Theoretical Modeling of Water and Aqueous Systems

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Michael von Domaros
aus
Leipzig

Bonn, 2018

Angefertigt mit Genehmigung
der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Erstgutachterin:	Prof. Dr. Barbara Kirchner
Zweitgutachter:	Prof. Dr. Thomas Bredow
Tag der Promotion:	04. Juli 2018
Erscheinungsjahr:	2018

Abstract

Thermodynamic and kinetic properties of liquid water and a variety of aqueous systems were investigated by means of theoretical methods. Focus was set on the prediction of mixing-induced changes in thermodynamic potentials, as well as on interfacial water dynamics.

One of the employed methods is Quantum Cluster Equilibrium (QCE) theory, which introduces quantum chemistry to the well-established class of mixture theories. In this thesis, the origin of the two empirical QCE parameters was analyzed in great detail and their tight connection to the van der Waals equation of state was established. This connection was exploited to refine the binary mixture version of QCE theory. The proposed measures can eliminate the need for binary reference data and were proven to be viable approximations in a study on three different amide/water mixtures. Predicting miscibilities without need for binary reference data or resorting to empiricism is an important and open task in modern theoretical research on binary mixtures that is now within the reach of QCE theory.

Like in any mixture theory, the success of QCE calculations depends sensitively on the choice of representative clusters, thus a systematic and empiricism-free scheme to generate cluster sets was proposed. Therein, only global minimum structures, obtained from a genetic-algorithm-based global geometry optimization, are employed. Such cluster sets reach almost chemical accuracy in the prediction of excess enthalpies of mixing. However, correctly describing excess entropies turned out to be out of their reach, due to the lack of energetically less stable, yet entropically important cluster structures. Suitable extensions of the scheme addressing this issue have been outlined.

QCE theory can go beyond the prediction of thermodynamic potentials and offer structural insight into liquids, as well. In this thesis, the ionic product of water was calculated by QCE theory, and thus the path

to the investigation of various other acid-base related phenomena has been opened.

The proposed theoretical improvements and the performed QCE calculations were made possible by the Peacemaker QCE software, which was rewritten from scratch as part of this work. The design, implementation, and use of the code are documented herein.

Some of the most fundamental properties of water and aqueous systems are dynamic in their nature, and thus beyond the reach of QCE theory. In the spirit of modern multimethod research, atomistic simulation was the second method of choice employed in this thesis and was applied to a selection of problems that occur on the nanoscale. A large part of this work was devoted to hydrogen bond and allied dynamics, which are a major driving force for processes occurring in liquid water. Further focus was set on electric field effects, which influence various applications ranging from nanofluidic devices to membrane ion channels.

On the nanoscale, the spontaneous orientational polarization of water can couple with electric field alignment, resulting in an asymmetric behavior at opposing surfaces—a situation that has previously been described as field-induced Janus interface. Here, a new and significant field polarity (sign) dependence of the dipolar reorientation dynamics in water hydration layers was uncovered. Imposition of an electric field across a nanopore can lead to differences in response times of interfacial water polarization of up to two orders of magnitude, with typical time scales being in the picosecond regime. Coupling between interfacial polarization and interfacial density relaxations was revealed, as well. The surprisingly strong asymmetry in the dynamic response at opposing surfaces is even more pronounced than the known static properties of a field-induced Janus interface.

Cavities found in nature and technology are often spherical, and water dynamics in such nanoconfinement was investigated, as well. A diffusive model was constructed by Bayesian inference from simulation data, which describes the single-particle dynamics of water molecules inside spherical cavities (fullerenes). The propagators of the diffusion model show good agreement with simulation data over four orders of magnitude, instilling great confidence in the model. There was no a priori

reason to believe in the existence of such a diffusion model, but after having established its validity, hydrogen bond kinetics could be meaningfully treated within the same diffusion model that applies to bulk water. Overall, hydrogen bond lifetimes slow down with decreasing cavity size. An attempt was made to predict hydrogen bond time correlation functions from a simple pair diffusion equation with sink and source terms corresponding to hydrogen bond breaking and formation, but the model could not be found to be reliable in spherical nanoconfinement. Various ways to improve upon this procedure have been proposed for follow-up studies.

