(r)$) integrals), i.e., the value of X in the thermodynamics limit, we consider an analytic model for the RDF [355, 356, 362]:

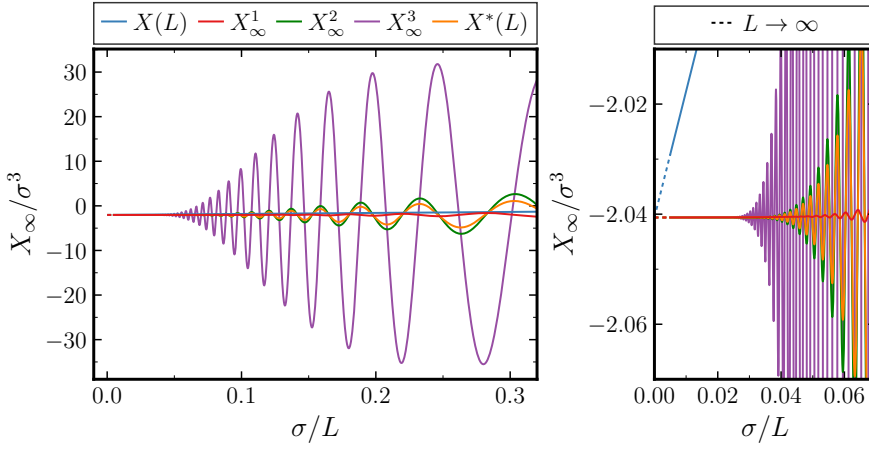
$$g(r) = \begin{cases} 1 + \frac{3/2}{r/\sigma} \exp\left[\frac{1-r/\sigma}{\chi}\right] \cos\left[2\pi\left(\frac{r}{\sigma} - \frac{21}{20}\right)\right], & \frac{r}{\sigma} \geq \frac{19}{20}, \\ 0, & \frac{r}{\sigma} < \frac{19}{20}, \end{cases} \quad (5.13)$$

in which σ is the diameter of the particles and χ is the dimensionless length scale, in units of σ , over which the oscillations of $g(r)$ decay. The parameter χ therefore controls the range of the correlations. In this section, all lengths are expressed in units of σ , so the size L of the volume is reported as L/σ and the integrals X as X/σ^3 . This approach allows one to separately consider finite-size effects of the integral and the other finite-size effects. Fig. 5.1 shows different orders of approximation of X_∞ , along with exact ($X(L)$) and running integral ($X^*(L)$), for $q_{\text{KB}}(r)$ and $q_{\text{S}}(r)$ using the analytic expression of $g(r)$ for $\chi = 2$. As shown in Fig. 5.1a for KB integrals, the exact expression $X(L)$ is free from oscillations but achieves convergence only at a very large length scale compared to $X^*(L)$ and different order of approximation. The third-order approximation of the KB integral has large oscillations and poor convergence compared to other approximations, as shown in Fig. 5.1a. The amplitude of these oscillations decreases with lower approximation orders, while the running KB integral ($X^*(L)$) has an oscillation amplitude slightly less than the second-order approximation. X_∞^3 reaches an asymptotic value at $L/\sigma \approx 40$, while X_∞^2 and $X^*(L)$ converge at $L/\sigma \approx 30$. The first-order approximation has the lowest amplitude of oscillations and converges to an asymptotic value at $L/\sigma \approx 20$. The observation that the first-order approximation converges better than $X(L)$ and $X^*(L)$ aligns with the work of Krüger et al. [359] for KB integrals. For excess entropy integrals shown in Fig. 5.1b, it is clear that the $X(L)$ suffers from poor convergence compared to different order of approximations and $X^*(L)$. Unlike KB integrals, the first and second-order approximations show no oscillations, while the third-order approximation does suffer from oscillations. The running excess entropy integral, $X^*(L)$,

has minor oscillations, with its function intersecting the local minima of X_∞^3 . X_∞^1 of the excess entropy integral reaches an asymptotic value at $L/\sigma \approx 50$, while X_∞^2 reaches at $L/\sigma \approx 40$. $X^*(L)$ and X_∞^3 approach an asymptotic value at $L/\sigma \approx 10$, but X_∞^3 has minor oscillations for $L/\sigma < 10$. Convergence of the excess entropy integral increases with increasing order of approximations, and $X^*(L)$ has better convergence compared to different order of approximations (X_∞^1 , X_∞^2 , X_∞^3). The convergence of different approximations for KB integrals shown in Fig. 5.2a and Fig. 5.3a for $\chi = 10$ and $\chi = 20$ respectively follows the order $X_\infty^1 > X^*(L) > X_\infty^2 > X_\infty^3$, similar to $\chi = 2$. X_∞^1 converges to an asymptotic value at $L/\sigma \approx 130$ and $L/\sigma > 200$ for $\chi = 10$ and $\chi = 20$ respectively. Similarly, for the excess entropy, the convergence is achieved at a smaller length scale L in the following order $X^*(L) > X_\infty^3 > X_\infty^2 > X_\infty^1$ for both $\chi = 10$ and $\chi = 20$ as shown in Fig. 5.2b and Fig. 5.3b. $X^*(L)$ converges to an asymptotic value at $L/\sigma \approx 50$ and $L/\sigma \approx 100$ for $\chi = 10$ and $\chi = 20$ respectively. The foregoing comparison of the minimum L/σ value needed for integral convergence shows that the minimum L/σ value needed for the KB integral (q_{KB} with X_∞^1) is always at least twice the minimum L/σ value needed for the excess entropy (q_{S} with $X^*(L)$).

We have observed that for the KB integrals, i.e., the function $q_{\text{KB}}(r)$, the first-order extrapolation X_∞^1 has the best convergence properties; in particular, it improves with respect to $X^*(L)$. This finding has been discussed before [359, 363] and can be understood from the fact that the weight function $4\pi r^2$ in $X^*(L)$ considerably amplifies the sign-changing oscillations of $q_{\text{KB}}(r)$ (see Fig. 5.4, blue line) [364]. Therefore, simple truncation of the integral at L gives rise to large oscillations of $X^*(L)$ and thus slow convergence (Fig. 5.1a). In contrast, X_∞^1 is based on the exact and almost oscillation-free finite volume KB integral $X(L)$. The difference $X(L) - X_\infty$, which may be considered as a surface term [365], is known to scale as $1/L$ for large L [359, 366]. In X_∞^1 , the leading error of $X(L)$ is corrected without much deteriorating the smoothness inherited from $X(L)$, because the weight function $4\pi r^2[1 - (r/L)^3]$ (see Eq. 5.10) is continuous at $r = L$ [356]. From the present analysis, it is seen that the extrapolations based on higher-order Taylor expansion X_∞^2 and X_∞^3 lead to a quite strong amplification of oscillations in $q_{\text{KB}}(r)$, and thus to a slower convergence than X_∞^1 and $X^*(L)$. For the excess entropy, however, we see that the fastest convergence is obtained with the running integral $X^*(L)$ rather than X_∞^1 . So the question arises why the above reasoning, which explains the convergence behavior of integrals over $q_{\text{KB}}(r)$, does not hold when the integrand is $q_{\text{S}}(r)$. As seen from Fig. 5.4, the function $q_{\text{S}}(r)$ oscillates, but with a much weaker amplitude than $q_{\text{KB}}(r)$. $q_{\text{S}}(r)$ is an essentially positive function, while $q_{\text{KB}}(r)$ changes sign at each oscillation. Both differences can easily be understood in the limit of sufficiently large r (typically $r/\sigma > 2$), where $|g(r) - 1| < 1$. We define $h(r) = g(r) - 1$ and have $q_{\text{S}} = (1 + h)\ln(1 + h) - h \approx h^2 + O(h^3)$. So $q_{\text{S}}(r)$ is essentially

(a)



(b)

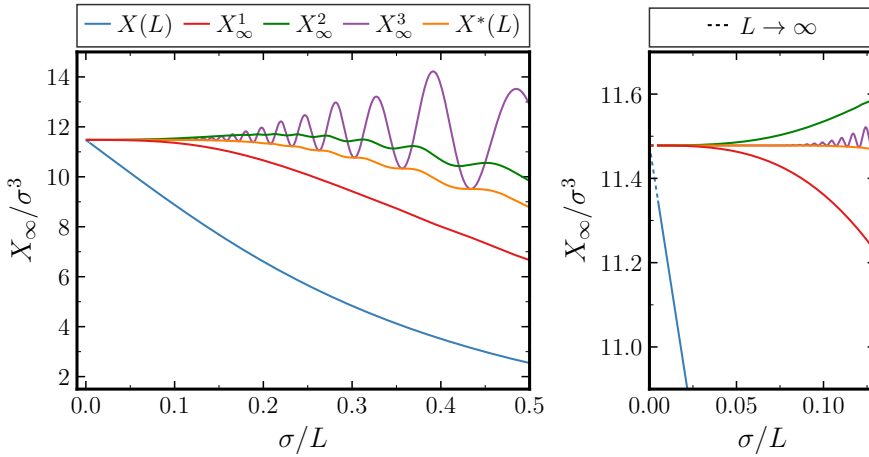
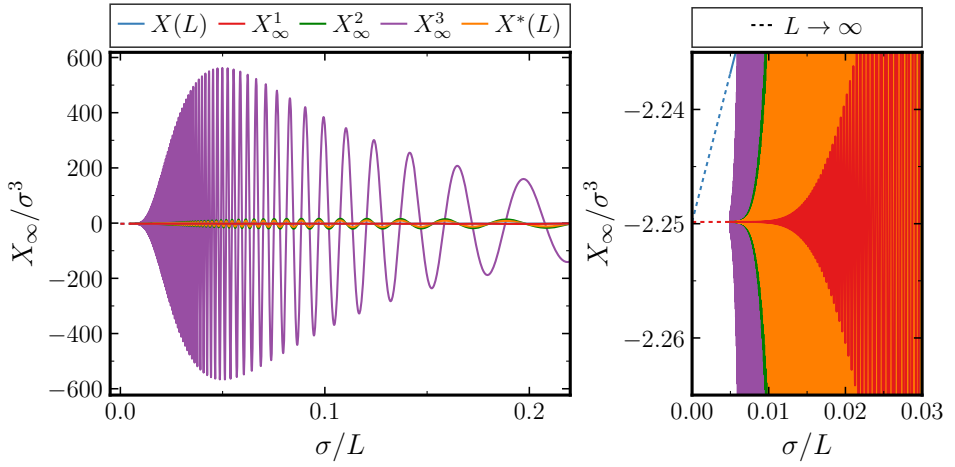


Figure 5.1. Comparison of different approximations $X(L)$, X_∞^1 , X_∞^2 , X_∞^3 , and $X^*(L)$ to compute (a) KB integrals and (b) excess entropy (S^{ex}) integrals in the thermodynamic limit ($L \rightarrow \infty$) obtained from the analytic RDF [355, 362] for $\chi = 2$.

positive and, for $r \rightarrow \infty$, where $h(r) \rightarrow 0$, its amplitude is much smaller than that of $q_{\text{KB}}(r) \equiv h(r)$. Since $0 < q_{\text{S}}(r) < 1$ and $0 < 1 - (r/L)^3 < 1$, it follows that the integrand of X_∞^1 is smaller than that of $X^*(L)$ for all r . As a consequence both X_∞^1 and $X^*(L)$ are strictly increasing and we have $0 < X_\infty^1 < X^*(L) < X_\infty$. This proves that $X^*(L)$ converges faster to X_∞ than X_∞^1 , as also seen in the numerical result of Fig. 5.1b.

(a)



(b)

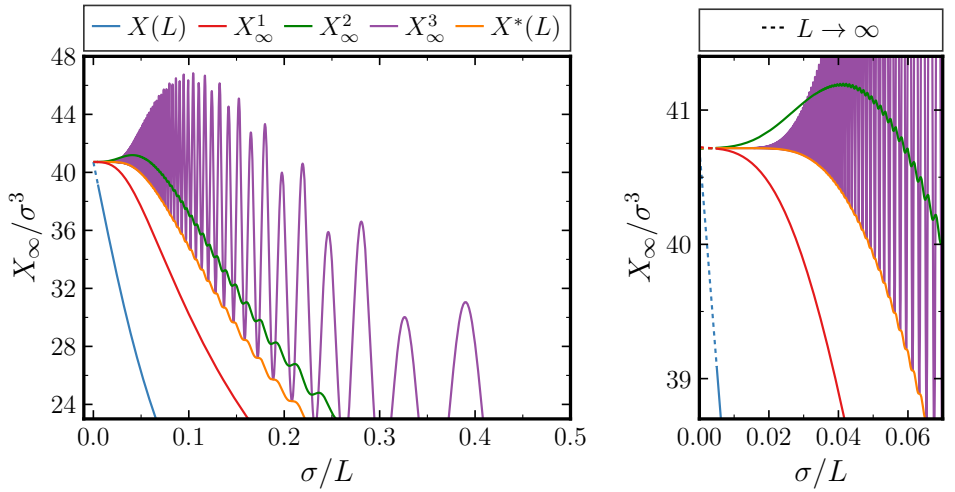
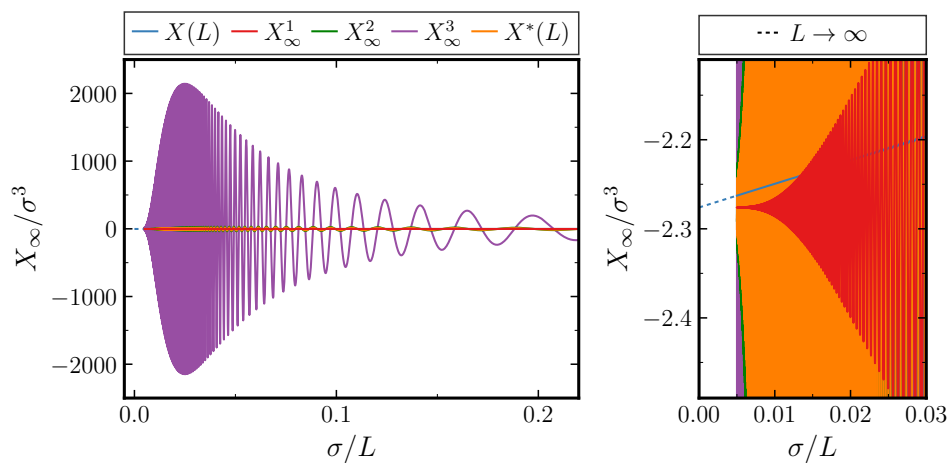


Figure 5.2. Comparison of different approximations $X(L)$, X_∞^1 , X_∞^2 , X_∞^3 , and $X^*(L)$ to compute (a) KB integrals and (b) excess entropy (S^{ex}) integrals in the thermodynamic limit ($L \rightarrow \infty$) obtained from the analytic RDF [355, 362] for $\chi = 10$.

(a)



(b)

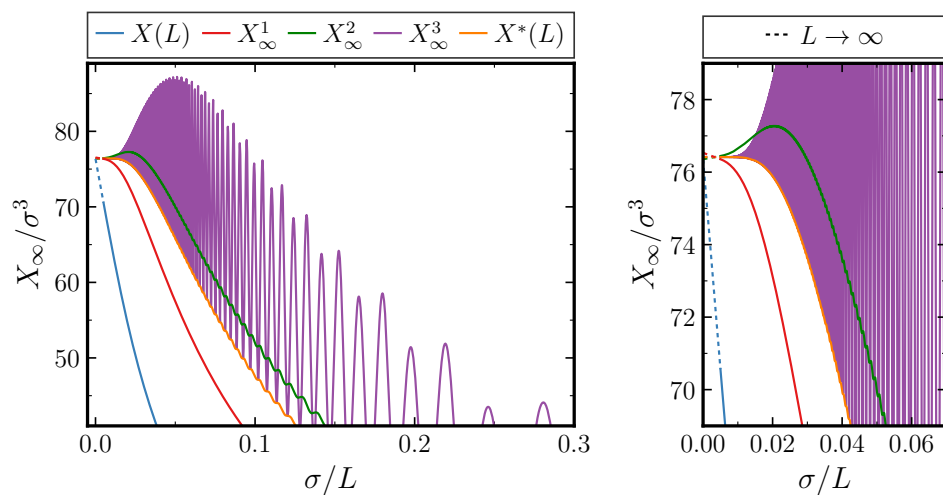


Figure 5.3. Comparison of different approximations $X(L)$, X_∞^1 , X_∞^2 , X_∞^3 , and $X^*(L)$ to compute (a) KB integrals and (b) excess entropy (S^{ex}) integrals in the thermodynamic limit ($L \rightarrow \infty$) obtained from the analytic RDF [355, 362] for $\chi = 20$.

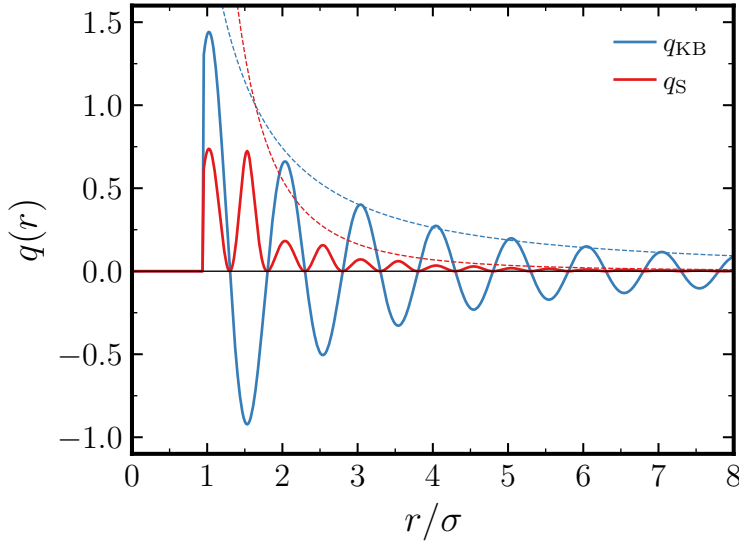


Figure 5.4. Functions $q_{\text{KB}}(r)$ (thick blue line) and $q_{\text{S}}(r)$ (thick red line) for the analytic RDF with $\chi = 10$. The dotted lines are guides to the eye to indicate the amplitude decrease. The blue dotted line is $2.1/(r/\sigma)^{3/2}$ (fit line to the maxima) and the red dotted line is the square of the blue one.

5.3. Simulation Details

We consider a system of molecules in the NVT ensemble that interact via the Wang–Ramírez–Dobnikar–Frenkel (WF) pair interaction potential [367]

$$u_{\text{WF}}(r) = \epsilon \left[\left(\frac{\sigma}{r} \right)^2 - 1 \right] \left[\left(\frac{r_c}{r} \right)^2 - 1 \right]^2 \quad r < r_c \quad (5.14)$$

and $u_{\text{WF}}(r) = 0$ otherwise. In this equation, r is the distance between two interacting molecules, σ is the size parameter, ϵ is a measure of the well-depth of the potential energy, and r_c is the cutoff radius. In the remainder of this chapter, all quantities are expressed in reduced units based on the WF parameters, with σ as the unit of length, ϵ as the unit of energy and $k_{\text{B}} = 1$, so that the pair potential becomes

$$u_{\text{WF}}(r) = \left[\left(\frac{1}{r} \right)^2 - 1 \right] \left[\left(\frac{r_c}{r} \right)^2 - 1 \right]^2 \quad r < r_c. \quad (5.15)$$

Densities, distances, temperatures, potential energies per particle and excess entropies per particle are reported in reduced form as $\rho\sigma^3$, r/σ , $k_{\text{B}}T/\epsilon$, $U/(N\epsilon)$ and $S^{\text{ex}}/(Nk_{\text{B}})$, respectively, in which $\rho = N/V$ is the number density, so all tabulated and plotted quantities are dimensionless.

For $r_c = 2$ and $u_{\text{WF}}(r) = 0$ for $r \geq r_c$, the WF pair potential resembles the Lennard-Jones potential [79, 367]. In contrast, for $r_c = 1.2$, it resembles typical short-range interactions between colloids [79, 367]. The advantage of this interaction potential (e.g., compared to Lennard-Jones) is that one does not need truncation or tail corrections [79]. To compute the excess free energy of the WF system, we adopt a soft-core version of Eq. 5.15

$$u_{\text{WF}}(r, \lambda) = \lambda \left[\frac{1}{r^2 + \alpha(1 - \lambda)} - 1 \right] \left[\frac{r_c^2 + \alpha(1 - \lambda)}{r^2 + \alpha(1 - \lambda)} - 1 \right]^2 \quad r < r_c, \quad (5.16)$$

where λ is the scaling parameter that scales the strength of the WF potential. It is easy to see that for $\lambda = 0$, we have an ideal gas ($u_{\text{WF}}(r, \lambda = 0) = 0$), while for $\lambda = 1$, the original WF interaction potential is recovered. For all λ , we have $u_{\text{WF}}(r_c, \lambda) = 0$. The parameter α is chosen such that one does not have a singularity for $r \rightarrow 0$ unless $\lambda = 1$. Fig. 5.5a shows the soft-core WF interaction potential, $u_{\text{WF}}(r)$, plotted as a function of r/σ , with $r_c = 2$ and $\alpha = 1$ for $\lambda \in [0, 1]$. For decreasing λ , the repulsive interactions become less steep and increase the interaction range over a broad distance, r . Similarly, as shown in Fig. 5.5b for varying α with $r_c = 2$ and $\lambda = 0.5$, decreasing the value of α also reduces the steepness of the repulsive interactions. The effective interaction range and strength can modify the conditions for a possible vapor-liquid phase transition [124, 125]. The excess free energy of a system is computed by Thermodynamic Integration (TI) by scaling the interactions of all particle pairs in the system [79]

$$\begin{aligned} A^{\text{ex}} &= A(T, \rho) - A^{\text{ig}}(T, \rho) \\ &= \int_0^1 \left\langle \frac{\partial U(r, \lambda)}{\partial \lambda} \right\rangle d\lambda \end{aligned} \quad (5.17)$$

with

$$U_{\text{WF}}(r, \lambda) = \sum_{i < j} u_{\text{WF}}(r_{ij}, \lambda) \quad (5.18)$$

and

$$\frac{\partial u_{\text{WF}}(r, \lambda)}{\partial \lambda} = \begin{cases} \frac{A(r, \lambda)B(r, \lambda)}{[\alpha(1 - \lambda) + r^2]^8}, & r < r_c, \\ 0, & r \geq r_c, \end{cases} \quad (5.19)$$

$$\begin{aligned} A(r, \lambda) &= (r_c^2 - r^2) \left[\alpha \lambda (r_c^2 - r^2) \right. \\ &\quad \left. - 2\alpha \lambda (r_c^2 - r^2) (\alpha(1 - \lambda) + r^2 - 1) \right] \end{aligned} \quad (5.20)$$

$$\begin{aligned} B(r, \lambda) &= -(r_c^2 - r^2)^2 (\alpha(1 - \lambda) + r^2) \\ &\quad \times (\alpha(1 - \lambda) + r^2 - 1). \end{aligned} \quad (5.21)$$

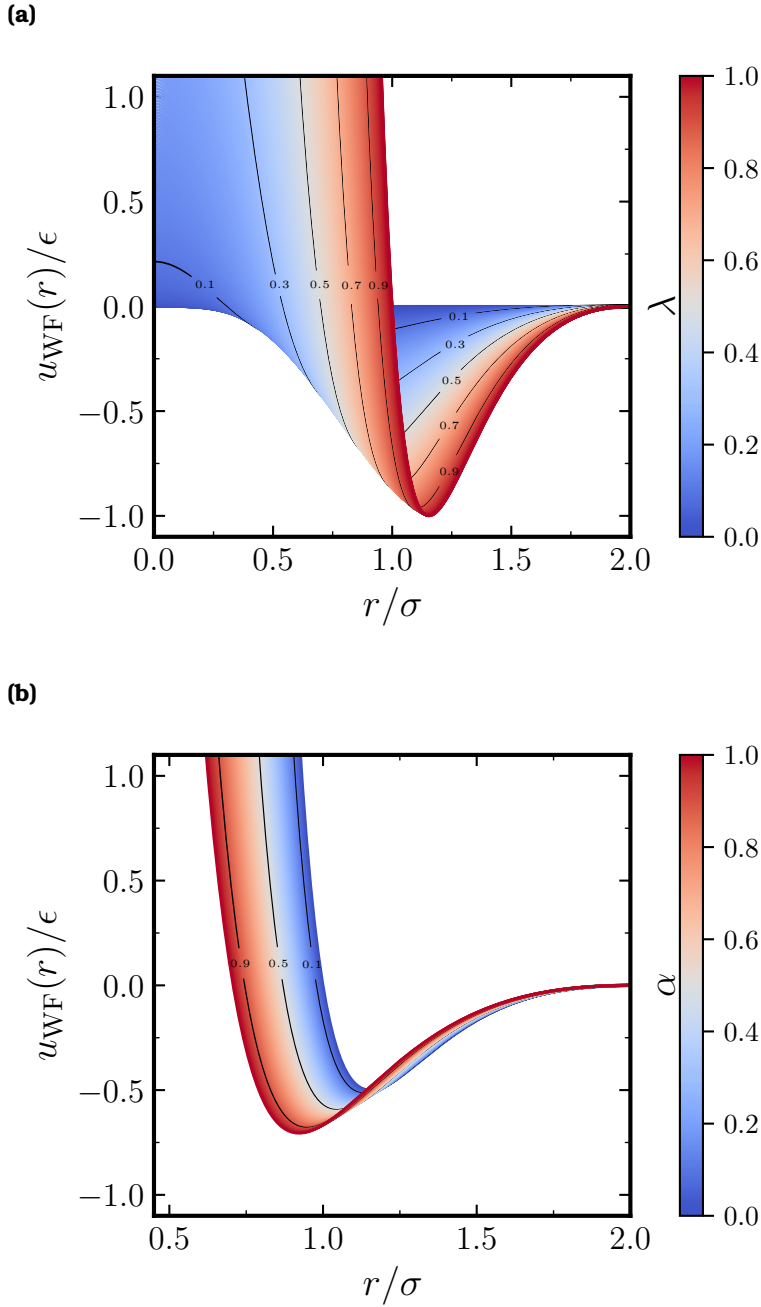


Figure 5.5. The potential energy function $u_{\text{WF}}(r, \lambda)$ as a function of r/σ for the WF pair potential (Eq. 5.16) [367], with $r_c = 2$. (a) Contour plot of WF potential for λ ranging from 0 to 1 with $\alpha = 1$. (b) Contour plot of WF potential for α ranging from 0 to 1 when $\lambda = 0.5$.

Since both $A(r, \lambda)$ and $B(r, \lambda)$ carry a factor $(r_c^2 - r^2)^2$, the product $A(r, \lambda)B(r, \lambda)$ vanishes as $r \rightarrow r_c$, while the denominator of Eq. 5.19 remains finite. The two branches of Eq. 5.19 therefore match at the cutoff, so there is no discontinuity at $r = r_c$ for any value of λ . This is consistent with $u_{\text{WF}}(r_c, \lambda) = 0$ for all λ , and it is the reason why no truncation or tail corrections are needed. Finally, for $\lambda = 1$ we can write $U^{\text{ex}} = U$ and $A^{\text{ex}} = U^{\text{ex}} - TS^{\text{ex}}$, so the excess entropy follows directly from this.

All NVT simulations were performed using an in-house Monte Carlo (MC) code. MC trial moves consist of (randomly selected) particle displacements. Typically, 10^3 equilibration cycles (starting from a random initial configuration) were used, and 10^6 production cycles, with N trial moves per cycle. The maximum particle displacement was adjusted to have ca. 50% of all displacements accepted and was maximized to half the box size. Thermodynamic integration of Eq. 5.17 was performed by running 100 simulations between $\lambda = 0$ and $\lambda = 1$ and by fitting a spline function to $\langle \partial U_{\text{WF}} / \partial \lambda \rangle$ as a function of λ . It was carefully checked that the TI does not cross any vapor–liquid phase transition. For each density $\rho\sigma^3$, 10 independent simulations with a different initial configuration are performed to compute S^{ex} using Eq. 5.1 truncated to a finite size. These 10 simulations are divided into 5 blocks from which average values and uncertainties of S^{ex} are computed. The mean and standard deviation of 5 blocks is the average value and uncertainty of S^{ex} . Finite-size effects of $g(r)$ are corrected by the method of Ganguly and van der Vegt [357]

$$g^{\infty}(r) = g(r) \times \frac{N \left(1 - \frac{4}{3} \pi r^3 / V \right)}{N \left(1 - \frac{4}{3} \pi r^3 / V \right) - \frac{4\pi N}{V} \int_0^r [g(r') - 1] r^2 dr' - 1} \quad (5.22)$$

in which $g(r)$ is the RDF from a simulation of a finite system in the NVT ensemble, and $g^{\infty}(r)$ is its estimate in the thermodynamic limit. Essentially, this method corrects for the slightly different density outside a sphere with radius r around a central particle, compared to the average density N/V . Ganguly and van der Vegt [357] showed that the finite-size correction of RDF to the thermodynamic limit (Eq. 5.22) is effective for nonideal systems with a limited number of particles. One can show that this method corrects the RDF of an ideal gas ($g(r) = (N-1)/N$) to the result in the thermodynamic limit ($g(r) = 1$) and also provides a good estimation of $g(r)$ in the thermodynamic limit for nonideal systems [357, 359, 365].

5.4. Results and Discussion

All MC simulations were performed for densities $\rho\sigma^3$ ranging from 0.01 to 0.8 in the NVT ensemble. The computed average values of S^{ex} for different densities, temperatures, and system sizes are shown in Tables 5.1–5.6 for $r_c = 2$ and $r_c = 1.2$. The statistical uncertainties of S^{ex} , computed from all MC simulations are $< 10^{-3}$. As a result, uncertainties appear smaller than symbols in Fig. 5.6a, 5.6c, 5.8a, 5.8c, 5.9a and 5.9c. For the sake of clarity, the statistical uncertainties of S^{ex} are not included in Tables 5.1–5.6. Excess entropies computed from TI ($S_{\text{TI}}^{\text{ex}}$) and by integrating RDFs (S^{ex} using $g(r)$ and S^{ex} using $g^\infty(r)$) for $k_B T/\epsilon = 4$ with $r_c = 2$, $\alpha = 1$ and $N = 100$ particles are shown in Table 5.1. Simulations were performed for $N = 100$ and $N = 500$ particles to analyze the effects of system size, and the computed S^{ex} from TI and RDFs are shown in Table 5.2. The computed S^{ex} listed in Table 5.1 and 5.2 are plotted in Fig. 5.6a and 5.6c for $N = 100$ and $N = 500$ particles, respectively. From Fig. 5.6a and 5.6c it is clear that $S^{\text{ex}} \rightarrow 0$ for $\rho\sigma^3 \rightarrow 0$. The computed values of S^{ex} obtained by integrating RDFs ($g(r)$ and $g^\infty(r)$) appear to have excellent agreement at low densities. However, due to the extended axis range in Fig. 5.6a and 5.6c, the discrepancies in S^{ex} computed by integrating the uncorrected RDF ($g(r)$) are not observed distinctly in Fig. 5.6a and 5.6c. To analyze S^{ex} computed by integrating RDFs with $S_{\text{TI}}^{\text{ex}}$, Absolute Percentage Errors (APEs) of S^{ex} computed from RDFs are plotted in Fig. 5.6b and 5.6d for $N = 100$ and $N = 500$ particles, respectively. These are defined as

$$\text{Absolute Percentage Error (APE)} = \left| \frac{S_{\text{TI}}^{\text{ex}} - S^{\text{ex}}}{S_{\text{TI}}^{\text{ex}}} \right| \times 100\%. \quad (5.23)$$

From Fig. 5.6b, it is observed that at low densities, S^{ex} computed by integrating the RDF corrected to the thermodynamic limit, $g^\infty(r)$ provides accurate estimations of S^{ex} . S^{ex} computed by integrating the RDF without applying the correction for the thermodynamic limit ($g(r)$) has significant absolute percentage errors. These errors become more pronounced when densities decrease in a system with 100 particles. When the system size is increased from $N = 100$ to $N = 500$ particles, absolute percentage errors of S^{ex} obtained using $g(r)$ become negligible. Consequently, both $g(r)$ and $g^\infty(r)$ provide nearly identical estimations of S^{ex} for a system consisting of $N = 500$ particles. A minor deviation in S^{ex} (using $g(r)$) is observed at $\rho\sigma^3 = 10^{-2}$ in Fig. 5.6d, implying that for $\rho\sigma^3 < 10^{-2}$, accurate estimation of S^{ex} from $g(r)$ necessitates systems with sizes > 500 particles. This illustrates the finite-size effects of $g(r)$, and for very low densities, one needs to consider very large systems for S^{ex} computation using $g(r)$. Nevertheless, S^{ex} can be computed accurately with small system sizes (even with $N = 100$) by using the RDF corrected to thermodynamic limit ($g^\infty(r)$) at low densities. The difference between corrected and uncorrected RDFs computed at $\rho\sigma^3 = 10^{-2}$ for $k_B T/\epsilon = 4$, $r_c = 2$, and $\alpha = 1$ are plotted in Fig. 5.7a for a system with $N = 100$ and in Fig. 5.7b for $N = 500$ particles.

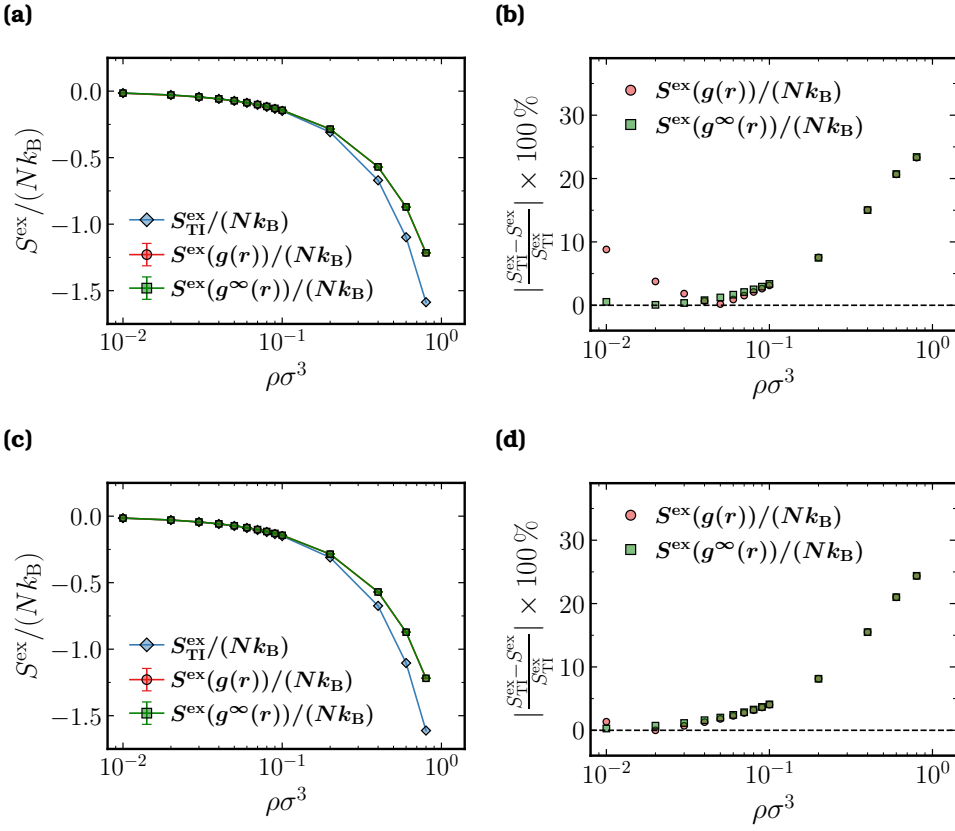


Figure 5.6. Comparison of excess entropies S^{ex} computed for various densities $\rho\sigma^3$ using TI and Eq. 5.1, with and without finite-size corrections to the RDF, $g^\infty(r)$ and $g(r)$, in the NVT ensemble for $k_B T/\epsilon = 4$, $r_c = 2$, $\alpha = 1$, and for different system sizes N . (a) Comparison of S^{ex} for $N = 100$. (b) Computed APE of S^{ex} for $N = 100$. (c) Comparison of S^{ex} for $N = 500$. (d) Computed APE of S^{ex} for $N = 500$.

Table 5.1. Excess entropies S^{ex} computed for various densities $\rho\sigma^3$ from TI using $A^{\text{ex}} = U - TS^{\text{ex}}$ and Eq. 5.1, with and without finite-size corrections to the RDF. The corrected RDF $g^\infty(r)$ (Eq. 5.22) uses the method proposed by Ganguly and van der Vegt [357], while the uncorrected one uses the RDF $g(r)$ directly. Simulations were performed in the NVT ensemble for $k_B T/\epsilon = 4$, $r_c = 2$, $\alpha = 1$, and $N = 100$.

$\rho\sigma^3$	$U/(N\epsilon)$	$S_H^{\text{ex}}/(Nk_B)$	$S^{\text{ex}}(g^\infty(r)) / (Nk_B)$	$S^{\text{ex}}(g(r)) / (Nk_B)$	Absolute Percentage Error - $S^{\text{ex}}(g^\infty(r))$	Absolute Percentage Error - $S^{\text{ex}}(g(r))$
0.01	-0.0391	-0.0144	-0.0145	-0.0157	0.54	8.81
0.02	-0.0782	-0.0290	-0.0290	-0.0301	0.07	3.76
0.03	-0.1170	-0.0436	-0.0434	-0.0444	0.36	1.84
0.04	-0.1558	-0.0583	-0.0578	-0.0587	0.80	0.67
0.05	-0.1944	-0.0731	-0.0722	-0.0730	1.23	0.19
0.06	-0.2328	-0.0880	-0.0865	-0.0872	1.67	0.91
0.07	-0.2712	-0.1030	-0.1009	-0.1014	2.09	1.53
0.08	-0.3093	-0.1181	-0.1151	-0.1156	2.51	2.10
0.09	-0.3473	-0.1333	-0.1294	-0.1298	2.94	2.63
0.10	-0.3852	-0.1486	-0.1436	-0.1440	3.37	3.14
0.20	-0.7561	-0.3083	-0.2852	-0.2850	7.49	7.55
0.40	-1.4341	-0.6701	-0.5694	-0.5691	15.02	15.07
0.60	-1.9117	-1.0982	-0.8707	-0.8712	20.71	20.67
0.80	-1.8920	-1.5865	-1.2153	-1.2168	23.40	23.30

It is clear from Fig. 5.7a that the RDF without finite-size correction differs from the RDF corrected to the thermodynamic limit. The difference lies in the convergence to an asymptotic value of 1, where $g^\infty(r)$ reaches 1 at $r_c = 2$, while $g(r)$ does not converge precisely to 1. In the case of a system with $N = 500$ particles seen in Fig. 5.7b, both RDFs are almost indistinguishable and converge to an asymptotic value of 1 at $r_c = 2$. Finite-size effects of $g(r)$ account for the observed differences in the computed S^{ex} in Fig. 5.6b at low densities. For $\rho\sigma^3 \gtrsim 0.1$ absolute percentage errors of S^{ex} computed from RDFs are $> 5\%$ for both $N = 100$ and 500 particles. For $\rho\sigma^3 \gtrsim 0.1$, absolute percentage errors of S^{ex} computed from RDFs tend to increase as $\rho\sigma^3$ increases. At $\rho\sigma^3 = 0.8$, absolute percentage errors of S^{ex} shown in Fig. 5.6b and 5.6d are $\approx 23\%$ and $\approx 24\%$ for $N = 100$ and $N = 500$ particles, respectively.

The values of S^{ex} computed at $k_B T/\epsilon = 2$ with $r_c = 2$ and $\alpha = 0.5$ for a system with $N = 100$ and $N = 500$ particles are presented in Table 5.3 and 5.4, respectively. For $k_B T/\epsilon = 2$, α was chosen as 0.5 instead of 1 to avoid a vapor-liquid phase transition during the thermodynamic integration. S^{ex}

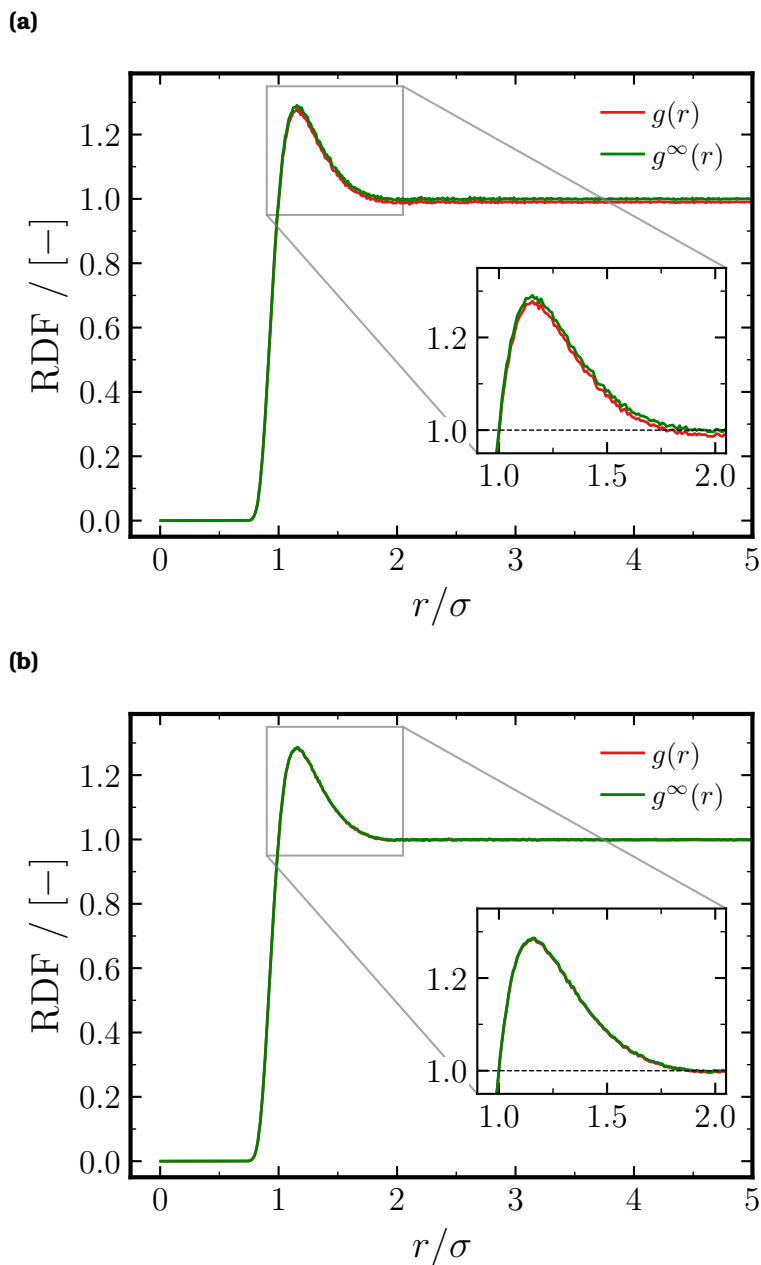


Figure 5.7. Comparison of RDFs computed in the NVT ensemble for $k_B T/\epsilon = 4$, $r_c = 2$, $\alpha = 1$, and $\rho\sigma^3 = 0.01$ without finite-size corrections ($g(r)$) and with finite-size corrections ($g^\infty(r)$) for different system sizes N . (a) Comparison of $g(r)$ and $g^\infty(r)$ for $N = 100$. (b) Comparison of $g(r)$ and $g^\infty(r)$ for $N = 500$.

Table 5.2. Excess entropies S^{ex} computed for various densities $\rho\sigma^3$ from TI using $A^{\text{ex}} = U - TS^{\text{ex}}$ and Eq. 5.1, with and without finite-size corrections to the RDF. The corrected RDF $g^\infty(r)$ (Eq. 5.22) uses the method proposed by Ganguly and van der Vegt [357], while the uncorrected one uses the RDF $g(r)$ directly. Simulations were performed in the NVT ensemble for $k_B T/\epsilon = 4$, $r_c = 2$, $\alpha = 1$, and $N = 500$.

$\rho\sigma^3$	$U/(N\epsilon)$	$S_{\text{TI}}^{\text{ex}}/(Nk_B)$	$S^{\text{ex}}(g^\infty(r)) / (Nk_B)$	$S^{\text{ex}}(g(r)) / (Nk_B)$	Absolute Percentage Error - $S^{\text{ex}}(g^\infty(r))$	Absolute Percentage Error - $S^{\text{ex}}(g(r))$
0.01	-0.0395	-0.0146	-0.0145	-0.0147	0.28	1.35
0.02	-0.0788	-0.0292	-0.0290	-0.0292	0.73	0.00
0.03	-0.1179	-0.0439	-0.0434	-0.0436	1.15	0.71
0.04	-0.1570	-0.0588	-0.0578	-0.0580	1.59	1.29
0.05	-0.1958	-0.0737	-0.0722	-0.0724	2.01	1.80
0.06	-0.2345	-0.0887	-0.0865	-0.0867	2.43	2.28
0.07	-0.2731	-0.1038	-0.1009	-0.1010	2.85	2.74
0.08	-0.3115	-0.1190	-0.1152	-0.1153	3.27	3.18
0.09	-0.3498	-0.1344	-0.1294	-0.1295	3.68	3.61
0.10	-0.3879	-0.1498	-0.1437	-0.1437	4.09	4.05
0.20	-0.7601	-0.3106	-0.2853	-0.2853	8.13	8.14
0.40	-1.4370	-0.6744	-0.5699	-0.5698	15.49	15.50
0.60	-1.9096	-1.1042	-0.8722	-0.8723	21.01	21.00
0.80	-1.8811	-1.6118	-1.2188	-1.2192	24.38	24.36

computed from TI and by integrating the RDFs are plotted in Fig. 5.8a ($N = 100$) and Fig. 5.8c ($N = 500$) including the computed absolute percentage error in Fig. 5.8b ($N = 100$) and Fig. 5.8d ($N = 500$). Similar to $k_B T/\epsilon = 4$, S^{ex} computed by integrating $g^\infty(r)$ is accurate compared to $S_{\text{TI}}^{\text{ex}}$ at low densities. S^{ex} computed by integrating the $g(r)$ for a system with $N = 100$ particles suffer from significant absolute percentage errors at low densities, as seen in Fig. 5.8b. For $N = 500$, both S^{ex} (using $g^\infty(r)$) and S^{ex} (using $g(r)$) are nearly identical as observed in Fig. 5.8d. Absolute percentage errors at $\rho\sigma^3 = 10^{-2}$ for $k_B T/\epsilon = 2$ and $k_B T/\epsilon = 4$ are ca. 6% and ca. 9% for 100 particles, indicating that APEs of S^{ex} computed by integrating the $g(r)$ is temperature dependent. Nevertheless, S^{ex} (using $g^\infty(r)$) provides accurate estimation independent of temperature. Absolute percentage errors of S^{ex} (using $g^\infty(r)$ and $g(r)$) at high densities are found to be increasing with increasing density (maximum of $\approx 20\%$ at $\rho\sigma^3 = 0.8$). MC simulations were also performed to compute S^{ex} for $r_c = 1.2$, where the WF potential behaves like colloid particles. The computed S^{ex} from TI and RDFs for $k_B T/\epsilon = 4$ with $N = 100$ particles are listed and

Table 5.3. Excess entropies S^{ex} computed for various densities $\rho\sigma^3$ from TI using $A^{\text{ex}} = U - TS^{\text{ex}}$ and Eq. 5.1, with and without finite-size corrections to the RDF. The corrected RDF $g^\infty(r)$ (Eq. 5.22) uses the method proposed by Ganguly and van der Vegt [357], while the uncorrected one uses the RDF $g(r)$ directly. Simulations were performed in the NVT ensemble for $k_B T/\epsilon = 2$, $r_c = 2$, $\alpha = 0.5$, and $N = 100$.