Contents

Preliminary

Introduction

Water is ubiquitous. Amongst others, water can be found extensively as solvent in chemical reactions, it is the most abundant compound on the surface of the earth, it moderates our planet's climate, and plays an exceptional role as biological life-support system.[1,2] Most often, water is not present as neat liquid—aqueous systems, such as solutions or interfaces, are at least of equal importance.

Unsurprisingly, scientific endeavors to study water date back to antiquity[a] and a huge body of literature has evolved. The "Web of Science" has record of more than 750 000 publications between 1945 and 2017 that contain "water" or "aqueous" in their title, which corresponds to roughly 30 publications per day during that period.[4] These numbers alone are probably proof for the fascination with water and its properties, which many scientists describe with adjectives such as unique, anomalous, or outstanding.[b]

Certain authors took up on the challenge of reviewing (at least parts of) the existing literature, and a selection of textbooks shall be highlighted here. In 1940, the first comprehensive compilation of the properties of water was published by Dorsey,[6] but it was not until 1969 that the first monograph trying to relate the structure and properties of water was published by Eisenberg and Kauzmann.[1] This book marked the beginning of modern era water research and was soon followed by other excellent textbooks such as those of Ben-Naim,[2,7] Franks,[8–14] and Robinson.[15]

This thesis is of theoretical nature. The modern theory of water and aqueous systems has been developed along several distinct paths,[7] two of which shall be mentioned here, because of their relevance to the work reported later. Essentially, all paths are based on the theory of statistical

[a] The Greek philosopher Thalos of Miletus, founder of the school of natural philosophy, hypothesized that water is the "arche", the first principle or element.[3]

[b] Though some authors consider water to be only intermediate.[5]

mechanics, aspects of which are explained in Chapter 1.

Realizing the extreme difficulties in explicitly treating a large number of strongly interacting molecules, the first water theories focused on aggregates of water molecules (clusters), in which all strong interactions (hydrogen bonds) are included. These so-called mixture theories transform a one-component system into a mixture of clusters, whose interactions are much weaker than those between single water molecules. Attempts to formulate such theories can be traced back to Röntgen, and over the years many different versions have emerged.[7] While initially frowned upon as ad-hoc models, lacking experimental and theoretical justification, their integration into the statistical mechanical framework has unambiguously demonstrated their legitimacy.[16]

Roughly half of this thesis is devoted to an advanced mixture theory, Quantum Cluster Equilibrium (QCE) theory,[17] which features a high quality, quantum chemical description of the clusters. QCE theory, as well as some novel improvements, are introduced in Chapter 2. These improvements are tightly connected to the development of a fast, stable, and modern QCE program, which is documented in Chapter 4. Finally, recent applications of QCE theory are presented in Chapter 5. These include the calculation of the ionic product of water and the prediction of thermodynamic potentials of aqueous mixtures.

A different perspective on water and aqueous systems can be offered by computer simulation techniques, notably Monte Carlo (MC) and molecular dynamics (MD).[18] Unlike mixture theories, where the structure of the liquid is an input, simulations derive the structure and many other system properties as output of a given interaction potential.[7] Since the first simulations of water in the early 1970s,[19,20] aqueous systems have continued to be some of the most interesting and difficult liquids to study by such methods. In the process, simulations have transcended their original role in the validation of statistical mechanical approximations targeted at liquid structure prediction, and have become established as the dominant tool in theoretical water research.[7] Molecular dynamics simulation is the second major method employed in this thesis and an elementary introduction is given in Chapter 3.