$\rho\sigma^3$	$U/(N\epsilon)$	$S_{\text{II}}^{\text{ex}}/(Nk_B)$	$S^{\text{ex}}(g^\infty(r)) / (Nk_B)$	$S^{\text{ex}}(g(r)) / (Nk_B)$	Absolute Percentage Error - $S^{\text{ex}}(g^\infty(r))$	Absolute Percentage Error - $S^{\text{ex}}(g(r))$
0.01	-0.0538	-0.0198	-0.0199	-0.0210	0.51	5.97
0.02	-0.1072	-0.0396	-0.0396	-0.0404	0.00	2.17
0.03	-0.1602	-0.0593	-0.0591	-0.0597	0.47	0.62
0.04	-0.2129	-0.0791	-0.0784	-0.0789	0.92	0.36
0.05	-0.2652	-0.0989	-0.0976	-0.0978	1.37	1.11
0.06	-0.3172	-0.1188	-0.1166	-0.1167	1.81	1.75
0.07	-0.3687	-0.1386	-0.1355	-0.1354	2.23	2.30
0.08	-0.4200	-0.1584	-0.1542	-0.1539	2.65	2.82
0.09	-0.4709	-0.1782	-0.1728	-0.1724	3.07	3.30
0.10	-0.5214	-0.1981	-0.1912	-0.1907	3.48	3.75
0.20	-1.0116	-0.4002	-0.3713	-0.3700	7.22	7.56
0.40	-1.9391	-0.8484	-0.7311	-0.7306	13.83	13.89
0.60	-2.7991	-1.4100	-1.1385	-1.1397	19.26	19.17
0.80	-3.3304	-2.0935	-1.6624	-1.6644	20.59	20.50

plotted in Table 5.5 and Fig. 5.9a, respectively. S^{ex} computed by integrating $g^\infty(r)$ were observed to be in agreement with $S_{\text{II}}^{\text{ex}}$ at low densities, with discrepancies reaching up to 20% at high densities. Absolute percentage error of S^{ex} (using $g(r)$) at $\rho\sigma^3 = 10^{-2}$ in Fig. 5.9b is $\gtrsim 21\%$, whereas in Fig. 5.6b ($r_c = 2$ and $N = 100$) the APE was $\gtrsim 9\%$. This indicates that S^{ex} (using $g(r)$) suffers substantial inaccuracies for colloid-like particles ($r_c = 1.2$) compared to Lennard-Jones-like particles ($r_c = 2$) at low densities. For a system size with $N = 500$ particles the computed S^{ex} are plotted in Fig. 5.9c (also listed in Table 5.6) and their corresponding absolute percentage errors in Fig. 5.9d. For extremely low densities ($\rho\sigma^3 = 10^{-2}$), even for a system of 500 particles S^{ex} (using $g(r)$) show notable errors ($\approx 4\%$) compared to $S_{\text{II}}^{\text{ex}}$ and S^{ex} (using $g^\infty(r)$) in Fig. 5.9d. It is clear from this that the corrected RDF $g^\infty(r)$ should be used in the S^{ex} computation for both $r_c = 2$ and $r_c = 1.2$ to obtain accurate $S_{\text{II}}^{\text{ex}}$ with small system sizes at low densities. Comparing the values of S^{ex} computed using TI for systems with $N = 100$ and 500 particles shows that $S_{\text{II}}^{\text{ex}}$ is independent of system size regardless of r_c . The maximum difference of $S_{\text{II}}^{\text{ex}}$ between different system

Table 5.4. Excess entropies S^{ex} computed for various densities $\rho\sigma^3$ from TI using $A^{\text{ex}} = U - TS^{\text{ex}}$ and Eq. 5.1, with and without finite-size corrections to the RDF. The corrected RDF $g^\infty(r)$ (Eq. 5.22) uses the method proposed by Ganguly and van der Vegt [357], while the uncorrected one uses the RDF $g(r)$ directly. Simulations were performed in the NVT ensemble for $k_B T/\epsilon = 2$, $r_c = 2$, $\alpha = 0.5$, and $N = 500$.

$\rho\sigma^3$	$U/(N\epsilon)$	$S_{\text{H}}^{\text{ex}}/(Nk_B)$	$S^{\text{ex}}(g^\infty(r)) / (Nk_B)$	$S^{\text{ex}}(g(r)) / (Nk_B)$	Absolute Percentage Error - $S^{\text{ex}}(g^\infty(r))$	Absolute Percentage Error - $S^{\text{ex}}(g(r))$
0.01	-0.0542	-0.0199	-0.0199	-0.0201	0.34	0.74
0.02	-0.1081	-0.0399	-0.0396	-0.0398	0.81	0.38
0.03	-0.1616	-0.0599	-0.0591	-0.0592	1.26	1.04
0.04	-0.2146	-0.0798	-0.0785	-0.0785	1.70	1.59
0.05	-0.2674	-0.0998	-0.0977	-0.0977	2.13	2.08
0.06	-0.3198	-0.1198	-0.1167	-0.1167	2.57	2.55
0.07	-0.3718	-0.1398	-0.1356	-0.1356	2.99	3.00
0.08	-0.4234	-0.1598	-0.1544	-0.1543	3.40	3.43
0.09	-0.4748	-0.1799	-0.1730	-0.1729	3.80	3.85
0.10	-0.5258	-0.1999	-0.1915	-0.1914	4.20	4.25
0.20	-1.0193	-0.4039	-0.3722	-0.3719	7.86	7.92
0.40	-1.9470	-0.8547	-0.7330	-0.7328	14.24	14.25
0.60	-2.8001	-1.4171	-1.1431	-1.1434	19.33	19.32
0.80	-3.3181	-2.0996	-1.6758	-1.6763	20.19	20.16

sizes was found to be $\approx 1.6\%$ for $k_B T/\epsilon = 4$, $\rho\sigma^3 = 0.8$, and $r_c = 2$. The discrepancies of S^{ex} computed using RDFs observed at high densities in Figs. 5.6–5.9 can be attributed to the low-density approximation inherent in the second-order density expansion of entropy (Eq. 5.1). The higher-order density expansion of entropy can be used to compute S^{ex} at high densities by following the approximations proposed by Huang and Widom [354]. However, these approximations lose validity in the vicinity of the liquid-to-solid transition ($\rho\sigma^3 \gtrsim 0.8$) [354]. Including higher-order terms could lead to high discrepancies compared to second-order near the liquid-to-solid transition [354].

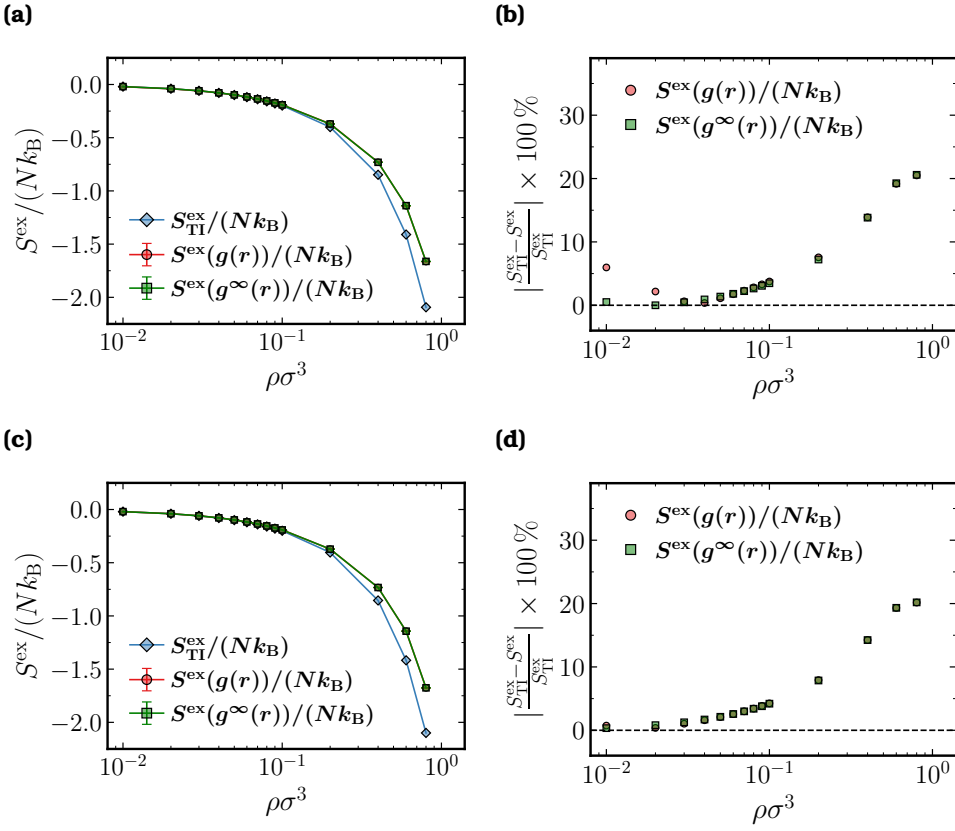


Figure 5.8. Comparison of excess entropies S^{ex} computed for various densities $\rho\sigma^3$ using TI and Eq. 5.1, with and without finite-size corrections to the RDF, $g^\infty(r)$ and $g(r)$, in the NVT ensemble for $k_B T/\epsilon = 2$, $r_c = 2$, $\alpha = 0.5$, and for different system sizes N . (a) Comparison of S^{ex} for $N = 100$. (b) Computed APE of S^{ex} for $N = 100$. (c) Comparison of S^{ex} for $N = 500$. (d) Computed APE of S^{ex} for $N = 500$.

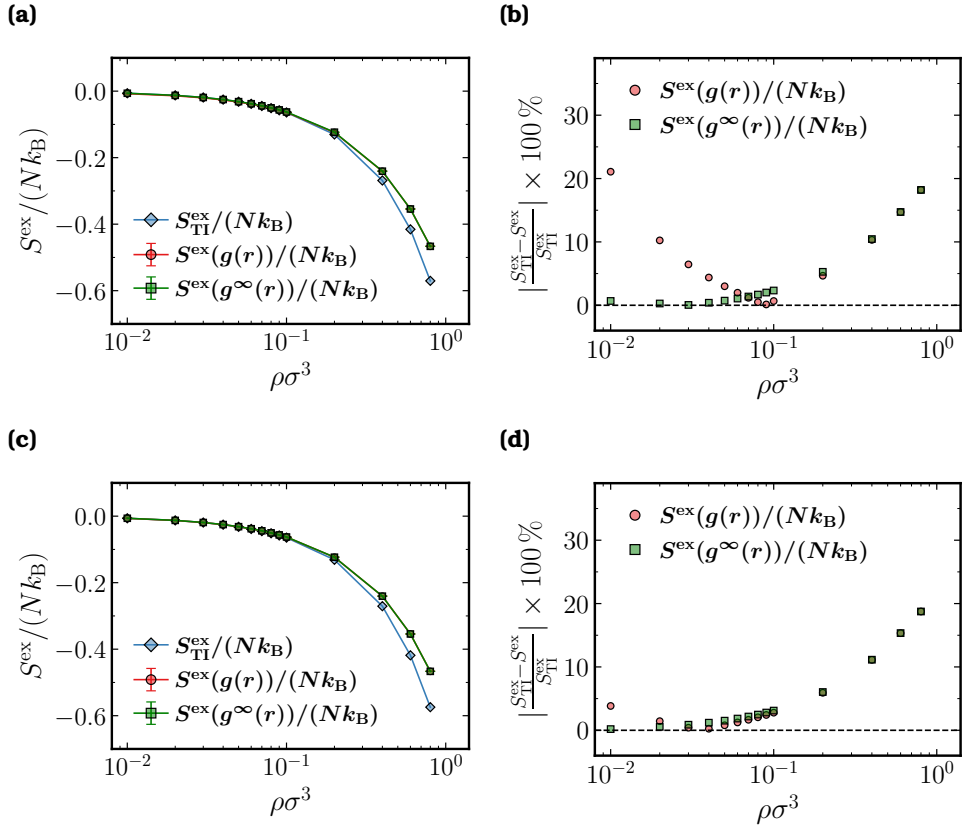


Figure 5.9. Comparison of excess entropies S^{ex} computed for various densities $\rho\sigma^3$ using TI and Eq. 5.1, with and without finite-size corrections to the RDF, $g^{\infty}(r)$ and $g(r)$, in the NVT ensemble for $k_B T / \epsilon = 4$, $r_c = 1.2$, $\alpha = 1$, and for different system sizes N . (a) Comparison of S^{ex} for $N = 100$. (b) Computed APE of S^{ex} for $N = 100$. (c) Comparison of S^{ex} for $N = 500$. (d) Computed APE of S^{ex} for $N = 500$.

Table 5.5. Excess entropies S^{ex} computed for various densities $\rho\sigma^3$ from TI using $A^{\text{ex}} = U - TS^{\text{ex}}$ and Eq. 5.1, with and without finite-size corrections to the RDF. The corrected RDF $g^\infty(r)$ (Eq. 5.22) uses the method proposed by Ganguly and van der Vegt [357], while the uncorrected one uses the RDF $g(r)$ directly. Simulations were performed in the NVT ensemble for $k_B T/\epsilon = 4$, $r_c = 1.2$, $\alpha = 1$, and $N = 100$.

$\rho\sigma^3$	$U/(N\epsilon)$	$S_{\text{H}}^{\text{ex}}/(Nk_B)$	$S^{\text{ex}}(g^\infty(r)) / (Nk_B)$	$S^{\text{ex}}(g(r)) / (Nk_B)$	Absolute Percentage Error - $S^{\text{ex}}(g^\infty(r))$	Absolute Percentage Error - $S^{\text{ex}}(g(r))$
0.01	0.0087	-0.0063	-0.0064	-0.0077	0.68	21.08
0.02	0.0175	-0.0127	-0.0127	-0.0140	0.28	10.24
0.03	0.0264	-0.0190	-0.0190	-0.0202	0.05	6.45
0.04	0.0353	-0.0254	-0.0253	-0.0265	0.40	4.38
0.05	0.0444	-0.0318	-0.0316	-0.0328	0.73	3.00
0.06	0.0536	-0.0382	-0.0378	-0.0390	1.06	1.98
0.07	0.0628	-0.0447	-0.0441	-0.0452	1.38	1.18
0.08	0.0722	-0.0511	-0.0503	-0.0514	1.69	0.49
0.09	0.0817	-0.0576	-0.0565	-0.0575	2.01	0.13
0.10	0.0912	-0.0641	-0.0626	-0.0637	2.33	0.67
0.20	0.1928	-0.1303	-0.1234	-0.1242	5.28	4.66
0.40	0.4312	-0.2687	-0.2406	-0.2410	10.45	10.32
0.60	0.7257	-0.4155	-0.3543	-0.3544	14.72	14.71
0.80	1.0882	-0.5705	-0.4667	-0.4666	18.20	18.21

Table 5.6. Excess entropies S^{ex} computed for various densities $\rho\sigma^3$ from TI using $A^{\text{ex}} = U - TS^{\text{ex}}$ and Eq. 5.1, with and without finite-size corrections to the RDF. The corrected RDF $g^\infty(r)$ (Eq. 5.22) uses the method proposed by Ganguly and van der Vegt [357], while the uncorrected one uses the RDF $g(r)$ directly. Simulations were performed in the NVT ensemble for $k_B T/\epsilon = 4$, $r_c = 1.2$, $\alpha = 1$, and $N = 500$.

$\rho\sigma^3$	$U/(N\epsilon)$	$S_{\text{H}}^{\text{ex}}/(Nk_B)$	$S^{\text{ex}}(g^\infty(r)) / (Nk_B)$	$S^{\text{ex}}(g(r)) / (Nk_B)$	Absolute Percentage Error - $S^{\text{ex}}(g^\infty(r))$	Absolute Percentage Error - $S^{\text{ex}}(g(r))$
0.01	0.0088	-0.0064	-0.0064	-0.0066	0.19	3.84
0.02	0.0176	-0.0128	-0.0127	-0.0129	0.54	1.44
0.03	0.0266	-0.0192	-0.0190	-0.0193	0.88	0.41
0.04	0.0356	-0.0256	-0.0253	-0.0255	1.20	0.25
0.05	0.0447	-0.0321	-0.0316	-0.0318	1.52	0.79
0.06	0.0540	-0.0385	-0.0378	-0.0380	1.85	1.24
0.07	0.0633	-0.0450	-0.0440	-0.0443	2.17	1.66
0.08	0.0728	-0.0515	-0.0503	-0.0505	2.48	2.05
0.09	0.0823	-0.0581	-0.0564	-0.0567	2.80	2.42
0.10	0.0919	-0.0646	-0.0626	-0.0628	3.11	2.78
0.20	0.1941	-0.1313	-0.1233	-0.1235	6.04	5.92
0.40	0.4339	-0.2707	-0.2405	-0.2406	11.14	11.12
0.60	0.7296	-0.4184	-0.3542	-0.3543	15.34	15.33
0.80	1.0932	-0.5744	-0.4667	-0.4667	18.75	18.75

Table 5.7. Summary of underlying key assumptions to consider for excess entropy (S^{ex}) computation using molecular simulations.

S. No.	Observations and remarks
1.	For the excess entropy, the running integral $X^*(L)$ (Eq. 5.5), i.e. simple truncation of Eq. 5.1 at a finite L , converges fastest to the thermodynamic limit. It outperforms the Taylor extrapolations of first, second and third order, X_∞^1 , X_∞^2 and X_∞^3 (Eqs. 5.10, 5.11 and 5.12). For KB integrals the ordering is reversed, with X_∞^1 converging fastest and $X^*(L)$, X_∞^2 and X_∞^3 converging more slowly.
2.	Finite-size corrected RDFs suggested by Ganguly and van der Vegt [357] (Eq. 5.22) must be used for computing the excess entropy, irrespective of the system size.
3.	Eq. 5.1 is a low-density approximation, for $\rho\sigma^3 > 0.1$ thermodynamic integration or other approximation methods suggested by Huang and Widom [354] using higher-order density expansion of entropy are preferred.

5.5. Conclusions

In this chapter, we have investigated the computation of S^{ex} and KB integrals extrapolated to the thermodynamic limit ($L \rightarrow \infty$) in a finite volume using analytic RDFs. Expressions to compute S^{ex} and KB integral at $L \rightarrow \infty$ were derived based on a Taylor expansion in $1/L$ for different orders. We observed that the running integral $X^*(L)$ (Eq. 5.5) and first-order approximation X_∞^1 (Eq. 5.10) converge faster than other approximations for S^{ex} and KB integrals for different ranges of $g(r)$. We noticed that the $X^*(L)$ approximation integral for S^{ex} converged much faster than the X_∞^1 approximation of the KB integral, irrespective of the range of $g(r)$. We showed that truncation of S^{ex} and KB integrals is possible, provided the appropriate approximated expression and L are chosen based on the range of $g(r)$. We also investigated finite-size effects of the RDF in computing S^{ex} from MC simulations using the WF potential. We found that S^{ex} computed by integrating RDF corrected to the thermodynamic limit ($g^\infty(r)$) agrees with S^{ex} computed from thermodynamic integration (S_{T}^{ex}) for both Lennard-Jones-like ($r_c = 2$) and colloid-like ($r_c = 1.2$) particles at low densities. This agreement holds for systems with 100 and 500 particles. We noticed that S_{T}^{ex} computed by thermodynamic integration showed no significant difference in the values of S^{ex} for different system sizes. S^{ex} computed by integrating the standard RDF ($g(r)$) shows significant discrepancies at low densities for a system with 100 particles (for both $r_c = 2$ and $r_c = 1.2$). For

a system size of 500 particles, S^{ex} computed by integrating $g(r)$ showed minor discrepancies at extremely low densities for both $r_c = 2$ and $r_c = 1.2$ suggesting that $g^\infty(r)$ should always be used in the computation of S^{ex} . At high densities ($\rho\sigma^3 > 0.1$), S^{ex} computed by integrating RDFs ($g^\infty(r)$ and $g(r)$) yields identical values. Comparing $S_{\text{II}}^{\text{ex}}$ with S^{ex} computed from RDFs shows significant differences for $\rho\sigma^3 > 0.1$. Discrepancies at high densities are due to the second-order approximation of the excess entropy integral (Eq. 5.1) used for computing S^{ex} from MC simulations. The computation of S^{ex} using Eq. 5.1 truncated to a finite size and $g^\infty(r)$ captures 95% of the S^{ex} for $\rho\sigma^3 < 0.1$. This level of accuracy in S^{ex} computation holds for both Lennard-Jones-like and colloid-like particles for $\rho\sigma^3 < 0.1$, even for a small system size of 100 particles. Our simulation results indicate that accurate estimations of S^{ex} can be obtained from Eq. 5.1 and TI for $\rho\sigma^3 < 0.1$ for a system with 100 particles. The computation of S^{ex} using Eq. 5.1 and $g^\infty(r)$ can result in errors of 20% at high densities, regardless of r_c . A summary of the comparison and observations related to the excess entropy computation investigated in this chapter is presented in Table 5.7. Our approach enables the efficient and computationally inexpensive computation of S^{ex} by addressing the underlying approximations.

6

Conclusions

Large-scale deployment of Carbon Capture, Utilization, and Storage (CCUS) requires the safe and efficient transport of captured CO_2 by pipeline, ship, train, or truck. The captured streams, however, are never pure. Depending on the capture technology, the CO_2 stream contains impurities that significantly alter its thermodynamic, transport, and interfacial properties. These changes in turn affect the design, capacity, and safe operation of the CO_2 transport and storage infrastructure. Experimental data for impure CO_2 streams at transport-relevant conditions are scarce. Measurements covering the vast range of compositions, temperatures, and pressures are impractical, time-consuming, and costly. Engineering models, meanwhile, require property evaluations that are both accurate and sufficiently fast for repeated use in process and safety simulations. This thesis addresses this issue by developing accurate and computationally efficient descriptions of the thermophysical properties of impure CO_2 streams using molecular simulations, thermodynamic modeling, and Machine Learning (ML). Force field-based Monte Carlo (MC) and Molecular Dynamics (MD) simulations were first validated with reference Equations of State (EoSs) and available experimental data. The validated force fields were then used to generate a large thermophysical-property dataset, spanning pure CO_2 to multicomponent mixtures at conditions relevant to pipeline and modular transport. The results show that impurities de-densify the stream, lowering the amount of CO_2 that can be transported for a given volumetric flow rate. Impurities also modify the heat capacity, Joule–Thomson coefficient, speed of sound, and viscosity. These properties govern the temperature response and the propagation of pressure and decompression waves during compression, expansion, and depressurization. Phase changes during fast transient events, such as depressurization and accidental release, cannot be described without accounting for the delayed onset of nucleation. The fluid can temporarily persist in a metastable state until a critical nucleus is formed. To characterize this behavior, MD simulations and the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) EoS combined with

classical Density Functional Theory (cDFT) were used to compute the phase equilibria, the Interfacial Tension (IFT), and the interfacial structure of CO₂-rich mixtures. The computed interfacial properties were then combined with Classical Nucleation Theory (CNT) to estimate the metastability limits of the liquid and vapor phase of binary CO₂-rich mixtures. These limits define how far a stream can be expanded beyond its equilibrium phase boundary before vaporization or condensation begins. The superheating limit marks the onset of vapor-bubble nucleation in a metastable liquid, whereas the supercooling limit marks the onset of liquid-droplet nucleation in a metastable vapor. The PC-SAFT + cDFT framework reproduced the MD results with comparable accuracy at a considerably lower computational cost. The results show that impurities reduce the IFT and increase the interfacial thickness, thereby influencing the nucleation behavior. Physics-based computations remain too slow for direct use in flow simulations. ML surrogate models were therefore trained on a large dataset of bubble- and dew-point pressures, IFTs, and interfacial thicknesses of multicomponent CO₂-rich mixtures, computed using the Perturbed-Chain Polar Statistical Associating Fluid Theory (PCP-SAFT) EoS combined with cDFT. Of the fully data-driven models, Tabular Prior-data Fitted Network (TabPFN) combined with Hyperparameter Optimization (HPO) provided the best balance between accuracy and robustness, and Gaussian Process Regression (GPR)-based active learning proved the most data-efficient sampling strategy. The highest accuracy for the IFT was achieved by a physics-informed hybrid model, in which the semi-empirical Winterfeld–Scriven–Davis (WSD) model provides a physical baseline and an ML model learns its residual. The trained surrogates predict a state point within a few milliseconds. This speed-up of several orders of magnitude, obtained with negligible loss of accuracy, enables the use of these surrogates in multiphase Computational Fluid Dynamics (CFD) simulations and transient pipeline simulation tools. Finally, the accurate and computationally efficient estimation of the excess entropy, a key input for entropy-scaling predictions of transport properties, was investigated using the straightforward two-body density expansion of entropy. The results show that simple truncation of the excess entropy integral at a finite box length converges fastest to the thermodynamic limit, whereas the Kirkwood–Buff (KB) integrals, which have a similar form, are best obtained from a first-order extrapolation. Finite-size corrected Radial Distribution Functions (RDFs) are required irrespective of the system size. The two-body expansion remains a low-density approximation, and thermodynamic integration or higher-order density expansions are preferred at higher densities. Within these limits, the excess entropy can be estimated accurately and inexpensively from small systems.

Overall, this thesis bridges the gap between molecular-level insight and process-scale simulation by translating force field-based simulations into an accurate and computationally affordable description of impure CO₂

streams. It delivers validated force fields for CO_2 and its main noncondensable impurities N_2 , Ar, H_2 , and CH_4 , together with thermodynamic and transport property data for the unary, binary, ternary, and quaternary mixtures of these components. It further provides the interfacial properties and metastability limits of binary CO_2 -rich mixtures, fast ML surrogates for the phase and interfacial properties of multicomponent CO_2 -rich mixtures, and practical guidelines for computing the excess entropy from molecular simulations. These outputs can be used directly in process, safety, and CFD simulations and thereby support the large-scale deployment of CO_2 transport and storage. The present work also has limitations. Water and reactive species such as SO_2 and NO_2 were not considered and remain to be included. The metastability limits assume homogeneous nucleation and rely on the capillarity approximation, in which the IFT of the planar interface is also used for the curved nucleus [237]. Curvature corrections, described by the Tolman length and the bending rigidity constants, were therefore neglected, although these corrections become significant for nuclei of only a few nanometers [368–370]. The ML surrogates are reliable only within the range of their training data. These limitations point to several research directions that build naturally on this foundation. The most immediate is the prediction of the transport properties of impure CO_2 mixtures. Entropy scaling provides a single framework in which the viscosity and the thermal conductivity follow from the same excess entropy that the EoS already delivers [74, 75, 142]. A separate empirical correlation for every property, component, and composition is then no longer needed. The thermal conductivity remains the more demanding target, because equilibrium MD relies on the slowly decaying and inherently noisy heat-flux autocorrelation function and lacks an established finite-size correction [371]. Thermal diffusion falls outside the scope of this thesis and remains unexplored, although the Soret effect couples temperature and composition gradients in buried pipelines, across throttling valves, and in the near-wellbore region. Moreover, the thermodiffusion coefficients of CO_2 mixtures show an extremum near the Widom line, exactly where dense-phase pipelines operate [372]. Nonequilibrium thermodynamics describes this coupling between heat and mass fluxes, together with the transfer resistance of the vapor–liquid interface itself [373], and nonequilibrium MD is the natural tool for computing the required coefficients. Combined with the interfacial models developed in this thesis, such a description would clarify the interfacial heat and mass transfer during depressurization, vapor collapse, and accidental release. Closely related is the extension of the nucleation framework from homogeneous to heterogeneous nucleation. In a real pipeline, the pipe wall, weld seams, corrosion products, and suspended particles provide surfaces on which nuclei form at a much lower energy barrier. Incorporating the contact angle and the surface energy of these substrates into the CNT framework would show how much of the metastable

region survives under realistic operating conditions. Another direction concerns chemically reactive impurities. Realistic captured streams contain H_2O , O_2 , SO_2 , NO_2 , and H_2S , which react with each other even at parts-per-million levels. Nitrogen dioxide catalyzes the oxidation of SO_2 , and both species combine with water to form nitric and sulfuric acid. These acids condense as a separate aqueous phase, attack the pipe wall, and shift the damage from uniform corrosion to localized pitting [48–50]. Whether such an acid phase appears within the residence time of a pipeline is a question of reaction kinetics rather than of equilibrium alone. Reaction ensemble MC simulations can provide reaction equilibria [161], whereas the rate constants require reactive force fields or first-principles methods. Predicting the composition of the condensed phase and the interfacial tension between an acid droplet and the CO_2 -rich phase is a direct extension of the framework developed here. Water deserves particular attention, because a free aqueous phase also enables gas hydrates, which can block valves, small-bore instrument lines, and injection wells at temperatures well above 273.15 K [51, 52]. The hydrate equilibrium curve of an impure stream is therefore an operating limit in the same sense as the bubble-point curve. Molecular simulation can supply the water solubility and hydrate stability data that remain scarce. Once the thermodynamic, transport, and interfacial properties of impure CO_2 are available for a sufficiently wide range of conditions, the data can be regressed into a dedicated mixture EoS, following the route already demonstrated for model fluids [374]. Such an EoS could describe the solid, liquid, vapor, and supercritical regions in a single formulation, which the mixture models used at present do not. The solid region is precisely where dry-ice formation during depressurization must be predicted. The ML surrogates can then be retrained on this broader property set and extended to transport properties and trace impurities, widening the applicability to realistic captured streams. The practical value of all these models is realized when the models are embedded in the simulators used in industry. Transient multiphase flow simulators such as Olga™ [338] and LedaFlow® [339, 340] evaluate fluid properties at every grid point and at every time step. Fast surrogates or precomputed lookup tables generated with the framework of this thesis can therefore replace the built-in property routines, bringing molecular-level accuracy to pipeline-scale simulations at an acceptable computational cost. The final step is validation with flow-loop experiments. Flow loops reproduce depressurization, phase change, and two-phase flow in a controlled pipe geometry, and provide the pressure, temperature, and phase-fraction data needed to test the property models and the flow solver together [214, 375]. Dosing a flow loop with controlled amounts of impurities would reveal whether the predicted shifts in the phase envelope, the metastability limits, and the interfacial properties translate into measurable changes in flow behavior. Such experiments would connect the molecular simulations of this thesis to the design and operation of CO_2 transport systems.

A

Appendix

A.1. Supporting Information

The Supporting Information for the chapters of this thesis is available online through the associated publications. These materials are extensive, so the full content is not reproduced within the thesis. This improves the readability, reduces the size of the thesis, and helps lower the associated environmental footprint. Chapters 1 and 6 are the introduction and the conclusions of this thesis and have no associated publication. Chapter 5 (Ref. [341]) has no Supporting Information. The simulation details and all numerical results of Ref. [341] are reported in full in the chapter itself. A summary of the Supporting Information of the remaining chapters is summarized below, together with the links through which the full documents can be accessed.

Chapter 2

The Supporting Information of Ref. [147] is organized as follows. Section S1 lists the force field parameters of the molecules used in the MC and MD simulations (Tables S1 to S7). Section S2 shows the VLE curves of the unary systems computed from MC simulations (Table S8 and Figures S1 to S3). Sections S3 and S4 report the VLE data of the unary and the binary systems (Tables S9 to S27). Sections S5 to S10 present the derivations of the heat capacity at constant volume (C_V), the heat capacity at constant pressure (C_P), the isothermal compressibility (β_T), the thermal expansivity (α_P), the speed of sound (c), and the Joule–Thomson coefficient (μ_{JT}). Section S11 describes the computation of the uncertainties Δc and $\Delta\mu_{JT}$. Section S12 provide the ideal gas heat capacity at constant volume, c_V^{ideal} , in Tables S28 to S32. Section S13 shows the plots of the thermodynamic and transport properties of the unary systems (Figures S4 to S23). Sections S14 to S17 tabulate the thermodynamic and transport properties of the unary, binary, ternary, and quaternary systems

(Tables S33 to S458). The full document can be accessed at <https://pubs.acs.org/doi/10.1021/acs.jced.4c00268>.

Chapter 3

The Supporting Information of Ref. [211] is organized as follows. Section S1 lists the force field parameters of the molecules used in the MD simulations (Tables S1 to S6). Section S2 lists the PC-SAFT EoS parameters used in the VLE and interfacial property computations in FeO_s (Table S7). Section S3 presents the empirical correlation model used to compute the IFTs of the pure components (Table S8 and Figure S1). Section S4 reports the VLE of CO_2 computed from MD simulations (Tables S9 to S10). Section S5 describes the computation of the IFT from MD simulations (Figure S2). Section S6 contains the VLE and interfacial property data of pure CO_2 (Tables S11 to S16). Section S7 lists the experimental phase equilibria data of the binary CO_2 mixtures (Tables S17 to S20). Section S8 reports the BIPs used in the PC-SAFT EoS (Tables S21 to S24 and Figures S3 to S15). Section S9 shows the interfacial properties of the binary CO_2 mixtures (Figures S16 to S18). Section S10 reports the phase equilibria and interfacial property data of the binary CO_2 mixtures (Tables S25 to S37). Section S11 contains the SHL and SCL data of pure CO_2 and of the binary CO_2 mixtures (Tables S38 to S44). The full document can be accessed at <https://pubs.acs.org/doi/10.1021/acs.iecr.5c04932>.

Chapter 4

The Supporting Information of Ref. [279] is organized as follows. Section S1 lists the PCP-SAFT EoS parameters of the pure components used to compute the phase envelopes and the interfacial properties of the multicomponent CO_2 mixtures in FeO_s (Table S1). Section S2 lists the temperature-dependent BIPs of the dissimilar molecule pairs used in the PCP-SAFT EoS, together with the polynomial fit expressions, the experimental VLE sources, and the resulting temperature dependence (Tables S2 to S3 and Figures S1 to S7). Section S3 describes the stratified composition space sampling algorithm and the clustering diagnostics underlying its strata (Figures S8 to S11). Section S4 presents the full data-driven ML models. The parity plots of the four target properties obtained with the RF, SVR, XGB, and TabPFN models are shown in Figures S12 to S15. The candidate and the selected hyperparameters are listed in Table S4, and the TabPFN Bayesian optimization search space and the selected values are listed in Table S5. The RMSE sweep curves of the RF, SVR, and XGB models are shown in Figures S16 to S25. Section S5 presents the SR of the IFT. The pure-component IFT correlations are listed in Tables S6 to S7 and Figures S26 to S30. The hybrid cDFT-WSD residual-correction model for the CO_2 -rich mixtures is reported in Tables S8 to S12 and Figures S31 to S32. The full document can be accessed at <https://pubs.acs.org/doi/10.1021/acs.iecr.6c03388>.

A.2. Data Availability

Sample simulation input and output files for each chapter of this thesis, except Chapters 1 and 6, are available in the 4TU.ResearchData repository to ensure the transparency and reproducibility of the presented results. The corresponding DOIs are listed below.

Chapter 2

<https://doi.org/10.4121/2bb3be7b-9e7d-4476-8865-a9754dff1de4>

Chapter 3

<https://doi.org/10.4121/6e9f4e5d-8332-4d4a-ab2a-97efe95ab194>

Chapter 4

The simulation data will be made available with the corresponding publication.

Chapter 5

<https://doi.org/10.4121/4e49f701-89f0-4805-93b2-6dff3d71e05f>

Bibliography

- [1] *Climate Change 2023: Synthesis Report, Summary for Policymakers*, Intergovernmental Panel on Climate Change (IPCC). 2023. Accessed: 9th July 2026. URL: <https://www.ipcc.ch/report/ar6/syr/summary-for-policymakers/>.
- [2] *Climate Change 2022: Mitigation of Climate Change, Working Group III Contribution to the Sixth Assessment Report, Summary for Policymakers*, Intergovernmental Panel on Climate Change (IPCC). 2022. Accessed: 9th July 2026. URL: <https://www.ipcc.ch/report/ar6/wg3/>.
- [3] *Progress in Reducing Emissions: 2025 Report to Parliament, Climate Change Committee*. 2025. Accessed: 9th July 2026. URL: <https://www.theccc.org.uk/publication/progress-in-reducing-emissions-2025-report-to-parliament/>.
- [4] *The EU Has Cut Its Greenhouse Gas Emissions by 40 Percent Since 1990*, European Environment Agency. 2026. Accessed: 9th July 2026. URL: <https://www.eea.europa.eu/en/newsroom/news/the-eu-has-cut-its-greenhouse-gas-emissions-by-40-since-1990>.
- [5] *GHG Emissions of All World Countries, JRC/IEA 2025 Report, EDGAR Emissions Database*, European Commission Joint Research Centre. 2025. Accessed: 9th July 2026. URL: https://edgar.jrc.ec.europa.eu/report_2025.
- [6] *CCUS Projects Around the World Are Reaching New Milestones*, International Energy Agency. 2025. Accessed: 9th July 2026. URL: <https://www.iea.org/commentaries/ccus-projects-around-the-world-are-reaching-new-milestones>.
- [7] M. E. Boot-Handford, J. C. Abanades, E. J. Anthony, M. J. Blunt, S. Brandani, N. Mac Dowell, J. R. Fernández, M.-C. Ferrari, R. Gross, J. P. Hallett, S. R. Haszeldine, P. Heptonstall, A. Lyngfelt, Z. Makuch, E. Mangano, R. T. Porter, M. Pourkashanian, G. T. Rochelle, N. Shah, J. G. Yao and P. S. Fennell. 'Carbon capture and storage update'. *Energy and Environmental Science* 7 (2014), pp. 130–189.
- [8] M. Bui, C. S. Adjiman, A. Bardow, E. J. Anthony, A. Boston, S. Brown, P. S. Fennell, S. Fuss, A. Galindo, L. A. Hackett, J. P. Hallett, H. J. Herzog, G. Jackson, J. Kemper, S. Krevor, G. C. Maitland, M. Matuszewski, I. S. Metcalfe, C. Petit, G. Puxty, J. Reimer, D. M. Reiner, E. S. Rubin, S. A. Scott, N. Shah, B. Smit, J. P. M. Trusler,

- P. Webley, J. Wilcox and N. Mac Dowell. 'Carbon capture and storage (CCS): the way forward'. *Energy and Environmental Science* 11 (2018), pp. 1062–1176.
- [9] T. Wilberforce, A. Olabi, E. T. Sayed, K. Elsaid and M. A. Abdelkareem. 'Progress in carbon capture technologies'. *Science of The Total Environment* 761 (2021), p. 143203.
- [10] E. S. Sanz-Pérez, C. R. Murdock, S. A. Didas and C. W. Jones. 'Direct Capture of CO₂ from Ambient Air'. *Chemical Reviews* 116 (2016), pp. 11840–11876.
- [11] O. Al Yafiee, F. Mumtaz, P. Kumari, G. N. Karanikolos, A. Decarlis and L. F. Dumée. 'Direct air capture (DAC) vs. Direct ocean capture (DOC)—A perspective on scale-up demonstrations and environmental relevance to sustain decarbonization'. *Chemical Engineering Journal* 497 (2024), p. 154421.
- [12] *CCUS Projects Database*, International Energy Agency. 2025. Accessed: 9th July 2026. URL: <https://www.iea.org/data-and-statistics/data-product/ccus-projects-database>.
- [13] *Commercial Carbon Capture Plants: International Operating CCS Projects*, International CCS Knowledge Centre. 2025. Accessed: 9th July 2026. URL: <https://ccsknowledge.com/carbon-management-101/international-ccs-projects/>.
- [14] A. C. Jones and D. J. Marples. *The Section 45Q Tax Credit for Carbon Sequestration*. Tech. rep. Washington DC, USA: Congressional Research Service, 2023. Accessed: 28th July 2026. URL: <https://www.congress.gov/crs-product/IF11455>.
- [15] *Carbon Capture and Storage in Europe: Slow but Significant Progress in 2025*, Clean Air Task Force. 2025. Accessed: 9th July 2026. URL: <https://www.catf.us/2025/11/carbon-capture-storage-europe-slow-but-significant-progress-2025/>.
- [16] *The Northern Lights Project*, Equinor. 2025. Accessed: 9th July 2026. URL: <https://www.equinor.com/energy/northern-lights>.
- [17] Northern Lights. *About the Longship Project*. 2026. Accessed: 12th May 2026. URL: <https://norlights.com/about-the-longship-project/>.
- [18] *The Porthos Project*, Porthos CO₂ Transport and Storage. 2026. Accessed: 28th July 2026. URL: <https://www.porthosco2.nl/en/project/>.
- [19] Porthos. *Porthos CO₂ Specifications 2021*. 2021. Accessed: 25th August 2025. URL: <https://www.porthosco2.nl/wp-content/uploads/2021/09/CO2-specifications.pdf>.

- [20] *Aramis: Large-Scale CO₂ Transport and Storage Infrastructure*, Aramis CCS. 2025. Accessed: 9th July 2026. URL: <https://www.aramis-ccs.com/>.
- [21] Greensand Future. *Greensand Future: Offshore CO₂ Storage in Denmark*. 2025. Accessed: 13th July 2026. URL: <https://greensandfuture.com/>.
- [22] Department for Energy Security and Net Zero. *Track-1 Carbon Capture, Usage and Storage Programme: HyNet and East Coast Cluster*. 2026. Accessed: 13th July 2026. URL: <https://www.gov.uk/government/publications/department-for-energy-security-and-net-zero-major-projects-accounting-officer-assessments/track-1-carbon-capture-usage-and-storage-programme-hynet-and-east-coast-cluster-accounting-officer-assessment-2024>.
- [23] European Union. *Regulation (EU) 2024/1735 of the European Parliament and of the Council of 13 June 2024 on Establishing a Framework of Measures for Strengthening Europe's Net-Zero Technology Manufacturing Ecosystem*. 2024. Accessed: 28th July 2026. URL: <https://eur-lex.europa.eu/eli/reg/2024/1735/oj>.
- [24] *Direct Air Capture Project Awarded Funding under Government Plans to Make the UK a World Leader in Greenhouse Gas Removals*. By Storegga. 2021. Accessed: 27th July 2026. URL: <https://storegga.earth/news/2021/news/direct-air-capture-project-awarded-funding-under-government-plans-to-make-uk-world-leader-in-greenhouse-gas-removals>.
- [25] *Sizewell C and Partners Awarded Direct Air Capture Funding*. By Sizewell C. 2022. Accessed: 27th July 2026. URL: <https://www.sizewellc.com/news-views/sizewell-c-and-partners-awarded-dac-funding/>.
- [26] *Agreement with Carbon Removal for Large-Scale DAC in Norway*. By Phlair. 2025. Accessed: 27th July 2026. URL: <https://phlair.com/news/carbon-removal-nordac>.
- [27] K. R. Simonsen, D. S. Hansen and S. Pedersen. 'Challenges in CO₂ transportation: Trends and perspectives'. *Renewable and Sustainable Energy Reviews* 191 (2024), pp. 114149–114149.
- [28] R. Span and W. Wagner. 'A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa'. *Journal of Physical and Chemical Reference Data* 25 (1996), pp. 1509–1596.
- [29] M. Raju, D. T. Banuti, P. C. Ma and M. Ihme. 'Widom Lines in Binary Mixtures of Supercritical Fluids'. *Scientific Reports* 7 (2017), p. 3027.

- [30] L. Xu, P. Kumar, S. V. Buldyrev, S.-H. Chen, P. H. Poole, F. Sciortino and H. E. Stanley. 'Relation between the Widom line and the dynamic crossover in systems with a liquid-liquid phase transition'. *Proceedings of the National Academy of Sciences* 102 (2005), pp. 16558–16562.
- [31] G. G. Simeoni, T. Bryk, F. A. Gorelli, M. Krisch, G. Ruocco, M. Santoro and T. Scopigno. 'The Widom line as the crossover between liquid-like and gas-like behaviour in supercritical fluids'. *Nature Physics* 6 (2010), pp. 503–507.
- [32] D. T. Banuti. 'Crossing the Widom-line – Supercritical pseudo-boiling'. *Journal of Supercritical Fluids* 98 (2015), pp. 12–16.
- [33] F. Maxim, C. Contescu, P. Boillat, B. Niceno, K. Karalis, A. Testino and C. Ludwig. 'Visualization of supercritical water pseudo-boiling at Widom line crossover'. *Nature Communications* 10 (2019), p. 4114.
- [34] F. Gorelli, M. Santoro, T. Scopigno, M. Krisch and G. Ruocco. 'Liquidlike behavior of supercritical fluids'. *Physical Review Letters* 97 (2006), p. 245702.
- [35] S. T. Munkejord, M. Hammer and S. W. Løvseth. 'CO₂ transport: Data and models—A review'. *Applied Energy* 169 (2016), pp. 499–523.
- [36] G. M. Kontogeorgis, R. Dohrn, I. G. Economou, J.-C. de Hemptinne, A. Ten Kate, S. Kuitunen, M. Mooijer, L. F. Žilnik and V. Vesović. 'Industrial requirements for thermodynamic and transport properties: 2020'. *Industrial & Engineering Chemistry Research* 60 (2021), pp. 4987–5013.
- [37] V. E. Onyebuchi, A. Kolios, D. P. Hanak, C. Biliyok and V. Manovic. 'A systematic review of key challenges of CO₂ transport via pipelines'. *Renewable and Sustainable Energy Reviews* 81 (2018), pp. 2563–2583.
- [38] G. Chinello, Y. Arellano, R. Span, D. van Putten, A. Abdulrahman, E. Joonaki, K. Arrhenius and A. Murugan. 'Toward Standardized Measurement of CO₂ Transfer in the CCS Chain'. *Nexus* 1 (2024), p. 100013.
- [39] J. Jimba, G. Chinello, R. Brown, S. Higgins and M. M. Maroto-Valer. 'Assessing Coriolis Meter Performance in Multicomponent Carbon Dioxide-Rich Mixtures'. *International Journal of Greenhouse Gas Control* 136 (2024), p. 104191.
- [40] S. T. Munkejord, J. P. Jakobsen, A. Austegard and M. J. Mølnvik. 'Thermo- and Fluid-Dynamical Modelling of Two-Phase Multi-Component Carbon Dioxide Mixtures'. *International Journal of Greenhouse Gas Control* 4 (2010), pp. 589–596.

- [41] M. Hammer, H. Deng, L. Liu, M. Langsholt and S. T. Munkejord. 'Upward and Downward Two-Phase Flow of CO₂ in a Pipe: Comparison between Experimental Data and Model Predictions'. *International Journal of Multiphase Flow* 138 (2021), p. 103590.
- [42] B. Tavagh Mohammadi, A. Jahanbani Ghahfarokhi and A.-A. Grimstad. 'A Review of Modeling Approaches for CO₂ Injection into Depleted Gas Reservoirs: Coupling Transient Wellbore and Reservoir Dynamics'. *International Journal of Greenhouse Gas Control* 145 (2025), p. 104406.
- [43] E. de Visser, C. Hendriks, M. Barrio, M. J. Mølnvik, G. de Koeijer, S. Liljemark and Y. Le Gallo. 'Dynamis CO₂ quality recommendations'. *International Journal of Greenhouse Gas Control* 2 (2008), pp. 478–484.
- [44] H. Li, J. P. Jakobsen, Ø. Wilhelmsen and J. Yan. 'PVTxy properties of CO₂ mixtures relevant for CO₂ capture, transport and storage: Review of available experimental data and theoretical models'. *Applied Energy* 88 (2011), pp. 3567–3579.
- [45] G. T. Rochelle. 'Amine Scrubbing for CO₂ Capture'. *Science* 325 (2009), pp. 1652–1654.
- [46] P. Shirley and P. Myles. *Quality Guidelines for Energy System Studies: CO₂ Impurity Design Parameters*. Tech. rep. 2019. Accessed: 10th October 2023. URL: <https://www.osti.gov/servlets/purl/1566771>.
- [47] P. A. Brownsort. *1st Report of the Thematic Working Group on: CO₂ transport, storage and networks. EU CCUS PROJECTS NETWORK (No ENER/C2/2017-65/SI2. 793333)*. 2019. Accessed: 22nd March 2024. URL: https://www.ccusnetwork.eu/sites/default/files/TG3_Briefing-CO2-Specifications-for-Transport.pdf.
- [48] M. Halseid, A. Dugstad and B. Morland. 'Corrosion and Bulk Phase Reactions in CO₂ Transport Pipelines with Impurities: Review of Recent Published Studies'. *Energy Procedia* 63 (2014), pp. 2557–2569.
- [49] B. H. Mørland, A. Dugstad and G. Svenningsen. 'Corrosion of Carbon Steel in Dense Phase CO₂ with Water Above and Below the Solubility Limit'. *Energy Procedia* 114 (2017), pp. 6752–6765.
- [50] L. Q. Hoa, R. Baessler and D. Bettge. 'On the Corrosion Mechanism of CO₂ Transport Pipeline Steel Caused by Condensate: Synergistic Effects of NO₂ and SO₂'. *Materials* 12 (2019), p. 364.
- [51] A. Chapoy, R. Burgass, B. Tohidi and I. Alsiyabi. 'Hydrate and Phase Behavior Modeling in CO₂-Rich Pipelines'. *Journal of Chemical & Engineering Data* 60 (2015), pp. 447–453.