With the advent of nanotechnology, liquid water research has faced new challenges. Interfaces and confinement are omnipresent at such

length scales and the high surface-to-bulk ratios render experiments challenging and push macroscopic theories beyond their limits. Computer simulations, on the other hand, offer atomistic insight and are ideally suited to study novel phenomena at the nanoscale. Chapter 6 is dedicated to such phenomena. Focus is set on hydrogen bond and allied dynamics, which are the ultimate driving force of liquid water dynamics.[21,22] In order to tackle nanoscale dynamics, a multitude of advanced simulation and analysis techniques were applied and are discussed in Chapter 6.

Symbols and Notation

Whenever possible, quantities are assigned symbols that are common in the respective field. While this hopefully helps the reader to find easier access to the topics covered in this thesis, it involuntarily leads to a certain overlap of symbols, due to the broad spectrum of theories and methods covered herein. For example, q may refer to a generalized coordinate in the context of classical mechanics, whereas it denotes a single-particle canonical partition function when discussing the Quantum Cluster Equilibrium theory, or an atomic charge in the description of classical force fields. For this reason, a general directory of symbols will not be given here. However, symbols are kept unique within each chapter and are introduced when they first occur.

Nevertheless, certain typographic conventions are being followed throughout all chapters. Vectors and matrices are represented by bold face letters ($\mathbf{x}$). For vectors $\mathbf{v}$, the corresponding normal face symbol $v = |\mathbf{v}|$ shall denote the vector's magnitude. Vector and matrix products are marked by a center dot ($\cdot$), whereas the cross product is indicated by a cross ($\times$). Products of scalar quantities do not receive any special multiplication symbol. The same conventions apply to vector operators, thus $\mathrm{grad}(\mathbf{x}) = \nabla\,\mathbf{x}$, $\mathrm{div}(\mathbf{x}) = \nabla\cdot\,\mathbf{x}$, and $\mathrm{rot}(\mathbf{x}) = \nabla\times\,\mathbf{x}$.

Ensemble averages are embraced by chevrons $\langle A \rangle$ and time averages receive an overline $\overline{A}$. Sets are specified by braces and an index referring to each element $\{A_i\}$. Finally, conditional probabilities are denoted by a vertical bar, which means that $p(A|B)$ describes the probability of observing A, given B.

Acknowledgment

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Barbara Kirchner, for her constant support in all scientific and administrative matters. Over the years, she has continued to provide me with time and financial support for unburdened research, opportunities for international collaboration, and scientific guidance. I am also exceptionally grateful to Prof. Dr. Alenka Luzar, for giving me several unique chances to expand my research horizon outside of Europe, for providing me with highly interesting research topics, and for sharing her extensive expertise on water and its dynamics.

Certain individuals have contributed to the completion of this work through fruitful discussions, excellent advice, and other input. Prof. Dr. Dusan Bratko's counsel was invaluable for the completion of the work on the Janus interface (Section 6.1) and he performed the GCMC simulations reported in Section 6.2. Dr. Gerhard Hummer hosted me several times, shared knowledge and computer codes for the calculation of diffusion coefficients in spherical confinement, and assisted in the interpretation of the results. My wife, Dr. Eva von Domaros, helped me in finding a common frame for presenting the broad range of topics, motivated me constantly in continuing the work, and gave considerable input on all topics. Together with Dr. Martin Brehm, she read the proofs of this thesis. My special thanks to all mentioned above.

I would also like to thank Prof. Dr. Thomas Bredow for accepting to review this thesis, as well as Prof. Dr. Peter Vöhringer and Prof. Dr. Karl G. Wagner for joining the committee. Finally, I am delighted that I could be a part of the pleasant research environment provided by the Mulliken Center for Theoretical Chemistry and its members.

Some results reported in this work have already been published in scientific journals. This applies to the prediction of the ionic product (Section 5.1)[23] and the reorientation dynamics at a Janus interface (Sec-

tion 6.1).[24] The associated text has been rephrased and partially restructured, so that it fits better into the present context. All figures were recreated, unless noted otherwise. The project on water dynamics in spherical nanoconfinement (Section 6.2) is based on my master's thesis, but has been significantly improved and extended.