- [52] R. Burgass, A. Chapoy and X. Li. 'Direct Measurement of Hydrate Equilibrium Temperature in CO₂ and CO₂-Rich Fluids with Low Water Content'. *Fluid Phase Equilibria* 581 (2024), p. 114063.
- [53] A. Chapoy, M. Nazeri, M. Kapateh, R. Burgass, C. Coquelet and B. Tohidi. 'Effect of impurities on thermophysical properties and phase behaviour of a CO₂-rich system in CCS'. *International Journal of Greenhouse Gas Control* 19 (2013), pp. 92–100.
- [54] P. Aursand, M. Hammer, S. T. Munkejord and Ø. Wilhelmsen. 'Pipeline Transport of CO₂ Mixtures: Models for Transient Simulation'. *International Journal of Greenhouse Gas Control* 15 (2013), pp. 174–185.
- [55] R. A. W. M. Henkes, J. E. Ellepola and A. Rao. 'Simulations of fast transients for two-phase CO₂ flow in pipes'. *ISAVFT International Conference on Multiphase Production Technology*. ISAVFT. 2025, ISAVFT-2025–471. Accessed: 9th June 2026. URL: <https://onepetro.org/isavftmpt/proceedings/MPT25/MPT25/ISAVFT-2025-471/795645>.
- [56] A. M. Log. 'Depressurization of CO₂ in pipes: Analysis of experiments and non-equilibrium flashing flow models'. PhD thesis. Norwegian University of Science and Technology (NTNU), 2023. Accessed: 27th July 2026. URL: <https://hdl.handle.net/11250/3107027>.
- [57] M. Hammer, A. M. Log, H. Deng, A. Austegard and S. T. Munkejord. 'Decompression-induced condensation of carbon dioxide: Experiments, and prediction of the supercooling limit using classical nucleation theory'. *Chemical Engineering Science* 309 (2025), p. 121415.
- [58] O. Kunz and W. Wagner. 'The GERG-2008 wide-range equation of state for natural gases and other mixtures: an expansion of GERG-2004'. *Journal of Chemical and Engineering Data* 57 (2012), pp. 3032–3091.
- [59] J. Gernert and R. Span. 'EOS-CG: A Helmholtz energy mixture model for humid gases and CCS mixtures'. *Journal of Chemical Thermodynamics* 93 (2016), pp. 274–293.
- [60] Ø. Wilhelmsen, A. Aasen, G. Skaugen, P. Aursand, A. Austegard, E. Aursand, M. A. Gjennestad, H. Lund, G. Linga and M. Hammer. 'Thermodynamic Modeling with Equations of State: Present Challenges with Established Methods'. *Industrial and Engineering Chemistry Research* 56 (2017), pp. 3503–3515.
- [61] A. R. H. Goodwin, J. V. Sengers and C. J. Peters, eds. *Applied Thermodynamics of Fluids*. 1st ed. Cambridge, UK: Royal Society of Chemistry, 2010.

- [62] A. Laesecke and C. D. Muzny. 'Reference correlation for the viscosity of carbon dioxide'. *Journal of Physical and Chemical Reference Data* 46 (2017), p. 013107.
- [63] Á. Mulero, I. Cachadiña and M. I. Parra. 'Recommended Correlations for the Surface Tension of Common Fluids'. *Journal of Physical and Chemical Reference Data* 41 (2012), p. 043105.
- [64] V. Vesovic, W. Wakeham, G. Olchow, J. Sengers, J. Watson and J. Millat. 'The transport properties of carbon dioxide'. *Journal of Physical and Chemical Reference Data* 19 (1990), pp. 763–808.
- [65] D.-Y. Peng and D. B. Robinson. 'A new two-constant equation of state'. *Industrial & Engineering Chemistry Fundamentals* 15 (1976), pp. 59–64.
- [66] O. Redlich and J. N. S. Kwong. 'On the thermodynamics of solutions. V. An equation of state. Fugacities of gaseous solutions'. *Chemical Reviews* 44 (1949), pp. 233–244.
- [67] G. Soave. 'Equilibrium constants from a modified Redlich-Kwong equation of state'. *Chemical Engineering Science* 27 (1972), pp. 1197–1203.
- [68] J. Gross and G. Sadowski. 'Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules'. *Industrial & Engineering Chemistry Research* 40 (2001), pp. 1244–1260.
- [69] T. Lafitte, A. Apostolakou, C. Avendaño, A. Galindo, C. S. Adjiman, E. A. Müller and G. Jackson. 'Accurate statistical associating fluid theory for chain molecules formed from Mie segments'. *Journal of Chemical Physics* 139 (2013), p. 154504.
- [70] P. Rehner. 'Interfacial properties using classical density functional theory: curved interfaces and surfactants'. PhD thesis. University of Stuttgart, Institute of Thermodynamics and Thermal Process Engineering, 2021. Accessed: 27th July 2026. URL: <https://elib.uni-stuttgart.de/server/api/core/bitstreams/2de1098b-4778-4245-8fe6-cb8d0a77deaf/content>.
- [71] J. Mairhofer and J. Gross. 'Numerical aspects of classical density functional theory for one-dimensional vapor-liquid interfaces'. *Fluid Phase Equilibria* 444 (2017), pp. 1–12.
- [72] J. Mairhofer and J. Gross. 'Modeling of interfacial properties of multicomponent systems using density gradient theory and PCP-SAFT'. *Fluid Phase Equilibria* 439 (2017), pp. 31–42.
- [73] S. E. Quiñones-Cisneros, C. K. Zéberg-Mikkelsen and E. H. Stenby. 'One parameter friction theory models for viscosity'. *Fluid Phase Equilibria* 178 (2001), pp. 1–16.

- [74] J. C. Dyre. 'Perspective: Excess-entropy scaling'. *Journal of Chemical Physics* 149 (2018), p. 210901.
- [75] I. H. Bell, R. Messerly, M. Thol, L. Costigliola and J. C. Dyre. 'Modified entropy scaling of the transport properties of the Lennard-Jones fluid'. *Journal of Physical Chemistry B* 123 (2019), pp. 6345–6363.
- [76] J. F. Ely and H. Hanley. 'Prediction of transport properties. 1. Viscosity of fluids and mixtures'. *Industrial & Engineering Chemistry Fundamentals* 20 (1981), pp. 323–332.
- [77] J. F. Ely and H. Hanley. 'Prediction of transport properties. 2. Thermal conductivity of pure fluids and mixtures'. *Industrial & Engineering Chemistry Fundamentals* 22 (1983), pp. 90–97.
- [78] M. L. Huber, A. Laesecke and R. A. Perkins. 'Model for the viscosity and thermal conductivity of refrigerants, including a new correlation for the viscosity of R134a'. *Industrial & Engineering Chemistry Research* 42 (2003), pp. 3163–3178.
- [79] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. 3rd ed. San Diego, CA, USA: Academic Press, Elsevier, 2023.
- [80] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. 2nd ed. Oxford, UK: Oxford University Press, 2017.
- [81] M. G. Martin and J. I. Siepmann. 'Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes'. *Journal of Physical Chemistry B* 102 (1998), pp. 2569–2577.
- [82] J. J. Potoff and J. I. Siepmann. 'Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen'. *AIChE Journal* 47 (2001), pp. 1676–1682.
- [83] D. Marx and J. Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. 1st ed. Cambridge, UK: Cambridge University Press, 2009.
- [84] J. Behler and M. Parrinello. 'Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces'. *Physical Review Letters* 98 (2007), p. 146401.
- [85] V. L. Deringer, M. A. Caro and G. Csányi. 'Machine Learning Interatomic Potentials as Emerging Tools for Materials Science'. *Advanced Materials* 31 (2019), p. 1902765.
- [86] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller. 'Machine Learning Force Fields'. *Chemical Reviews* 121 (2021), pp. 10142–10186.

- [87] R. Mathur, M. C. Muniz, S. Yue, R. Car and A. Z. Panagiotopoulos. 'First-principles-based machine learning models for phase behavior and transport properties of CO₂'. *Journal of Physical Chemistry B* 127 (2023), pp. 4562–4569.
- [88] D. M. Anstine and O. Isayev. 'Machine Learning Interatomic Potentials and Long-Range Physics'. *Journal of Physical Chemistry A* 127 (2023), pp. 2417–2431.
- [89] J. D. van der Waals. 'Over de Continuïteit van den Gas- en Vloeistof-toestand'. PhD thesis. Leiden, The Netherlands: Leiden University, 1873. Accessed: 22nd September 2026. URL: https://www.lorentz.leidenuniv.nl/history/proefschriften/sources/vanderWaals_1873.pdf.
- [90] G. M. Kontogeorgis, R. Privat and J.-N. Jaubert. 'Taking another look at the van der Waals equation of state—almost 150 years later'. *Journal of Chemical & Engineering Data* 64 (2019), pp. 4619–4637.
- [91] S. Lasala, P. Chiesa, R. Privat and J.-N. Jaubert. 'VLE properties of CO₂-based binary systems containing N₂, O₂ and Ar: experimental measurements and modelling results with advanced cubic equations of state'. *Fluid Phase Equilibria* 428 (2016), pp. 18–31.
- [92] A. Péneloux, E. Rauzy and R. Fréze. 'A consistent correction for Redlich-Kwong-Soave volumes'. *Fluid Phase Equilibria* 8 (1982), pp. 7–23.
- [93] N. I. Diamantonis and I. G. Economou. 'Evaluation of statistical associating fluid theory (SAFT) and perturbed chain-SAFT equations of state for the calculation of thermodynamic derivative properties of fluids related to carbon capture and sequestration'. *Energy & Fuels* 25 (2011), pp. 3334–3343.
- [94] M. S. Shin, Y. Lee and H. Kim. 'Estimation of second-order derivative thermodynamic properties using the crossover cubic equation of state'. *Journal of Chemical Thermodynamics* 40 (2008), pp. 688–694.
- [95] N. I. Diamantonis, G. C. Boulougouris, E. Mansoor, D. M. Tsangaris and I. G. Economou. 'Evaluation of cubic, SAFT, and PC-SAFT equations of state for the vapor-liquid equilibrium modeling of CO₂ mixtures with other gases'. *Industrial & Engineering Chemistry Research* 52 (2013), pp. 3933–3942.
- [96] J. Buchenfeld, M. Thol and R. Span. 'Classical Ideal Curves of Fluid Mixtures and Their Role for the Development of Equations of State'. *International Journal of Thermophysics* 47 (2026), p. 64.
- [97] M. S. Wertheim. 'Fluids with Highly Directional Attractive Forces. I. Statistical Thermodynamics'. *Journal of Statistical Physics* 35 (1984), pp. 19–34.

- [98] M. S. Wertheim. 'Fluids with Highly Directional Attractive Forces. II. Thermodynamic Perturbation Theory and Integral Equations'. *Journal of Statistical Physics* 35 (1984), pp. 35–47.
- [99] M. S. Wertheim. 'Fluids with Highly Directional Attractive Forces. III. Multiple Attraction Sites'. *Journal of Statistical Physics* 42 (1986), pp. 459–476.
- [100] M. S. Wertheim. 'Fluids with Highly Directional Attractive Forces. IV. Equilibrium Polymerization'. *Journal of Statistical Physics* 42 (1986), pp. 477–492.
- [101] T. L. Hill. *Statistical Mechanics: Principles and Selected Applications*. 1st ed. New York, USA: McGraw-Hill, 1956.
- [102] J. Gross and G. Sadowski. 'Application of the perturbed-chain SAFT equation of state to associating systems'. *Industrial & Engineering Chemistry Research* 41 (2002), pp. 5510–5515.
- [103] J. Gross. 'An Equation-of-State Contribution for Polar Components: Quadrupolar Molecules'. *AIChE Journal* 51 (2005), pp. 2556–2568.
- [104] J. Gross and J. Vrabec. 'An Equation-of-State Contribution for Polar Components: Dipolar Molecules'. *AIChE Journal* 52 (2006), pp. 1194–1204.
- [105] C. Avendano, T. Lafitte, A. Galindo, C. S. Adjiman, G. Jackson and E. A. Müller. 'SAFT- γ force field for the simulation of molecular fluids. 1. A single-site coarse grained model of carbon dioxide'. *Journal of Physical Chemistry B* 115 (2011), pp. 11154–11169.
- [106] W. G. Chapman, K. E. Gubbins, G. Jackson and M. Radosz. 'New Reference Equation of State for Associating Liquids'. *Industrial & Engineering Chemistry Research* 29 (1990), pp. 1709–1721.
- [107] T. van Westen. 'Development of an Equation of State for Nematic Liquid Crystals'. PhD thesis. Delft, The Netherlands: Delft University of Technology, 2015. URL: <https://resolver.tudelft.nl/uuid:23a7e664-c2e5-471e-a768-0e0a1a5ff909>.
- [108] P. Tarazona. 'Free-energy density functional for hard spheres'. *Molecular Physics* 55 (1985), pp. 557–572.
- [109] F. Llovel, A. Galindo, F. J. Blas and G. Jackson. 'Classical density functional theory for the prediction of the surface tension and interfacial properties of fluids mixtures of chain molecules based on the statistical associating fluid theory for potentials of variable range'. *Journal of Chemical Physics* 133 (2010), p. 024704.
- [110] J. Algaba, B. Mendiboure, P. Gomez-Alvarez and F. J. Blas. 'Density functional theory for the prediction of interfacial properties of molecular fluids within the SAFT- γ coarse-grained approach'. *RSC Advances* 12 (2022), pp. 18821–18833.

- [111] J. M. Garrido, A. Mejía, M. M. Pineiro, F. J. Blas and E. A. Müller. 'Interfacial tensions of industrial fluids from a molecular-based square gradient theory'. *AIChE Journal* 62 (2016), pp. 1781–1794.
- [112] C. G. Aimoli, E. J. Maginn and C. R. Abreu. 'Force field comparison and thermodynamic property calculation of supercritical CO₂ and CH₄ using molecular dynamics simulations'. *Fluid Phase Equilibria* 368 (2014), pp. 80–90.
- [113] J. Vrabec, J. Stoll and H. Hasse. 'A set of molecular models for symmetric quadrupolar fluids'. *Journal of Physical Chemistry B* 105 (2001), pp. 12126–12133.
- [114] J. G. Harris and K. H. Yung. 'Carbon dioxide's liquid-vapor coexistence curve and critical properties as predicted by a simple molecular model'. *Journal of Physical Chemistry* 99 (1995), pp. 12021–12024.
- [115] A. Köster, M. Thol and J. Vrabec. 'Molecular models for the hydrogen age: hydrogen, nitrogen, oxygen, argon, and water'. *Journal of Chemical and Engineering Data* 63 (2018), pp. 305–320.
- [116] E. García-Pérez, J. B. Parra, C. O. Ania, D. Dubbeldam, T. J. H. Vlugt, J. M. Castillo, P. J. Merkling and S. Calero. 'Unraveling the argon adsorption processes in MFI-type zeolite'. *Journal of Physical Chemistry C* 112 (2008), pp. 9976–9979.
- [117] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller. 'Equation of state calculations by fast computing machines'. *Journal of Chemical Physics* 21 (1953), pp. 1087–1092.
- [118] M. Lagache, P. Ungerer, A. Boutin and A. Fuchs. 'Prediction of thermodynamic derivative properties of fluids by Monte Carlo simulation'. *Physical Chemistry Chemical Physics* 3 (2001), pp. 4333–4339.
- [119] C. Colina, C. Olivera-Fuentes, F. Siperstein, M. Lisal and K. Gubbins. 'Thermal properties of supercritical carbon dioxide by Monte Carlo simulations'. *Molecular Simulation* 29 (2003), pp. 405–412.
- [120] A. Z. Panagiotopoulos. 'Monte Carlo methods for phase equilibria of fluids'. *Journal of Physics: Condensed Matter* 12 (2000), R25.
- [121] W. Shi and E. J. Maginn. 'Continuous Fractional Component Monte Carlo: an adaptive biasing method for open system atomistic simulations'. *Journal of Chemical Theory and Computation* 3 (2007), pp. 1451–1463.
- [122] W. Shi and E. J. Maginn. 'Improvement in molecule exchange efficiency in Gibbs ensemble Monte Carlo: development and implementation of the Continuous Fractional Component move'. *Journal of Computational Chemistry* 29 (2008), pp. 2520–2530.

- [123] A. Rahbari, R. Hens, M. Ramdin, O. Moulτος, D. Dubbeldam and T. J. H. Vlugt. 'Recent advances in the Continuous Fractional Component Monte Carlo methodology'. *Molecular Simulation* 47 (2021), pp. 804–823.
- [124] R. Hens, A. Rahbari, S. Caro-Ortiz, N. Dawass, M. Erdős, A. Pour-saeidesfahani, H. S. Salehi, A. T. Celebi, M. Ramdin, O. A. Moulτος, D. Dubbeldam and T. J. H. Vlugt. 'Brick-CFCMC: Open source software for Monte Carlo simulations of phase and reaction equilibria using the Continuous Fractional Component Method'. *Journal of Chemical Information and Modeling* 60 (2020), pp. 2678–2682.
- [125] H. M. Polat, H. S. Salehi, R. Hens, D. O. Wasik, A. Rahbari, F. De Meyer, C. Houriez, C. Coquelet, S. Calero, D. Dubbeldam and T. J. H. Vlugt. 'New features of the open source Monte Carlo software Brick-CFCMC: Thermodynamic integration and hybrid trial moves'. *Journal of Chemical Information and Modeling* 61 (2021), pp. 3752–3757.
- [126] E. A. Müller, Å. Ervik and A. Mejía. 'A Guide to Computing Interfacial Properties of Fluids from Molecular Simulations [Article v1.0]'. *Living Journal of Computational Molecular Science* 2 (2020), p. 21385.
- [127] S. Stephan. 'Vapor-liquid interfaces: Molecular simulation, density gradient theory, and experiments'. PhD thesis. Dissertation, Kaiserslautern, Technische Universität Kaiserslautern, 2020. Accessed: 27th July 2026. URL: <https://kluedo.ub.rptu.de/frontdoor/deliver/index/docId/6089/file/Dissertation+Simon+Stephan.pdf>.
- [128] J. G. Kirkwood and F. P. Buff. 'The statistical mechanical theory of surface tension'. *Journal of Chemical Physics* 17 (1949), pp. 338–343.
- [129] J. H. Irving and J. G. Kirkwood. 'The Statistical Mechanical Theory of Transport Processes. IV. The Equations of Hydrodynamics'. *Journal of Chemical Physics* 18 (1950), pp. 817–829.
- [130] J.-P. Hansen and I. R. McDonald. *Theory of Simple Liquids: With Applications to Soft Matter*. 4th ed. Oxford, UK: Academic Press, Elsevier, 2013.
- [131] D. P. Landau and K. Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. 5th ed. Cambridge, UK: Cambridge University Press, 2021.
- [132] M. E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. 2nd ed. Oxford Graduate Texts. Oxford, UK: Oxford University Press, 2023.
- [133] L. Verlet. 'Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules'. *Physical Review* 159 (1967), p. 98.

- [134] S. H. Jamali, L. Wolff, T. M. Becker, M. De Groen, M. Ramdin, R. Hartkamp, A. Bardow, T. J. H. Vlugt and O. A. Moultos. 'OCTP: A tool for on-the-fly calculation of transport properties of fluids with the order-n algorithm in LAMMPS'. *Journal of Chemical Information and Modeling* 59 (2019), pp. 1290–1294.
- [135] F. Müller-Plathe. 'A Simple Nonequilibrium Molecular Dynamics Method for Calculating the Thermal Conductivity'. *Journal of Chemical Physics* 106 (1997), pp. 6082–6085.
- [136] F. Müller-Plathe. 'Reversing the Perturbation in Nonequilibrium Molecular Dynamics: An Easy Way to Calculate the Shear Viscosity of Fluids'. *Physical Review E* 59 (1999), pp. 4894–4898.
- [137] P. Wirnsberger, D. Frenkel and C. Dellago. 'An Enhanced Version of the Heat Exchange Algorithm with Excellent Energy Conservation Properties'. *Journal of Chemical Physics* 143 (2015), p. 124104.
- [138] D. J. Evans and G. P. Morriss. *Statistical Mechanics of Nonequilibrium Liquids*. 2nd ed. Cambridge, UK: Cambridge University Press, 2008.
- [139] B. Hess. 'Determining the Shear Viscosity of Model Liquids from Molecular Dynamics Simulations'. *Journal of Chemical Physics* 116 (2002), pp. 209–217.
- [140] Y. Rosenfeld. 'Relation between the Transport Coefficients and the Internal Entropy of Simple Systems'. *Physical Review A* 15 (1977), pp. 2545–2549.
- [141] I. H. Bell, J. C. Dyre and T. S. Ingebrigtsen. 'Excess-entropy scaling in supercooled binary mixtures'. *Nature Communications* 11 (2020), p. 4300.
- [142] M. Hopp and J. Gross. 'Thermal conductivity from entropy scaling: A group-contribution method'. *Industrial & Engineering Chemistry Research* 58 (2019), pp. 20441–20449.
- [143] B. Bursik, R. Stierle, A. Schlaich, P. Rehner and J. Gross. 'Viscosities of inhomogeneous systems from generalized entropy scaling'. *Physics of Fluids* 36 (2024), p. 042007.
- [144] S. A. Ghaffarizadeh and G. J. Wang. 'A Picture is Worth a Thousand Timesteps: Excess Entropy Scaling for Rapid Estimation of Diffusion Coefficients in Molecular-Dynamics Simulations of Fluids'. *Journal of Chemical Theory and Computation* 20 (2024), pp. 6839–6851.
- [145] A. Ferguson. 'The Parachor and Molecular Volume'. *Nature* 125 (1930), p. 597.
- [146] P. Winterfeld, L. Scriven and H. Davis. 'An approximate theory of interfacial tensions of multicomponent systems: Applications to binary liquid-vapor tensions'. *AIChE Journal* 24 (1978), pp. 1010–1014.

- [147] D. Raju, M. Ramdin and T. J. H. Vlugt. 'Thermophysical Properties and Phase Behavior of CO₂ with Impurities: Insight from Molecular Simulations'. *Journal of Chemical & Engineering Data* 69 (2024), pp. 2735–2755.
- [148] H. Lee, K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barrett, G. Blanco, W. W. L. Cheung, S. L. Connors, F. Denton, A. Diongue-Niang, D. Dodman, M. Garschagen, O. Geden, B. Hayward, C. Jones, F. Jotzo, T. Krug, R. Lasco, J.-Y. Lee, V. Masson-Delmotte, M. Meinshausen, K. Mintenbeck, A. Mokssit, F. E. L. Otto, M. Pathak, A. Pirani, E. Poloczanska, H.-O. Pörtner, A. Revi, D. C. Roberts, J. Romero, J. Roy, A. C. Ruane, J. Skeea, P. R. Shukla, R. Slade, A. Slangen, Y. Sokona, A. A. Sörensson, M. Tignor, D. van Vuuren, Y.-M. Wei, H. Winkler, P. Zhai and Z. Zommers. *Climate change 2023: synthesis report. Contribution of working groups I, II and III to the sixth assessment report of the intergovernmental panel on climate change*. Tech. rep. Intergovernmental Panel on Climate Change, 2023. Accessed: 26th September 2025. URL: https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_FullVolume.pdf.
- [149] C. Olivia, G. Victor and R. Arthur. *CO₂ Emissions in 2022*, IEA. 2023. Accessed: 22nd March 2024. URL: <https://iea.blob.core.windows.net/assets/3c8fa115-35c4-4474-b237-1b00424c8844/CO2Emissionsin2022.pdf>.
- [150] M. Mohitpour, P. Seevam, K. K. Botros, B. Rothwell and C. Ennis. *Pipeline transportation of carbon dioxide containing impurities*. 1st ed. New York, USA: ASME Press, 2012.
- [151] *Statistical Review of World Energy*, Energy Institute. 2023. Accessed: 10th October 2023. URL: https://www.energyinst.org/_data/assets/pdf_file/0004/1055542/EI_Stat_Review_PDF_single_3.pdf.
- [152] A. I. Osman, L. Chen, M. Yang, G. Msigwa, M. Farghali, S. Fawzy, D. W. Rooney and P.-S. Yap. 'Cost, environmental impact, and resilience of renewable energy under a changing climate: a review'. *Environmental Chemistry Letters* 21 (2023), pp. 741–764.
- [153] T. A. Demetriades and R. S. Graham. 'A new equation of state for CCS pipeline transport: Calibration of mixing rules for binary mixtures of CO₂ with N₂, O₂ and H₂'. *Journal of Chemical Thermodynamics* 93 (2016), pp. 294–304.
- [154] J. Serpa, J. Morbee and E. Tzimas. *Technical and economic characteristics of a CO₂ transmission pipeline infrastructure*. Tech. rep. 2011. Accessed: 24th April 2024. URL: <https://publications.jrc.ec.europa.eu/repository/handle/JRC62502>.

- [155] Z. Yang, A. Fahmi, M. Drescher, L. Teberikler, C. Merat, S. Solvang, O. J. Rinde, J. G. Norstrøm, W. Dijkhuizen, T. Haugset, A. Brigadeaue, M. Langsholtf and L. Liu. 'Improved understanding of flow assurance for CO₂ transport and injection'. *Proceedings of the 15th Greenhouse Gas Control Technologies Conference* (2021), pp. 15–18.
- [156] H. Lund, T. Flåtten and S. T. Munkejord. 'Depressurization of carbon dioxide in pipelines—models and methods'. *Energy Procedia* 4 (2011), pp. 2984–2991.
- [157] M. Nazeri, H. Haghighi, C. McKay, D. Erickson and S. Zhai. 'Impact of CO₂ Specifications on Design and Operation Challenges of CO₂ Transport and Storage Systems in CCUS'. 2021, SPE-205472-MS. Accessed: 22nd March 2024. URL: <https://doi.org/10.2118/205472-MS>.
- [158] A. J. Cresswell, R. J. Wheatley, R. D. Wilkinson and R. S. Graham. 'Molecular simulation of the thermophysical properties and phase behaviour of impure CO₂ relevant to CCS'. *Faraday Discussions* 192 (2016), pp. 415–436.
- [159] A. Sacconi and H. Mahgerefteh. 'Modelling start-up injection of CO₂ into highly-depleted gas fields'. *Energy* 191 (2020), p. 116530.
- [160] M. Ramdin, T. M. Becker, S. H. Jamali, M. Wang and T. J. H. Vlugt. 'Computing equation of state parameters of gases from Monte Carlo simulations'. *Fluid Phase Equilibria* 428 (2016), pp. 174–181.
- [161] S. P. Balaji, S. Gangarapu, M. Ramdin, A. Torres-Knoop, H. Zuilhof, E. L. Goetheer, D. Dubbeldam and T. J. H. Vlugt. 'Simulating the reactions of CO₂ in aqueous monoethanolamine solution by reaction ensemble Monte Carlo using the Continuous Fractional Component method'. *Journal of Chemical Theory and Computation* 11 (2015), pp. 2661–2669.
- [162] Y. Lee, M. S. Shin and H. Kim. 'Estimation of 2nd-order derivative thermodynamic properties using the crossover lattice equation of state'. *Journal of Chemical Thermodynamics* 40 (2008), pp. 1580–1587.
- [163] B. Fazelabdolabadi and A. Bahramian. 'Molecular simulation prediction of sound velocity for a binary mixture near miscible conditions'. *Industrial and Engineering Chemistry Research* 49 (2010), pp. 3469–3473.
- [164] B. Fazelabdolabadi and A. Bahramian. 'Prediction of sound velocity and compressibility via molecular simulation at fixed entropy'. *Fluid Phase Equilibria* 293 (2010), pp. 262–264.
- [165] B. S. Shin, W. G. Rho, S.-S. You, J. W. Kang and C. S. Lee. 'Evaluation of thermodynamic models for predicting phase equilibria of CO₂ + impurity binary mixture'. *International Journal of Thermophysics* 39 (2018), p. 44.

- [166] Ø. Wilhelmsen, G. Skaugen, O. Jørstad and H. Li. 'Evaluation of SPUNG* and other equations of state for use in carbon capture and storage modelling'. *Energy Procedia* 23 (2012), pp. 236–245.
- [167] A. Rahbari, J. C. Garcia-Navarro, M. Ramdin, L. J. Van Den Broeke, O. A. Moulτος, D. Dubbeldam and T. J. H. Vlugt. 'Effect of water content on thermodynamic properties of compressed hydrogen'. *Journal of Chemical and Engineering Data* 66 (2021), pp. 2071–2087.
- [168] H. M. Polat, C. van der Geest, F. de Meyer, C. Houriez, T. J. H. Vlugt and O. A. Moulτος. 'Densities, viscosities, and diffusivities of loaded and unloaded aqueous CO₂/H₂S/MDEA mixtures: A molecular dynamics simulation study'. *Fluid Phase Equilibria* 575 (2023), p. 113913.
- [169] H. M. Polat, F. de Meyer, C. Houriez, C. Coquelet, O. A. Moulτος and T. J. H. Vlugt. 'Transport properties of mixtures of acid gases with aqueous monoethanolamine solutions: A molecular dynamics study'. *Fluid Phase Equilibria* 564 (2023), p. 113587.
- [170] M. Gecht, M. Siggel, M. Linke, G. Hummer and J. Köfinger. 'MD-Benchmark: A toolkit to optimize the performance of molecular dynamics simulations'. *Journal of Chemical Physics* 153 (2020), p. 144105.
- [171] G. Ciccotti, C. Dellago, M. Ferrario, E. Hernández and M. Tuckerman. 'Molecular simulations: past, present, and future (a Topical Issue in EPJB)'. *European Physical Journal B* 95 (2022), p. 3.
- [172] A. Bergant and A. R. Simpson. 'Pipeline column separation flow regimes'. *Journal of Hydraulic Engineering* 125 (1999), pp. 835–848.
- [173] I. Al-Siyabi. 'Effect of impurities on CO₂ stream properties'. PhD thesis. Edinburgh, Scotland, UK: Heriot-Watt University, 2013. Accessed: 22nd March 2024. URL: <http://hdl.handle.net/10399/2643>.
- [174] C. Coquelet, P. Stringari, M. Hajiw, A. Gonzalez, L. Pereira, M. Nazeri, R. Burgass and A. Chapoy. 'Transport of CO₂: presentation of new thermophysical property measurements and phase diagrams'. *Energy Procedia* 114 (2017), pp. 6844–6859.
- [175] M. J. Moran, H. N. Shapiro, D. D. Boettner and M. B. Bailey. *Fundamentals of Engineering Thermodynamics*. 10th ed. Hoboken, NJ, USA: John Wiley & Sons, 2020.
- [176] B. E. Poling, J. M. Prausnitz and J. P. O'Connell. *The Properties of Gases and Liquids*. 5th ed. New York, USA: McGraw-Hill, 2001.

- [177] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox. 'Gaussian 09, Revision D. 01, Gaussian, Inc., Wallingford CT'. *Gaussian 09*. Gaussian Inc Wallingford, 2009. Accessed: 27th July 2026. URL: <https://gaussian.com/>.
- [178] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. 2nd ed. New York, USA: John Wiley & Sons, 1985.
- [179] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton. 'LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales'. *Computer Physics Communications* 271 (2022), p. 108171.
- [180] H. Higashi, Y. Iwai, H. Uchida and Y. Arai. 'Diffusion coefficients of aromatic compounds in supercritical carbon dioxide using molecular dynamics simulation'. *Journal of Supercritical Fluids* 13 (1998), pp. 93–97.
- [181] M. E. Perez-Blanco and E. J. Maginn. 'Molecular dynamics simulations of CO₂ at an ionic liquid interface: Adsorption, ordering, and interfacial crossing'. *Journal of Physical Chemistry B* 114 (2010), pp. 11827–11837.
- [182] Z. Zhang and Z. Duan. 'An optimized molecular potential for carbon dioxide'. *Journal of Chemical Physics* 122 (2005), p. 214507.
- [183] R. T. Cygan, V. N. Romanov and E. M. Myshakin. 'Molecular simulation of carbon dioxide capture by montmorillonite using an accurate and flexible force field'. *Journal of Physical Chemistry C* 116 (2012), pp. 13079–13091.
- [184] C. Murthy, S. O'Shea and I. McDonald. 'Electrostatic interactions in molecular crystals: lattice dynamics of solid nitrogen and carbon dioxide'. *Molecular Physics* 50 (1983), pp. 531–541.

- [185] G. Galassi and D. Tildesley. 'Phase diagrams of diatomic molecules using the Gibbs ensemble Monte Carlo method'. *Molecular Simulation* 13 (1994), pp. 11–24.
- [186] R. F. Cracknell. 'Molecular simulation of hydrogen adsorption in graphitic nanofibres'. *Physical Chemistry Chemical Physics* 3 (2001), pp. 2091–2097.
- [187] V. Buch. 'Path integral simulations of mixed para-D₂ and ortho-D₂ clusters: The orientational effects'. *Journal of Chemical Physics* 100 (1994), pp. 7610–7629.
- [188] J. O. Hirschfelder, C. F. Curtiss and R. B. Bird. *Molecular Theory of Gases and Liquids*. 1st ed. New York, USA: John Wiley & Sons, 1954.
- [189] D. Marx and P. Nielaba. 'Path-integral Monte Carlo techniques for rotational motion in two dimensions: Quenched, annealed, and no-spin quantum-statistical averages'. *Physical Review A* 45 (1992), p. 8968.
- [190] E. W. Lemmon, I. H. Bell, M. L. Huber and M. O. McLinden. *NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 10.0*, National Institute of Standards and Technology. 2018. doi: [10.18434/T4/1502528](https://doi.org/10.18434/T4/1502528). Accessed: 27th July 2026. URL: <https://www.nist.gov/srd/refprop>.
- [191] A. Z. Panagiotopoulos. 'Molecular simulation of phase coexistence: Finite-size effects and determination of critical parameters for two- and three-dimensional Lennard-Jones fluids'. *International Journal of Thermophysics* 15 (1994), pp. 1057–1072.
- [192] M. Dinpajooh, P. Bai, D. A. Allan and J. I. Siepmann. 'Accurate and precise determination of critical properties from Gibbs ensemble Monte Carlo simulations'. *Journal of Chemical Physics* 143 (2015).
- [193] A. Poursaeidesfahani, A. Torres-Knoop, D. Dubbeldam and T. J. H. Vlugt. 'Direct free energy calculation in the Continuous Fractional Component Gibbs ensemble'. *Journal of Chemical Theory and Computation* 12 (2016), pp. 1481–1490.
- [194] A. Poursaeidesfahani, A. Rahbari, A. Torres-Knoop, D. Dubbeldam and T. J. H. Vlugt. 'Computation of thermodynamic properties in the Continuous Fractional Component Monte Carlo Gibbs ensemble'. *Molecular Simulation* 43 (2017), pp. 189–195.
- [195] L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez. 'PACKMOL: A package for building initial configurations for molecular dynamics simulations'. *Journal of Computational Chemistry* 30 (2009), pp. 2157–2164.

- [196] P. Agilio, G. Kateryna and G. Zheng. 'fftool v1.2.1: XML force field files'. *fftool v1.2.1*. Zenodo, 2021. doi: [10.5281/zenodo.4701065](https://doi.org/10.5281/zenodo.4701065). Accessed: 27th July 2026. URL: <https://doi.org/10.5281/zenodo.4701065>.
- [197] M. L. Huber, E. W. Lemmon, I. H. Bell and M. O. McLinden. 'The NIST REFPROP database for highly accurate properties of industrially important fluids'. *Industrial & Engineering Chemistry Research* 61 (2022), pp. 15449–15472.
- [198] C. Coquelet, A. Valtz, F. Dieu, D. Richon, P. Arpentinier and F. Lockwood. 'Isothermal P , x , y data for the argon + carbon dioxide system at six temperatures from 233.32 to 299.21 K and pressures up to 14 MPa'. *Fluid Phase Equilibria* 273 (2008), pp. 38–43.
- [199] J. Davalos, W. R. Anderson, R. E. Phelps and A. J. Kidnay. 'Liquid-vapor equilibria at 250.00K for systems containing methane, ethane, and carbon dioxide'. *Journal of Chemical and Engineering Data* 21 (1976), pp. 81–84.
- [200] M. S.-W. Wei, T. S. Brown, A. J. Kidnay and E. D. Sloan. 'Vapor+liquid equilibria for the ternary system methane+ ethane+ carbon dioxide at 230 K and its constituent binaries at temperatures from 207 to 270 K'. *Journal of Chemical and Engineering Data* 40 (1995), pp. 726–731.
- [201] 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'. *Chemical Engineering Science* 36 (1981), pp. 993–1000.
- [202] T. Brown, V. Niesen, E. Sloan and A. Kidnay. 'Vapor-liquid equilibria for the binary systems of nitrogen, carbon dioxide, and n-butane at temperatures from 220 to 344 K.' *Fluid Phase Equilibria* 53 (1989), pp. 7–14.
- [203] C. Cadena, Q. Zhao, R. Q. Snurr and E. J. Maginn. 'Molecular modeling and experimental studies of the thermodynamic and transport properties of pyridinium-based ionic liquids'. *Journal of Physical Chemistry B* 110 (2006), pp. 2821–2832.
- [204] G. Raabe. 'Molecular simulation studies on the thermophysical properties of the refrigerant blend R-445A'. *Journal of Chemical and Engineering Data* 58 (2013), pp. 3470–3476.
- [205] B. Agbodekhe, E. Marin-Rimoldi, Y. Zhang, A. W. Dowling and E. J. Maginn. 'Assessment and Ranking of Difluoromethane (R32) and Pentafluoroethane (R125) Interatomic Potentials Using Several Thermophysical and Transport Properties Across Multiple State Points'. *Journal of Chemical and Engineering Data* 69 (2023), pp. 427–444.

- [206] T. Hulikal Chakrapani, H. Hajibeygi, O. A. Moulton and T. J. H. Vlucht. 'Mutual Diffusivities of Mixtures of Carbon Dioxide and Hydrogen and Their Solubilities in Brine: Insight from Molecular Simulations'. *Industrial & Engineering Chemistry Research* 63 (2024), pp. 10456–10481.
- [207] R. Span, E. W. Lemmon, R. T. Jacobsen, W. Wagner and A. Yokozeki. 'A reference equation of state for the thermodynamic properties of nitrogen for temperatures from 63.151 to 1000 K and pressures to 2200 MPa'. *Journal of Physical and Chemical Reference Data* 29 (2000), pp. 1361–1433.
- [208] C. Tegeler, R. Span and W. Wagner. 'A new equation of state for argon covering the fluid region for temperatures from the melting line to 700 K at pressures up to 1000 MPa'. *Journal of Physical and Chemical Reference Data* 28 (1999), pp. 779–850.
- [209] J. W. Leachman, R. T. Jacobsen, S. Penoncello and E. W. Lemmon. 'Fundamental equations of state for parahydrogen, normal hydrogen, and orthohydrogen'. *Journal of Physical and Chemical Reference Data* 38 (2009), pp. 721–748.
- [210] U. Setzmann and W. Wagner. 'A new equation of state and tables of thermodynamic properties for methane covering the range from the melting line to 625 K at pressures up to 1000 MPa'. *Journal of Physical and Chemical Reference Data* 20 (1991), pp. 1061–1155.
- [211] D. Raju, R. Skartlien, M. Ramdin and T. J. H. Vlucht. 'Vapor-Liquid Interfacial Properties of CO₂ Mixtures for Sequestration Applications: Molecular Simulations, Classical Density Functional Theory, and Equations of State'. *Industrial & Engineering Chemistry Research* 65 (2026), pp. 5213–5238.
- [212] M. W. Jones, G. P. Peters, T. Gasser, R. M. Andrew, C. Schwingshackl, J. Gütschow, R. A. Houghton, P. Friedlingstein, J. Pongratz and C. Le Quéré. 'National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850'. *Scientific Data* 10 (2023), p. 155.
- [213] L. Klinkby, C. M. Nielsen, E. Krogh, I. E. Smith, B. Palm and C. Bernstone. 'Simulating rapidly fluctuating CO₂ flow into the Vedsted CO₂ pipeline, injection well and reservoir'. *Energy Procedia* 4 (2011), pp. 4291–4298.
- [214] S. T. Munkejord and M. Hammer. 'Depressurization of CO₂-rich mixtures in pipes: Two-phase flow modelling and comparison with experiments'. *International Journal of Greenhouse Gas Control* 37 (2015), pp. 398–411.

- [215] E. Hendriks, G. M. Kontogeorgis, R. Dohrn, J.-C. de Hemptinne, I. G. Economou, L. F. Zilnik and V. Vesovic. 'Industrial requirements for thermodynamics and transport properties'. *Industrial & Engineering Chemistry Research* 49 (2010), pp. 11131–11141.
- [216] P. Habibi, J. R. T. Postma, J. T. Padding, P. Dey, T. J. H. Vlugt and O. A. Moultos. 'Thermodynamic and Transport Properties of $\text{H}_2/\text{H}_2\text{O}/\text{NaB}(\text{OH})_4$ Mixtures Using the Delft Force Field ($\text{DFF}/\text{B}(\text{OH})_4^-$)'. *Industrial & Engineering Chemistry Research* 62 (2023), pp. 11992–12005.
- [217] S. E. Quiñones-Cisneros, U. K. Deiters, R. E. Rozas and T. Kraska. 'New model for the correlation of the surface tension based on friction theory'. *Journal of Physical Chemistry B* 113 (2009), pp. 3504–3511.
- [218] C. R. Wilke. 'A viscosity equation for gas mixtures'. *Journal of Chemical Physics* 18 (1950), pp. 517–519.
- [219] E. A. Guggenheim. 'The principle of corresponding states'. *Journal of Chemical Physics* 13 (1945), pp. 253–261.
- [220] Y. T. F. Chow, D. K. Eriksen, A. Galindo, A. J. Haslam, G. Jackson, G. C. Maitland and J. P. M. Trusler. 'Interfacial tensions of systems comprising water, carbon dioxide and diluent gases at high pressures: Experimental measurements and modelling with SAFT-VR Mie and square-gradient theory'. *Fluid Phase Equilibria* 407 (2016), pp. 159–176.
- [221] Y. T. F. Chow, G. C. Maitland, M. S. P. Stevar and J. P. M. Trusler. 'Correction to "Interfacial Tension of (Brines + CO_2): (0.864 NaCl + 0.136 KCl) at Temperatures between (298 and 448) K, Pressures between (2 and 50) MPa, and Total Molalities of (1 to 5) $\text{mol} \cdot \text{kg}^{-1}$ ".'. *Journal of Chemical & Engineering Data* 63 (2018), pp. 2333–2334.
- [222] Y. T. F. Chow, G. C. Maitland and J. P. M. Trusler. 'Interfacial tensions of ($\text{H}_2\text{O} + \text{H}_2$) and ($\text{H}_2\text{O} + \text{CO}_2 + \text{H}_2$) systems at temperatures of (298–448) K and pressures up to 45 MPa'. *Fluid Phase Equilibria* 475 (2018), pp. 37–44.
- [223] P. Lv, M. S. P. Stevar and J. P. M. Trusler. 'Interfacial tensions in the ($\text{CH}_4 + \text{CO}_2 + \text{H}_2\text{O}$) system under two- and three-phase conditions'. *Fluid Phase Equilibria* 522 (2020), p. 112760.
- [224] Y. T. F. Chow, G. C. Maitland and J. P. M. Trusler. 'Interfacial tensions of the ($\text{CO}_2 + \text{N}_2 + \text{H}_2\text{O}$) system at temperatures of (298 to 448) K and pressures up to 40 MPa'. *Journal of Chemical Thermodynamics* 93 (2016), pp. 392–403.
- [225] 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'. *Journal of Chemical Physics* 106 (1997), pp. 4168–4186.

- [226] J. W. Cahn and J. E. Hilliard. 'Free energy of a nonuniform system. I. Interfacial free energy'. *Journal of Chemical Physics* 28 (1958), pp. 258–267.
- [227] J. D. van der Waals. 'The thermodynamic theory of capillarity under the hypothesis of a continuous variation of density'. *Journal of Statistical Physics* 20 (1979), pp. 200–244.
- [228] F. Jiao, S. Z. A. Ghafri, K. N. Seneviratne, M. Akhfarsh, T. J. Hughes, M. L. Johns and E. F. May. 'Interfacial tension measurements of methane + propane binary and methane + propane + n-heptane ternary mixtures at low temperatures'. *Journal of Chemical Thermodynamics* 171 (2022), p. 106786.
- [229] R. Evans. 'The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids'. *Advances in Physics* 28 (1979), pp. 143–200.
- [230] M. Telo da Gama and R. Evans. 'The structure and surface tension of the liquid-vapour interface near the upper critical end point of a binary mixture of Lennard-Jones fluids: I. The two phase region'. *Molecular Physics* 48 (1983), pp. 229–250.
- [231] C. Miqueu, B. Mendiboure, A. Graciaa and J. Lachaise. 'Modelling of the surface tension of pure components with the gradient theory of fluid interfaces: a simple and accurate expression for the influence parameters'. *Fluid Phase Equilibria* 207 (2003), pp. 225–246.
- [232] J. S. Rowlinson and B. Widom. *Molecular theory of capillarity*. 1st ed. Mineola, NY, USA: Dover Publications, Courier Corporation, 2013.
- [233] S. Sugden. 'A relation between surface tension, density, and chemical composition'. *Journal of the Chemical Society, Transactions* 125 (1924), pp. 32–41.
- [234] A. M. Log, V. Diky and M. L. Huber. 'Assessment of a Parachor Model for the Surface Tension of Binary Mixtures'. *International Journal of Thermophysics* 44 (2023), Article 110.
- [235] H. L. Skarsvåg, M. Hammer, S. T. Munkejord, A. M. Log, S. Dumoulin and G. Gruben. 'Towards an engineering tool for the prediction of running ductile fractures in CO₂ pipelines'. *Process Safety and Environmental Protection* 171 (2023), pp. 667–679.
- [236] M. R. Schenk, T. Köddermann, K. N. Kirschner, S. Knauer and D. Reith. 'Molecular dynamics in the energy sector: Experiment and modeling of the CO₂/CH₄ mixture'. *Journal of Chemical & Engineering Data* 65 (2020), pp. 1117–1123.
- [237] P. G. Debenedetti. *Metastable Liquids: Concepts and Principles*. 1st ed. Princeton, NJ: Princeton University Press, 1996.
- [238] V. P. Skripov. *Metastable Liquids*. 1st ed. Translated from Russian. New York: Wiley, 1974.

- [239] S. Stephan, V. Bråten and H. Hasse. 'Mass transfer at vapor-liquid interfaces of $\text{H}_2\text{O} + \text{CO}_2$ mixtures studied by molecular dynamics simulation'. *Journal of Non-Equilibrium Thermodynamics* 49 (2024), pp. 441–461.
- [240] Ø. Wilhelmsen, T. T. Trinh, S. Kjelstrup, T. S. van Erp and D. Bedeaux. 'Heat and mass transfer across interfaces in complex nanogeometries'. *Physical Review Letters* 114 (2015), p. 065901.
- [241] Ø. Wilhelmsen and A. Aasen. 'Choked liquid flow in nozzles: Crossover from heterogeneous to homogeneous cavitation and insensitivity to depressurization rate'. *Chemical Engineering Science* 248 (2022), p. 117176.
- [242] Y. Liao and D. Lucas. 'Computational modelling of flash boiling flows: A literature survey'. *International Journal of Heat and Mass Transfer* 111 (2017), pp. 246–265.
- [243] M. Hammer, H. Deng, A. Austegard, A. M. Log and S. T. Munkejord. 'Experiments and modelling of choked flow of CO_2 in orifices and nozzles'. *International Journal of Multiphase Flow* 156 (2022), p. 104201.
- [244] P. Aursand, M. A. Gjennestad, E. Aursand, M. Hammer and Ø. Wilhelmsen. 'The spinodal of single- and multi-component fluids and its role in the development of modern equations of state'. *Fluid Phase Equilibria* 436 (2017), pp. 98–112.
- [245] P. Rehner, G. Bauer and J. Gross. 'FeO_s: An Open-Source Framework for Equations of State and Classical Density Functional Theory'. *Industrial and Engineering Chemistry Research* 62 (2023), pp. 5347–5357.
- [246] I. Cachadiña, A. Mulero and J. Tian. 'Surface Tension of Refrigerants—Selection of Data and Recommended Correlations'. *Journal of Physical and Chemical Reference Data* 44 (2015).
- [247] Á. Mulero, A. Hernández, V. Vadillo-Rodríguez and I. Cachadiña. 'Correlations of surface tension for mixtures of n-alkanes as a function of the composition: applicability and performance analysis of existing models'. *Physical Chemistry Chemical Physics* 27 (2025), pp. 12812–12836.
- [248] S. Wang, K. Hou and H. Heinz. 'Accurate and compatible force fields for molecular oxygen, nitrogen, and hydrogen to simulate gases, electrolytes, and heterogeneous interfaces'. *Journal of Chemical Theory and Computation* 17 (2021), pp. 5198–5213.
- [249] C. Vega and E. de Miguel. 'Surface tension of the most popular models of water by using the test-area simulation method'. *Journal of Chemical Physics* 126 (2007).