Theoretical Background

1 Statistical Thermodynamics

In this chapter, a short overview on statistical thermodynamics is given, which is the foundation for all theories that follow. Obviously, a complete introduction can fill textbooks and is beyond the scope of a thesis. Thus, the intent behind this chapter is to present the concepts and notations required to cover the theories, models, and results discussed later. Basic knowledge of general thermodynamics is assumed and some concepts may not be discussed in full depth for brevity's sake. Excellent textbooks on the subject exist and readers seeking deeper knowledge are referred to those.[25–28] The following presentation is largely based on the books of Reif and McQuarrie.[26,28]

1.1 Theory of Many-Particle Systems

Many-particle systems are diverse, ranging from gases, liquids, and solids to much more complex ones, such as plants or animals. The equations governing the motion of the particles that constitute these systems (electrons, atoms, molecules, …) can with great confidence be adequately described by quantum chemistry. Yet, despite of being able to write down these equations of motion, it is extremely hard to make any useful prediction for all but the most simple systems. Problems are not only of quantitative nature and thus cannot be overcome by increasing computational power or improved algorithms; the complexity arising from the subtle interplay between a large number of particles can cause striking qualitative difficulties, as well. It is, for example, hard to understand condensation processes of gases solely from the knowledge of their microscopic interactions, and complex systems (such as the aforementioned biological systems) are even more challenging. The main goal of theory is to gain insight into the essential characteristics of such systems, to identify important relations,

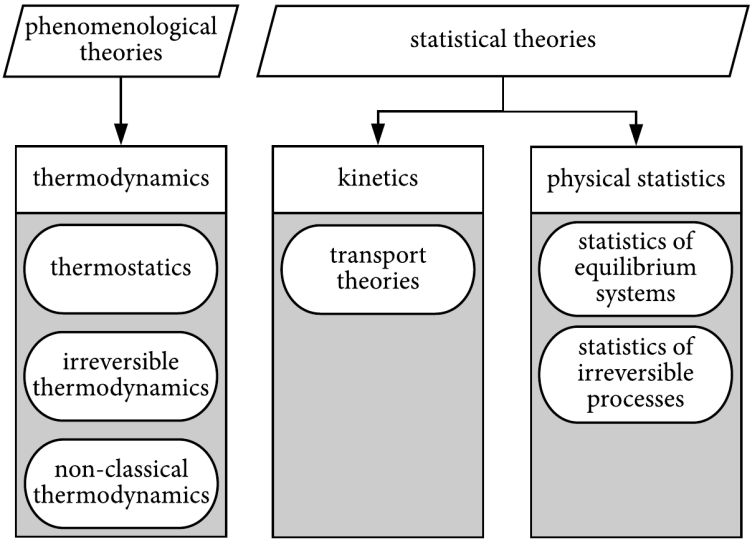


Figure 1.1: Classification of many-particle theories; scheme reproduced from ref. [26]. For further information, see original source.

and to make useful predictions at an accuracy that is not too detailed to run into any of those problems.

A system shall be called microscopic if it is of molecular dimensions ($\lesssim 1\,\mathrm{nm}$) and macroscopic if it is large enough to be visible by a conventional optical microscope ($\gtrsim 1\,\mu\mathrm{m}$). Such macroscopic systems contain a large number of atoms. Typically, they are not described in atomic detail, but by a small set of macroscopic parameters (e.g., pressure, volume, electric field strength) that fully characterize the system. If none of these macroscopic parameters change over time, the system is said to be in equilibrium.

Several theories have been developed to tackle many-particle systems. The boundaries between them are diffuse, nevertheless a classification such as the one shown in Figure 1.1 is often given. Therein, distinctions are made between phenomenological and statistical theories, or based on whether equilibrium systems are treated or non-equilibrium processes.

The majority of theories and results presented in this thesis falls into the domain of the statistics of equilibrium systems, more commonly called statistical thermodynamics. This theory deals with systems in equilibrium and its intent is to make general statements about the system based on the microscopic properties of the constituting particles and the laws of mechanics that describe their motions.