- [250] R. W. Hockney and J. W. Eastwood. *Computer simulation using particles*. 1st ed. Boca Raton, FL, USA: CRC Press, Taylor & Francis, 1988.
- [251] S. Nosé. 'A unified formulation of the constant temperature molecular dynamics methods'. *Journal of Chemical Physics* 81 (1984), pp. 511–519.
- [252] W. G. Hoover. 'Canonical dynamics: Equilibrium phase-space distributions'. *Physical Review A* 31 (1985), p. 1695.
- [253] W. Humphrey, A. Dalke and K. Schulten. 'VMD: Visual Molecular Dynamics'. *Journal of Molecular Graphics* 14 (1996), pp. 33–38.
- [254] J. Lekner and J. R. Henderson. 'Surface tension and energy of a classical liquid-vapour interface'. *Molecular Physics* 34 (1977), pp. 333–359.
- [255] T. Kawahara, T. Imai, T. Okumura and C. Kondou. 'Molecular simulation and experimental validation for surface tension of new Low-GWP refrigerant mixtures R32/CF3I and R1123/CF3I'. *Journal of Molecular Liquids* 386 (2023).
- [256] T. Esper, G. Bauer, P. Rehner and J. Gross. 'PCP-SAFT Parameters of Pure Substances Using Large Experimental Databases'. *Industrial & Engineering Chemistry Research* 62 (2023), pp. 15300–15310.
- [257] O. Fandiño, J. P. M. Trusler and D. Vega-Maza. 'Phase behavior of (CO₂ + H₂) and (CO₂ + N₂) at temperatures between (218.15 and 303.15) K at pressures up to 15 MPa'. *International Journal of Greenhouse Gas Control* 36 (2015), pp. 78–92.
- [258] T. A. Al-Sahhaf, A. J. Kidnay and E. D. Sloan. 'Liquid + vapor equilibriums in the N₂ + CO₂ + CH₄ system'. *Industrial & Engineering Chemistry Fundamentals* 22 (1983), pp. 372–380.
- [259] M. L. Huber. *Thermal conductivity, and surface tension of selected pure fluids as implemented in REFPROP v10. 0*. NIST Technical Note. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2017. Accessed: 3rd November 2025. URL: <https://doi.org/10.6028/NIST.IR.8209>.
- [260] M. L. Huber. *Preliminary Models for Viscosity, Thermal Conductivity, and Surface Tension of Pure Fluid Constituents of Selected Diesel Surrogate Fuels*. NIST Technical Note. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2017. Accessed: 3rd November 2025. URL: <https://doi.org/10.6028/NIST.TN.1949>.
- [261] M. E. Boudh-Hir and G. A. Mansoori. 'Statistical mechanics basis of Macleod's formula'. *Journal of Physical Chemistry* 94 (1990), pp. 8362–8364.

- [262] R. H. Fowler. 'A tentative statistical theory of Macleod's equation for surface tension, and the parachor'. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences* 159 (1937), pp. 229–246.
- [263] J. S. Rowlinson and F. Swinton. *Liquids and liquid mixtures: Butterworths monographs in chemistry*. 3rd ed. Oxford, UK: Butterworth-Heinemann, Elsevier, 2013.
- [264] A. Daryasafar, M. Masoudi, S. Kord and M. Madani. 'Evaluation of different thermodynamic models in predicting asphaltene precipitation: A comparative study'. *Fluid Phase Equilibria* 514 (2020), p. 112557.
- [265] A. Stukowski. 'Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool'. *Modelling and Simulation in Materials Science and Engineering* 18 (2010).
- [266] E. L. Quinn. 'The surface tension of liquid carbon dioxide'. *Journal of the American Chemical Society* 49 (1927), pp. 2704–2711.
- [267] J. J. Jasper. 'The surface tension of pure liquid compounds'. *Journal of Physical and Chemical Reference Data* 1 1 (1972), pp. 841–1010.
- [268] T. Merker, C. Engin, J. Vrabec and H. Hasse. 'Molecular model for carbon dioxide optimized to vapor-liquid equilibria'. *Journal of Chemical Physics* 132 (2010).
- [269] S. Stephan, F. Fleckenstein and H. Hasse. 'Vapor-Liquid Interfacial Properties of the Systems (Toluene + CO₂) and (Toluene + N₂): Experiments, Molecular Simulation, and Density Gradient Theory'. *Journal of Chemical and Engineering Data* 69 (2023), pp. 590–607.
- [270] M. C. Antoine. 'Nouvelle relation entre les tensions et les temperatures'. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences* 107 (1888), pp. 681–684.
- [271] F. Fleckenstein, S. Becker, H. Hasse and S. Stephan. 'Vapor-liquid interfacial properties of the system acetone + CO₂: Experiments, molecular simulation, density gradient theory, and density functional theory'. *Fluid Phase Equilibria* 596 (2025).
- [272] H. Vijayakumaran and T. A. Lemma. 'CFD modelling of non-equilibrium condensation of CO₂ within a supersonic nozzle using metastability approach'. *Journal of Natural Gas Science and Engineering* 85 (2021), p. 103715.
- [273] C. Tammone, A. Romei, G. Persico and F. Haglind. 'Extension of the delayed equilibrium model to flashing flows of organic fluids in converging-diverging nozzles'. *International Journal of Multiphase Flow* 171 (2024), p. 104661.

- [274] M. Hammer, A. Aasen and Ø. Wilhelmsen. 'Extrapolating into no man's land enables accurate estimation of surface properties with multiparameter equations of state'. *Fluid Phase Equilibria* 586 (2024), p. 114196.
- [275] A. Aasen, M. Hammer, D. Reguera and Ø. Wilhelmsen. 'Estimating metastable thermodynamic properties by isochoric extrapolation from stable states'. *Journal of Chemical Physics* 161 (2024).
- [276] J. C. Hesson and R. E. Peck. 'Flow of two-phase carbon dioxide through orifices'. *AIChE Journal* 4 (1958), pp. 207–210.
- [277] A. Keller, K. Langenbach and H. Hasse. 'Comparison of predictions of the PC-SAFT equation of state and molecular simulations for the metastable region of binary mixtures'. *Fluid Phase Equilibria* 444 (2017), pp. 31–36.
- [278] R. P. Coutinho and M. Hadjicharalambous. 'Demystifying Vapor Collapse in CO₂ Injection Wells'. *Proceedings of the 17th International Conference on Greenhouse Gas Control Technologies (GHGT-17)*. Calgary, Canada, 2024. Accessed: 27th July 2026. URL: <https://ssrn.com/abstract=5010518>.
- [279] D. Raju, R. Skartlien, M. Ramdin and T. J. H. Vlugt. 'Machine Learning-Based Prediction of Phase Equilibria and Interfacial Properties of Multicomponent CO₂ Mixtures with Impurities: Application to CO₂ Transportation'. *Industrial & Engineering Chemistry Research* 65 (2026), pp. 18863–18890.
- [280] P. Rehner and G. Bauer. 'Application of Generalized (Hyper-) Dual Numbers in Equation of State Modeling'. *Frontiers in Chemical Engineering* 3 (2021), p. 758090.
- [281] Z. Pan and J. P. M. Trusler. 'Measurement and modelling of the interfacial tensions of CO₂ + decane-iododecane mixtures at high pressures and temperatures'. *Fluid Phase Equilibria* 566 (2023), p. 113700.
- [282] J. Algaba, J. M. Miguez, B. Mendiboure and F. J. Blas. 'An accurate density functional theory for the vapor–liquid interface of chain molecules based on the statistical associating fluid theory for potentials of variable range for Mie chainlike fluids'. *Physical Chemistry Chemical Physics* 21 (2019), pp. 11937–11948.
- [283] O. Redlich and A. T. Kister. 'Algebraic Representation of Thermodynamic Properties and the Classification of Solutions'. *Industrial & Engineering Chemistry* 40 (1948), pp. 341–345.
- [284] S. Gupta, J. R. Elliott, A. Anderko, J. Crosthwaite, W. G. Chapman and C. T. Lira. 'Current practices and continuing needs in thermophysical properties for the chemical industry'. *Industrial & Engineering Chemistry Research* 62 (2023), pp. 3394–3427.

- [285] B. P. Agbodekhe, M. N. Carlozo, D. O. Abranches, K. D. Jones, A. W. Dowling and E. J. Maginn. 'Enhanced thermophysical property prediction with uncertainty quantification using group contribution-Gaussian process regression'. *Molecular Systems Design & Engineering* 11 (2026), pp. 85–106.
- [286] J. R. Rumble, ed. *CRC Handbook of Chemistry and Physics*. 105th ed. Boca Raton, FL, USA: CRC Press, Taylor & Francis, 2024.
- [287] J. Mouallem, A. Raza, G. Glatz, M. Mahmoud and M. Arif. 'Estimation of CO₂-brine interfacial tension using machine learning: implications for CO₂ geo-storage'. *Journal of Molecular Liquids* 393 (2024), p. 123672.
- [288] C. G. Albà, I. I. Alkhatib, L. F. Vega and F. Llovel. 'Machine learning integration in thermodynamics: Predicting CO₂ mixture saturation properties for sustainable refrigeration applications'. *Journal of CO₂ Utilization* 95 (2025), p. 103072.
- [289] S. Sharma, P. Yang, Y. Liu, K. R. Rossi, P. Bai, M. S. Rigutto, E. Zuidema, U. Agarwal, R. Baur, S. Calero, D. Dubbeldam and T. J. H. Vlucht. 'Machine Learning-Based Predictions of Henry Coefficients for Long-Chain Alkanes in One-Dimensional Zeolites: Application to Hydroisomerization'. *Journal of Physical Chemistry C* 129 (2025), pp. 18234–18249.
- [290] E. Johnsson, S. Sharma, A. G. Rao, D. Dubbeldam, S. Calero and T. J. H. Vlucht. 'Predicting the Maximum Loading in Zeolites for Hydroisomerization Applications: A Machine Learning Approach'. *Journal of Physical Chemistry C* 130 (2026), pp. 4299–4314.
- [291] J. Zhang, Q. Feng, S. Wang, X. Zhang and S. Wang. 'Estimation of CO₂-brine interfacial tension using an artificial neural network'. *Journal of Supercritical Fluids* 107 (2016), pp. 31–37.
- [292] A. Kamari, M. Pournik, A. Rostami, A. Amirlatifi and A. H. Mohammadi. 'Characterizing the CO₂-brine interfacial tension (IFT) using robust modeling approaches: A comparative study'. *Journal of Molecular Liquids* 246 (2017), pp. 32–38.
- [293] T. Song, W. Zhu, H. Emami-Meybodi, Y. Jiang, S. Chen, M. Yue, H. Mahani, Q. Liao, S. Iglaier and B. Pan. 'An explicit machine learning model for brine-gas interfacial tension prediction: Implications for H₂, CH₄, and CO₂ geo-storage'. *Fuel* 405 (2026), p. 136502.
- [294] M. H. Makke and K. Ghorayeb. 'Integrating physics and machine learning: Viscosity estimation in impure CO₂ transport systems'. *Fuel* 406 (2026), p. 137078.

- [295] M. H. Makke and K. Ghorayeb. 'Machine Learning-Based Assessment of Impurity Effects on Thermophysical Properties in CO₂-Rich Mixtures for CCUS Applications'. *IPTC Summit on AI for the Energy Industry*. International Petroleum Technology Conference, 2026, IPTC-25251-MS. Accessed: 27th July 2026. URL: <https://onepetro.org/IPTCONF/proceedings-abstract/26IPTC/26IPTC/D021S015R003/794891>.
- [296] B. Zhao, L. Ding and Y. Su. 'A Bayesian network-based robust framework for determining density, viscosity, and thermal conductivity of near-critical-state CO₂ applied in CCUS'. *Chemometrics and Intelligent Laboratory Systems* 269 (2026), p. 105636.
- [297] J. F. Lutsko. 'Recent Developments in Classical Density Functional Theory'. *Advances in Chemical Physics*. Ed. by S. A. Rice. Vol. 144. New York, USA: John Wiley & Sons, 2010. Chap. 1, pp. 1–92. Accessed: 27th July 2026. URL: <https://onlinelibrary.wiley.com/doi/10.1002/9780470564318.ch1>.
- [298] D. G. Anderson. 'Iterative procedures for nonlinear integral equations'. *Journal of the Association for Computing Machinery* 12 (1965), pp. 547–560.
- [299] G. Shen, Y. Sun, X. Zhang, X. Gao, Y. Qian, X. Lu and X. Ji. 'Partition and selectivity of electrolytes in cylindrical nanopores with heterogeneous surface charge'. *Journal of Molecular Liquids* 340 (2021), p. 116839.
- [300] P. Maidl, K. Langenbach and F. Frank. 'Achieving Convergence in Multiphase Multicomponent Density Gradient Theory Calculations through Regularization'. *Industrial & Engineering Chemistry Research* 63 (2024), pp. 14367–14380.
- [301] H. V. Thanh, M. Rahimi, S. Tangparitkul and N. Promsuk. 'Modeling the thermal transport properties of hydrogen and its mixtures with greenhouse gas impurities: A data-driven machine learning approach'. *International Journal of Hydrogen Energy* 83 (2024), pp. 1–12.
- [302] Y. Liu, W. Hong and B. Cao. 'Machine learning for predicting thermodynamic properties of pure fluids and their mixtures'. *Energy* 188 (2019), p. 116091.
- [303] F. Jirasek and H. Hasse. 'Combining machine learning with physical knowledge in thermodynamic modeling of fluid mixtures'. *Annual review of chemical and biomolecular engineering* 14 (2023), pp. 31–51.
- [304] F. Succetti, M. Panella, P. Giannitrapani, J.-C. Rigo and S. Colonnese. 'Machine Learning for Fluid Thermodynamics: State of the Art and Open Challenges'. *IEEE Access* 13 (2025), pp. 171073–171092.

- [305] N. Hollmann, S. Müller, L. Purucker, A. Krishnakumar, M. Körfer, S. B. Hoo, R. T. Schirrmeister and F. Hutter. ‘Accurate predictions on small data with a tabular foundation model’. *Nature* 637 (2025), pp. 319–326.
- [306] L. Grinsztajn, K. Flöge, O. Key, F. Birkel, P. Jund, B. Roof, B. Jäger, D. Safaric, S. Alessi, A. Hayler, M. Manium, R. Yu, F. Jablonski, S. B. Hoo, A. Garg, J. Robertson, M. Bühler, V. Moroshan, L. Purucker, C. Cornu, L. C. Wehrhahn, A. Bonetto, B. Schölkopf, S. Gambhir, N. Hollmann and F. Hutter. *TabPFN-2.5: Advancing the State of the Art in Tabular Foundation Models*. 2026. arXiv: [2511.08667 \[cs.LG\]](https://arxiv.org/abs/2511.08667). Accessed: 5th March 2026. URL: <https://arxiv.org/abs/2511.08667>.
- [307] J. Hensman, A. G. d. G. Matthews and Z. Ghahramani. ‘Scalable Variational Gaussian Process Classification’. *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2015, pp. 351–360.
- [308] C. Cortes and V. Vapnik. ‘Support-Vector Networks’. *Machine Learning* 20 (1995), pp. 273–297.
- [309] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola and V. Vapnik. ‘Support Vector Regression Machines’. *Advances in Neural Information Processing Systems* 9 (1997), pp. 155–161.
- [310] L. Breiman. ‘Random Forests’. *Machine Learning* 45 (2001), pp. 5–32.
- [311] T. Chen and C. Guestrin. ‘XGBoost: A Scalable Tree Boosting System’. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 785–794. Accessed: 5th March 2026. URL: <https://doi.org/10.1145/2939672.2939785>.
- [312] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, M. Li, J. Xie, M. Lin, Y. Geng, Y. Li, J. Yuan and D. Cortes. *xgboost: Extreme Gradient Boosting*. Python package version 3.2.1. 2026. Accessed: 27th July 2026. URL: <https://github.com/dmlc/xgboost>.
- [313] Prior Labs. *TabPFN Extensions: Community extensions for TabPFN*. 2025. Accessed: 16th May 2026. URL: <https://github.com/PriorLabs/tabpfm-extensions>.
- [314] Y. Freund and R. E. Schapire. ‘A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting’. *Journal of Computer and System Sciences* 55 (1997), pp. 119–139.
- [315] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush and A. Gulin. ‘CatBoost: Unbiased Boosting with Categorical Features’. *Advances in Neural Information Processing Systems* 31 (2018), pp. 6639–6649.

- [316] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay. 'Scikit-learn: Machine Learning in Python'. *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [317] M. Cranmer. *Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl*. 2023. DOI: [10.48550/arXiv.2305.01582](https://doi.org/10.48550/arXiv.2305.01582). arXiv: [2305.01582](https://arxiv.org/abs/2305.01582) [cs.LG]. Accessed: 9th May 2026.
- [318] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. 1st ed. Cambridge, MA, USA: MIT Press, 2006.
- [319] M. K. Titsias. 'Variational Learning of Inducing Variables in Sparse Gaussian Processes'. *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2009, pp. 567–574. Accessed: 12th March 2026. URL: <http://proceedings.mlr.press/v5/titsias09a.html>.
- [320] J. Hensman, N. Fusi and N. D. Lawrence. 'Gaussian Processes for Big Data'. *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*. 2013, pp. 282–290.
- [321] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola and L. K. Saul. 'An introduction to variational methods for graphical models'. *Machine Learning* 37 (1999), pp. 183–233.
- [322] D. M. Blei, A. Kucukelbir and J. D. McAuliffe. 'Variational Inference: A Review for Statisticians'. *Journal of the American Statistical Association* 112 (2017), pp. 859–877.
- [323] S. Kullback and R. A. Leibler. 'On information and sufficiency'. *Annals of Mathematical Statistics* 22 (1951), pp. 79–86.
- [324] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger and A. G. Wilson. 'GPYtorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration'. *Advances in Neural Information Processing Systems*. Vol. 31. 2018, pp. 7576–7586.
- [325] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin. *Bayesian Data Analysis*. 3rd ed. Boca Raton, FL, USA: Chapman & Hall/CRC, 2013.
- [326] T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York, NY, USA: Springer, 2009.
- [327] C. M. Bishop. *Pattern Recognition and Machine Learning*. 1st ed. New York, USA: Springer, 2006.
- [328] P. J. Rousseeuw. 'Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis'. *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65.

- [329] Y. Xiang, Y.-H. Tang, Z. Gong, H. Liu, L. Wu, G. Lin and H. Sun. 'Efficient exploration of chemical compound space using active learning for prediction of thermodynamic properties of alkane molecules'. *Journal of Chemical Information and Modeling* 63 (2023), pp. 6515–6524.
- [330] Y. Xiang, Y.-H. Tang, H. Liu, G. Lin and H. Sun. 'Predicting single-substance phase diagrams: A kernel approach on graph representations of molecules'. *Journal of Physical Chemistry A* 125 (2021), pp. 4488–4497.
- [331] V. Diky, C. D. Muzny, A. Y. Smolyanitsky, A. Bazyleva, R. D. Chirico, J. W. Magee, Y. Paulechka, A. F. Kazakov, S. A. Townsend and E. W. Lemmon. *ThermoData Engine (TDE) Version 10: Pure Compounds, Binary Mixtures, Ternary Mixtures, and Chemical Reactions*. NIST Standard Reference Database 103b. Gaithersburg, MD, USA, 2015.
- [332] C. Spearman. 'The proof and measurement of association between two things'. *American Journal of Psychology* 100 (1987), pp. 441–471.
- [333] F. J. Massey Jr. 'The Kolmogorov-Smirnov test for goodness of fit'. *Journal of the American statistical Association* 46 (1951), pp. 68–78.
- [334] M. Rosenblatt. 'Remarks on Some Nonparametric Estimates of a Density Function'. *Annals of Mathematical Statistics* 27 (1956), pp. 832–837.
- [335] E. Parzen. 'On Estimation of a Probability Density Function and Mode'. *Annals of Mathematical Statistics* 33 (1962), pp. 1065–1076.
- [336] F. Okoro, F. J. Owuna, A. Chapoy, P. Ahmadi and R. Burgass. 'Bubble points and densities of H₂ (up to 5%) in CO₂-rich binary systems'. *Fluid Phase Equilibria* 592 (2025), p. 114321.
- [337] S. M. Lundberg and S.-I. Lee. 'A unified approach to interpreting model predictions'. *Advances in Neural Information Processing Systems* 30 (2017), pp. 4765–4774.
- [338] SLB. *OLGA Dynamic Multiphase Flow Simulator*. SLB. 2026. Accessed: 25th June 2026. URL: <https://www.slb.com/products-and-services/delivering-digital-at-scale/software/olga>.
- [339] Kongsberg Digital. *LedaFlow: Transient Multiphase Flow Simulation Software*. Kongsberg Digital. 2026. Accessed: 25th June 2026. URL: <https://kongsbergdigital.com/products/ledaflow>.
- [340] SINTEF. *LedaFlow*. SINTEF. 2026. Accessed: 25th June 2026. URL: <https://www.sintef.no/en/software/ledaflow/>.
- [341] D. Raju, M. Ramdin, J.-M. Simon, P. Krüger and T. J. H. Vlugt. 'Finite-size effects of the excess entropy computed from integrating the radial distribution function'. *Molecular Physics* 123 (2025), e2456115.

- [342] B. B. Laird and A. Haymet. 'Calculation of the entropy from multiparticle correlation functions'. *Physical Review A* 45 (1992), pp. 5680–5689.
- [343] S. A. Ghaffarizadeh and G. J. Wang. 'Excess entropy scaling in active-matter systems'. *Journal of Physical Chemistry Letters* 13 (2022), pp. 4949–4954.
- [344] A. Saliou, P. Jarry and N. Jakse. 'Excess entropy scaling law: A potential energy landscape view'. *Physical Review E* 104 (2021), p. 044128.
- [345] H. J. Raveché. 'Entropy and molecular correlation functions in open systems. I. Derivation'. *Journal of Chemical Physics* 55 (1971), pp. 2242–2250.
- [346] A. Baranyai and D. J. Evans. 'Direct entropy calculation from computer simulation of liquids'. *Physical Review A* 40 (1989), p. 3817.
- [347] A. Samanta, S. M. Ali and S. K. Ghosh. 'Universal scaling laws of diffusion in a binary fluid mixture'. *Physical Review Letters* 87 (2001), p. 245901.
- [348] J. Hoyt, M. Asta and B. Sadigh. 'Test of the universal scaling law for the diffusion coefficient in liquid metals'. *Physical Review Letters* 85 (2000), p. 594.
- [349] Y. Zhang, L. Dong, L.-M. Wang, R.-P. Liu and S. Sanvito. 'Towards quantifying (meta-) stability of multi-principal element alloys: from configurational entropy to characteristic temperatures'. *Acta Materialia* 281 (2024), p. 120415.
- [350] P. M. Piaggi and M. Parrinello. 'Entropy based fingerprint for local crystalline order'. *Journal of Chemical Physics* 147 (2017), p. 114112.
- [351] P. M. Piaggi, O. Valsson and M. Parrinello. 'Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations'. *Physical Review Letters* 119 (2017), p. 015701.
- [352] D. C. Wallace. 'Statistical mechanical theory of liquid entropy'. *International Journal of Quantum Chemistry* 52 (1994), pp. 425–435.
- [353] D. C. Wallace. 'On the role of density fluctuations in the entropy of a fluid'. *Journal of Chemical Physics* 87 (1987), pp. 2282–2284.
- [354] Y. Huang and M. Widom. 'Entropy approximations for simple fluids'. *Physical Review E* 109 (2024), p. 034130.
- [355] J. G. Kirkwood and E. M. Boggs. 'The radial distribution function in liquids'. *Journal of Chemical Physics* 10 (1942), pp. 394–402.
- [356] P. Krüger and T. J. H. Vlugt. 'Size and shape dependence of finite-volume Kirkwood–Buff integrals'. *Physical Review E* 97 (2018), p. 051301.

- [357] P. Ganguly and N. F. van der Vegt. 'Convergence of sampling Kirkwood–Buff integrals of aqueous solutions with molecular dynamics simulations'. *Journal of Chemical Theory and Computation* 9 (2013), pp. 1347–1355.
- [358] J. Salacuse, A. Denton and P. Egelstaff. 'Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. I. Theoretical method'. *Physical Review E* 53 (1996), p. 2382.
- [359] P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, T. J. H. Vlugt and J.-M. Simon. 'Kirkwood–Buff integrals for finite volumes'. *Journal of Physical Chemistry Letters* 4 (2013), pp. 235–238.
- [360] J. C. Amazigo and L. A. Rubinfeld. *Advanced calculus and its applications to the engineering and physical sciences*. 1st ed. New York, USA: John Wiley & Sons, 1980.
- [361] A. Ben-Naim. *Molecular theory of solutions*. 1st ed. Oxford, UK: Oxford University Press, 2006.
- [362] L. Verlet. 'Computer experiments on classical fluids. II. Equilibrium correlation functions'. *Physical Review* 165 (1968), p. 201.
- [363] N. Dawass, P. Krüger, J.-M. Simon and T. J. H. Vlugt. 'Kirkwood–Buff integrals of finite systems: shape effects'. *Molecular Physics* 116 (2018), pp. 1573–1580.
- [364] A. Santos. 'Finite-size estimates of Kirkwood–Buff and similar integrals'. *Physical Review E* 98 (2018), p. 063302.
- [365] N. Dawass, P. Krüger, S. K. Schnell, O. A. Moulton, I. G. Economou, T. J. H. Vlugt and J.-M. Simon. 'Kirkwood–Buff integrals using molecular simulation: estimation of surface effects'. *Nanomaterials* 10 (2020), p. 771.
- [366] S. K. Schnell, X. Liu, J.-M. Simon, A. Bardow, D. Bedeaux, T. J. H. Vlugt and S. Kjelstrup. 'Calculating thermodynamic properties from fluctuations at small scales'. *Journal of Physical Chemistry B* 115 (2011), pp. 10911–10918.
- [367] X. Wang, S. Ramírez-Hinestrosa, J. Dobnikar and D. Frenkel. 'The Lennard–Jones potential: when (not) to use it'. *Physical Chemistry Chemical Physics* 22 (2020), pp. 10624–10633.
- [368] R. C. Tolman. 'The Effect of Droplet Size on Surface Tension'. *Journal of Chemical Physics* 17 (1949), pp. 333–337.
- [369] A. Aasen, D. Reguera and Ø. Wilhelmsen. 'Curvature Corrections Remove the Inconsistencies of Binary Classical Nucleation Theory'. *Physical Review Letters* 124 (2020), p. 045701.
- [370] A. Aasen, Ø. Wilhelmsen, M. Hammer and D. Reguera. 'Free energy of critical droplets—from the binodal to the spinodal'. *Journal of Chemical Physics* 158 (2023), p. 114108.

- [371] N. F. de Souza, L. F. M. Franco and B. Coasne. 'Consistency of Equilibrium and Nonequilibrium Molecular Dynamics to Assess Thermal Conductivity'. *Journal of Chemical & Engineering Data* 71 (2026), pp. 1022–1032.
- [372] G. Guevara-Carrion, D. Saric and J. Vrabec. 'Thermodiffusion of CO₂ Mixtures in the Extended Critical Region'. *Fluid Phase Equilibria* 596 (2025), p. 114438.
- [373] S. Kjelstrup and D. Bedeaux. *Non-Equilibrium Thermodynamics of Heterogeneous Systems*. Vol. 16. Series on Advances in Statistical Mechanics. Singapore: World Scientific, 2008.
- [374] M. Thol, G. Rutkai, A. Köster, R. Lustig, R. Span and J. Vrabec. 'Equation of State for the Lennard-Jones Fluid'. *Journal of Physical and Chemical Reference Data* 45 (2016), p. 023101.
- [375] W. Zhang, D. Shao, Y. Yan, S. Liu and T. Wang. 'Experimental Investigations into the Transient Behaviours of CO₂ in a Horizontal Pipeline during Flexible CCS Operations'. *International Journal of Greenhouse Gas Control* 79 (2018), pp. 193–199.

Summary

Carbon Capture, Utilization, and Storage (CCUS) is a key technology for reducing anthropogenic CO₂ emissions. Transporting CO₂ safely and efficiently from capture facilities to storage reservoirs is central to this chain. Captured CO₂ is rarely pure. The stream carries impurities such as N₂, Ar, H₂, CH₄, O₂, CO, and H₂S. Even at low concentrations, these impurities shift the phase behavior, the bulk thermophysical properties, and the interfacial properties of CO₂-rich streams. These impurities also make it harder to keep CO₂ in a dense, single phase that transport requires. Measuring these properties experimentally for every relevant composition, temperature, and pressure is impractical. In this thesis, these properties are therefore predicted using molecular simulations, classical thermodynamic theories, and machine learning. These techniques are fast enough to be used in large-scale process and flow simulations. Chapter 2 examines how impurities affect bulk thermodynamic and transport properties. Monte Carlo (MC) and Molecular Dynamics (MD) simulations were used to compute densities, thermal expansion coefficients, isothermal compressibilities, heat capacities, Joule–Thomson coefficients, speeds of sound, and viscosities. The systems covered pure CO₂ and its binary, ternary, and quaternary mixtures with N₂, Ar, H₂, and CH₄ at (1–10) mol%, for (253–313) K and (20–200) bar. Vapor–liquid equilibria computed with the Continuous Fractional Component Monte Carlo (CFCMC) method validate the force fields with the GERG-2008 and Span–Wagner Equations of State (EoS). Impurities lower the density in all cases. In the gas phase, impurities also lower the thermal expansion coefficient, isothermal compressibility, heat capacity, and Joule–Thomson coefficient, whereas these properties increase in the liquid and supercritical phases. The speed of sound follows the opposite trend. The result of this chapter is an extensive dataset of these properties for the safe design of CO₂ transport systems. Chapter 3 turns to interfacial properties. The phase equilibria and vapor–liquid interfacial properties of pure CO₂ and of CO₂–CH₄, CO₂–Ar, CO₂–N₂, and CO₂–H₂ mixtures are computed at (220–273) K. MD simulations are compared with the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) EoS coupled with classical Density Functional Theory (cDFT). Both approaches reproduce the phase equilibria and interfacial tensions (IFTs) of pure CO₂. For the mixtures, the PC-SAFT results depend strongly on a nonzero binary interaction parameter (k_{ij}), whereas MD systematically overpredicts the IFTs. The IFTs decrease with pressure, least strongly for H₂-containing mixtures, and the lighter component accumulates at the interface. These

findings refine predictions of the metastability limits used in multiphase CO₂ flow modeling. These limits mark how far a stream can be expanded beyond its equilibrium phase boundary before vaporization or condensation begins. Direct PC-SAFT + cDFT evaluation is accurate but computationally expensive. Chapter 4 therefore combines these techniques with machine learning (ML) surrogate models. A dataset of ca. 1.9×10^4 points was generated for 100 multicomponent CO₂-rich systems, ranging from 95% to 99.99% CO₂ and from binary to eight-component, using the polar variant PCP-SAFT coupled with cDFT and optimized k_{ij} correlations. Fully data-driven models (Support Vector Machines, Random Forests, XGBoost, and TabPFN) were trained on temperature, pressure, and composition. TabPFN provides the best accuracy and robustness, followed by XGBoost. Active learning based on Gaussian Process Regression (GPR) proved the most data-efficient strategy for sampling the training data. A physics-informed hybrid model, which learns the residual of a Winterfeld–Scriven–Davis (WSD) baseline, outperforms the best data-driven IFT model. The trained surrogates predict the properties at a given temperature, pressure, and composition within milliseconds, which is orders of magnitude faster than direct cDFT. This speed enables rapid property estimation for Computational Fluid Dynamics (CFD) and transient multiphase flow simulators. Chapter 5 steps back to a more fundamental question. Entropy-scaling laws are widely used to compute transport properties such as diffusion coefficients, viscosities, and thermal conductivities. These laws require the excess entropy S^{ex} , which can be estimated by integrating the Radial Distribution Function (RDF) over space. Using an analytic RDF model together with MC simulations of Lennard-Jones-like and colloid-like particles, this chapter shows that truncating this integral at a finite box length is justified when the range of the RDF is not too long and an appropriate finite-size expression is used. Integrating the thermodynamic limit-corrected RDF $g^\infty(r)$ yields excess entropies that agree with thermodynamic integration, even for systems of only 100 particles. At high densities, the second-order expansion of the excess entropy becomes the dominant error source, with deviations of up to 20%. Taken together, the chapters of this thesis provide a consistent, multiscale toolbox spanning molecular simulations, classical EoS and density functional theories, and machine learning. This toolbox quantifies how impurities alter the bulk, interfacial, and phase-equilibrium properties of CO₂-rich streams and makes accurate predictions cheap enough for process- and pipeline-scale simulations. The resulting models and data show how much CO₂ a pipeline or ship can carry, how the stream heats or cools during compression and expansion, and when bubbles or droplets start to form. This knowledge supports the safe, efficient, and reliable design and operation of CO₂ transport and injection systems within the CCUS chain. In a broader perspective, this work contributes to safer and more affordable transport of captured CO₂, and thereby supports CCUS as one of the tools available to society for reducing CO₂ emissions.

Samenvatting

Carbon Capture, Utilization, and Storage (CCUS, het afvangen, gebruiken en opslaan van CO₂) is een sleuteltechnologie voor het terugdringen van antropogene CO₂-emissies. Het veilige, efficiënte transport van CO₂ van afvanginstallaties naar opslaglocaties vormt een centraal onderdeel van deze keten. Afgelaten CO₂ is zelden zuiver. De stroom bevat onzuiverheden zoals N₂, Ar, H₂, CH₄, O₂, CO en H₂S. Zelfs bij lage concentraties beïnvloeden deze onzuiverheden het fasegedrag, de thermofysische bulkeigenschappen en de grensvlakeigenschappen van CO₂-rijke stromen. Deze onzuiverheden bemoeilijken ook het handhaven van de dichte één-fase toestand die het transport vereist. Het experimenteel bepalen van deze eigenschappen voor alle relevante samenstellingen, temperaturen en drukken is uiterst uitdagend. In dit proefschrift worden deze eigenschappen daarom voorspeld met moleculaire simulaties, klassieke thermodynamische theorieën en machine learning. Deze methodes zijn snel genoeg voor proces- en stromingssimulaties op grote schaal. Hoofdstuk 2 onderzoekt het effect van onzuiverheden op de thermofysische bulk- en transporteigenschappen. Met Monte Carlo (MC)- en Molecular Dynamics (MD)-simulaties zijn dichtheden, thermische uitzettingscoëfficiënten, isotherme compressibiliteiten, warmtecapaciteiten, Joule-Thomson-coëfficiënten, geluidssnelheden en viscositeiten berekend. De onderzochte systemen omvatten zuivere CO₂ en binaire, ternaire en quaternaire mengsels met N₂, Ar, H₂ en CH₄ bij (1–10) mol%, voor (253–313) K en (20–200) bar. Vloeistof-dampevenwichten, berekend met de Continuous Fractional Component Monte Carlo (CFCMC)-methode, valideren de force fields met de GERG-2008- en Span-Wagner-toestandsvergelijkingen (Equations of State, EoS). Onzuiverheden verlagen in het algemeen de dichtheid. In de gasfase verlagen onzuiverheden ook de thermische uitzettingscoëfficiënt, de isotherme compressibiliteit, de warmtecapaciteit en de Joule-Thomson-coëfficiënt, terwijl deze in de vloeibare en superkritische fase juist toenemen. De geluidssnelheid vertoont de tegenovergestelde trend. Het resultaat van dit hoofdstuk is een uitgebreide dataset van de genoemde eigenschappen voor het veilige ontwerp van CO₂-transportsystemen. Hoofdstuk 3 richt zich op grensvlakeigenschappen. De fase-evenwichten en vloeistof-damp-grensvlakeigenschappen van zuivere CO₂ en van CO₂-CH₄-, CO₂-Ar-, CO₂-N₂- en CO₂-H₂-mengsels zijn berekend bij (220–273) K. Daarbij worden MD-simulaties vergeleken met de Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) EoS gekoppeld aan klassieke dichtheidsfunctionaaltheorie (cDFT). Beide methodes reproduceren de fase-evenwichten en grensvlakspanningen (IFTs) van zuivere CO₂. Voor de

mengsels hangen de PC-SAFT-resultaten sterk af van de waarde van de binaire interactieparameter (k_{ij}), terwijl MD de IFTs systematisch overschat. De IFTs nemen af met de druk. Dit is het minst sterk voor H_2 -houdende mengsels. De lichtere component verrijkt zich aan het grensvlak. Deze bevindingen verfijnen de voorspellingen van de metastabiliteitsgrenzen die worden gebruikt bij het modelleren van meerfase CO_2 -stromingen. Deze grenzen geven aan hoe ver een stroom voorbij de evenwichtsfasegrens kan worden gebracht voordat verdamping of condensatie begint. Directe PC-SAFT + cDFT-evaluatie is nauwkeurig maar rekenkundig kostbaar. Hoofdstuk 4 combineert deze technieken daarom met machine learning (ML)-surrogaatmodellen. Er is een dataset van circa 1.9×10^4 datapunten gegenereerd voor 100 meercomponenten CO_2 -rijke systemen, van 95% tot 99,99% CO_2 en van binair tot acht componenten. De polaire variant PCP-SAFT, gekoppeld aan cDFT en geoptimaliseerde k_{ij} -correlaties, is hierbij gebruikt. Volledig datagedreven modellen (Support Vector Machines, Random Forests, XGBoost en TabPFN) zijn getraind op temperatuur, druk en samenstelling. TabPFN biedt de beste nauwkeurigheid en robuustheid, gevolgd door XGBoost. Actief leren op basis van Gaussian Process Regression (GPR) bleek de meest data-efficiënte strategie om de trainingsdata te samplen. Een fysisch geïnformeerd hybride model dat het residu van een Winterfeld-Scriven-Davis (WSD)-basiswaarde leert, presteert beter dan het beste datagedreven IFT-model. De getrainde surrogaatmodellen voorspellen de eigenschappen bij een gegeven temperatuur, druk en samenstelling binnen milliseconden, hetgeen ordes van grootte sneller is dan directe cDFT. Dit maakt snelle berekeningen van eigenschappen mogelijk voor Computational Fluid Dynamics (CFD) en transiënte meerfase stromingssimulatoren. Hoofdstuk 5 keert terug naar een fundamentele vraag. Entropie-schalingswetten worden veelvuldig gebruikt om transporteigenschappen zoals diffusiecoëfficiënten, viscositeiten en thermische geleidbaarheden te berekenen. Deze wetten vereisen de excesentropie S^{ex} , die kan worden geschat door de radiale distributiefunctie (RDF) over de ruimte te integreren. Met behulp van een analytisch RDF-model en MC-simulaties van Lennard-Jones-achtige en colloïdale deeltjes wordt in dit hoofdstuk aangetoond dat het afkappen van deze integraal bij een eindige boxlengte gerechtvaardigd is wanneer de dracht van de RDF niet te lang is en een geschikte finite-size correctie wordt toegepast. Integratie van de naar de thermodynamische limiet gecorrigeerde RDF $g^\infty(r)$ levert excesentropieën op die overeenkomen met thermodynamische integratie, zelfs voor systemen met slechts 100 deeltjes. Bij hoge dichtheden wordt de tweede-orde-expansie van de excesentropie de dominante foutbron, met afwijkingen tot 20%. Samenvattend vormen de hoofdstukken van dit proefschrift een consistente, multiscale gereedschapskist bestaande uit moleculaire simulaties, klassieke toestandsvergelijkingen en dichtheidsfunctionaaltheorieën, en machine learning. Deze gereedschapskist kwantificeert hoe onzuiverheden de bulk-, grensvlak- en fase-evenwichtseigenschappen van CO_2 -rijke stromen veranderen en maakt

nauwkeurige voorspellingen goedkoop genoeg voor simulaties op proces- en pijpleidingschaal. De resulterende modellen en data laten zien hoeveel CO₂ een pijpleiding of schip kan vervoeren, hoe de stroom opwarmt of afkoelt tijdens compressie en expansie, en wanneer zich bellen of druppels beginnen te vormen. Deze kennis ondersteunt het veilige, efficiënte en betrouwbare ontwerp en gebruik van CO₂-transport- en injectiesystemen binnen de CCUS-keten. In een breder perspectief draagt dit werk bij aan een veiliger en betaalbaarder transport van afgevangen CO₂, en ondersteunt het daarmee CCUS als een van de middelen die de samenleving heeft om CO₂-emissies terug te dringen.

Curriculum Vitæ

Darshan Raju was born on 25 December 1997 in Coimbatore, India. In 2015, he started his B.Tech. in Mechanical Engineering at Amrita University, India, and graduated with distinction in 2019. He then pursued an M.Sc. in Mechanical Engineering at Delft University of Technology, the Netherlands. For his M.Sc. thesis, supervised by Prof. dr. ir. Sikke Klein, he investigated flashback mechanisms in low-swirl premixed burners using computational fluid dynamics, focusing on the numerical analysis of turbulent premixed combustion with high H_2 content. In 2023, he continued his academic career as a Ph.D. candidate in the Process and Energy Department at Delft University of Technology, under the supervision of Prof. dr. ir. Thijs J.H. Vlugt and Dr. ir. Mahinder Ramdin. His Ph.D. research focuses on understanding the effects of impurities in CO_2 streams for CO_2 transportation. To this end, he uses molecular simulation methods, including Monte Carlo and molecular dynamics, together with thermodynamic modeling based on equations of state and machine-learning approaches. His work focuses on computing the thermodynamic, transport, phase-equilibrium, and interfacial properties of CO_2 mixtures with impurities. These properties include density, heat capacity, speed of sound, viscosity, interfacial tension, and excess entropy. He also developed efficient surrogate models based on advanced machine learning methods to predict phase-equilibrium and interfacial properties of multicomponent CO_2 -rich mixtures. During his Ph.D., he collaborated with Prof. dr. Peter Krüger (Chiba University, Japan), Prof. dr. Jean-Marc Simon (ICB, CNRS/University of Burgundy, France), and Dr. Roar Skartlien (Institute for Energy Technology, Norway) on topics related to molecular modeling, thermophysical-property prediction, and CO_2 transport. This dissertation presents the findings of his research.

List of Publications

1. D. Raju, M. Ramdin and T. J. H. Vlugt. 'Thermophysical Properties and Phase Behavior of CO₂ with Impurities: Insight from Molecular Simulations'. *Journal of Chemical & Engineering Data* 69 (2024), pp. 2735–2755. (Chapter 2)
2. D. Raju, M. Ramdin, J.-M. Simon, P. Krüger and T. J. H. Vlugt. 'Finite-size effects of the excess entropy computed from integrating the radial distribution function'. *Molecular Physics* 123 (2025), e2456115. (Chapter 5)
3. D. Raju, R. Skartlien, M. Ramdin and T. J. H. Vlugt. 'Vapor–Liquid Interfacial Properties of CO₂ Mixtures for Sequestration Applications: Molecular Simulations, Classical Density Functional Theory, and Equations of State'. *Industrial & Engineering Chemistry Research* 65 (2026), pp. 5213–5238. (Chapter 3)
4. D. Raju, R. Skartlien, M. Ramdin and T. J. H. Vlugt. 'Machine Learning-Based Prediction of Phase Equilibria and Interfacial Properties of Multicomponent CO₂ Mixtures with Impurities: Application to CO₂ Transportation'. *Industrial & Engineering Chemistry Research* 65 (2026), pp. 18863–18890. (Chapter 4)

Acknowledgements

My doctoral journey would not have been possible without many inspiring and supportive people, all of whom I would like to acknowledge in this thesis. First and foremost, I am sincerely grateful to Prof. dr. ir. Thijs J.H. Vlugt and Dr. ir. Mahinder Ramdin for believing in me and for providing invaluable guidance and support throughout these years. Thijs, thank you for your continuous encouragement and for being a constant source of inspiration. Your remarkable dedication has set an example that I will continue to admire and strive to follow. Mahinder, thank you for the many warm discussions, both scientific and personal, that made life in the lab meaningful and engaging. I greatly admire your experimental expertise and your ability to articulate ideas clearly, whether the discussion concerns thermodynamics or any other matter. I am deeply grateful for the trust you both placed in me, and I will always cherish the moments we shared while traveling to conferences and consortium meetings. Your mentorship has truly shaped me, both academically and personally.

I have also been fortunate to collaborate with industry partners from TNO and Shell, including Dr. ir. Dries van Nimwegen, Ir. Stefan Belfroid, Ir. Harmen Slot, and Prof. dr. ir. Ruud Henkes. Our discussions provided valuable perspectives on the practical relevance and industrial application of this research. I further wish to thank Prof. dr. Jean-Marc Simon, Prof. dr. Peter Krüger, and Dr. Roar Skartlien for a fruitful collaboration. Although we have not had the opportunity to meet in person, our scientific exchanges have contributed considerably to this work.

This research was made possible through the support of the ENCASE consortium and NWO. The computational work relied on the generous resources provided by the Snellius and DelftBlue supercomputers, and I am grateful to the Delft High Performance Computing Center for facilitating access to these facilities.

I would especially like to thank Shrinjay and Jelle for always being there to talk to and for joining me as fellow ice-cream makers. Shrinjay, thank you for encouraging me to travel to New York, an experience I would have deeply regretted missing, and for trusting me as one of your paranympths. Jelle, thank you for the countless discussions and for helping me whenever I got stuck. I have learned a lot, and these years would not have been as smooth without you both. I am equally thankful to my other colleagues in the Engineering Thermodynamics group, namely Panji, Parsa, Mert, Thejas, Georgia, Bowen, Irem, and Sebastian. Each of you made the working days

more enjoyable and left me with memories that I will cherish for a long time. Special thanks are due to Dr. ir. Tim Nijssen and Dr. Othonas Moulτος for their engaging discussions and valuable feedback during group meetings. My thanks also go to Prof. dr. Monique van der Veen and Chunyu Huang for the stimulating scientific conversations and our productive collaboration. I am also grateful to my office mates, Suriya, Marko, Heng, Mengmeng, Nagaraj, Mahmod, Zhaopeng, Karlijn, Ali, and Qi, for creating a pleasant work environment. It has been a pleasure to share an office with all of you. These acknowledgements would be incomplete without Abhirath, Asif, Sowmya, Nikhilesh, Faruk, Pepijn, Ravi, Arvind, Arjun, Fathaah, and Shahid, whose good company I valued throughout my time in the department.

Life in the Netherlands can be difficult and lonely when you are far from your family, especially during the Ph.D. journey. Fortunately, I was spared much of that, since the void was filled by many friends, well-wishers, and guardians who deserve a special mention here. My warm thanks go to the van Hulst family, in particular Jack, Henk, and Harold, for being more than just my landlords and for treating me as part of the family. I am forever indebted to you for that. Thanks also to my roommates, Sebastian and Roshan, for tolerating me. Thank you, Sebastian, for the much-needed breaks and for discussing different viewpoints on all aspects of life. Toon, thank you for the lunch breaks and for patiently listening to me talk about Tamil Nadu politics and Vijay. I owe a great deal to Teja, Sanjay, and Hem for being a rock-solid support system and for always being there in the toughest moments. Thank you, Teja, for being a constructive bully. Thank you, Sanjay, for the distractions from research, which were always welcome. Thank you, Vignesh, Atharv, Nishanth, Ranges, Srikar, and Sridhar, for the weekend cooking, movie nights, and borrels. Vignesh, I hope the good influences outweighed the bad ones. Vinay, Nikhil, Navin, Lokesh, Hari, Chakri, Kavin, Prasad, Prashanth, and Rajneesh, thank you for being good friends and for always wishing me well. I am also grateful to my friends back home, Maharajan and Anbu, who have brought me much joy and who are always there when I call to rant. I also thank my brother, Mithun, who has always wished me the best. To anyone I may have unintentionally left out, I apologize. Finally, I would like to express my heartfelt gratitude to my parents, Raju and Kavitha, for their unwavering support and patience. Your belief in me has been a lasting source of motivation, making this journey all the more memorable.