1.2 Describing the System State

A many-particle system of any complexity can be described by the laws of quantum chemistry. Thus, the system state is completely specified by a wave function $\Psi\left(q_1, \ldots, q_f\right)$, which is a function of all f coordinates (including spatial and spin variables) necessary to characterize the system. The system's description is complete if knowledge of Ψ at a given time t permits the prediction of Ψ at any other time.

This review is based on a quantum chemical description of the system, where system states are discrete. Nevertheless, it is sometimes useful to resort to a classical description, which can be an adequate approximation to the true mechanics of the system. In particular in molecular dynamics simulation (covered in Chapter 3), atoms are propagated according to the classical laws of mechanics. In such cases, the methodology presented here can still be applied after the following considerations.

A classical system is described by a set of f position coordinates $q_1, \ldots, q_f$ and f corresponding momenta $p_1, \ldots, p_f$. The number f of independent position coordinates is equal to the system's degrees of freedom. For a system consisting of N punctiform particles, for example, $f = 3N$. The set of numbers $\left\{q_1, \ldots, q_f, p_1, \ldots, p_f\right\}$ can be considered as cartesian coordinates spanning a $2f$-dimensional space called Γ-space, or phase space. Each pair (q_k, p_k) of cartesian coordinates can be discretized by dividing it into small intervals of size δq_k and δp_k, respectively. This partitions the phase space into small cells. The system state can now be characterized by numbering all cells and specifying in which one the representative point $\left(q_1, \ldots, q_f, p_1, \ldots, p_f\right)$ is located. This is formally equivalent to a quantum chemical description of the system, in which each possible quantum state is numbered ($r = 1, 2, 3, \ldots$), and the

system state is characterized by giving the number r. Such a microscopic description of a system state is called microstate. In contrast, specification of all macroscopic parameters that characterize the system is called macrostate.

1.3 Statistical Ensembles

In principle, the knowledge of the wave function Ψ at any time t permits a complete description of the system and calculation of all properties. In most many-particle systems, however, such a complete description is neither feasible nor desired and a statistical approach is more appropriate. For this purpose, an ensemble of a large number of identical systems is considered. Each of these copies of the system shall be subject to a certain number of known, external constraints. The most common statistical ensembles and the corresponding fixed external variables are summarized in Table 1.1. The microstates occupied by each member of the ensemble are usually different and so are the associated, unconstrained macroscopic parameters. It is possible, to give the probability that one of these parameters takes a specific value, which is the working principle of statistical thermodynamics. The theory allows the calculation of such probabilities based on a small number of postulates.

The fundamental postulate of statistical thermodynamics makes a statement about the a priori probabilities of finding the system in a certain microstate:

> An isolated equilibrium system can be found with equal probability in any of its accessible microstates.

A few clarifying notes are appropriate. An isolated system is a system that cannot exchange energy or mass with its surrounding. Equilibrium requires the probability to find the system in any of its states to be time-independent. Finally, only those microstates that agree with the externally applied conditions can be occupied; such states are called accessible.

Finding the probability of a microstate in the microcanonical ensemble is straightforward, given the fundamental principle. In this ensemble, the particle number N, volume V, and total energy E are fixed, in other

Table 1.1: Common statistical ensembles and the fixed macroscopic parameters that characterize them (indicated by check marks). The following symbols are used: particle number N, chemical potential μ, volume V, pressure P, total energy E, temperature T, and enthalpy H.

Ensemble	N	μ	V	P	E	T	H
microcanonical	✓		✓		✓		
canonical	✓		✓			✓	
grand canonical		✓	✓			✓	
isothermal-isobaric	✓			✓		✓	
isoenthalpic-isobaric	✓			✓			✓

words the system is isolated. Thus, all accessible states have an energy which falls into the range E and $E + \delta E$. If the energy corresponding to a state r is denoted as E_r, the probability of finding the system in such a state is given by

$$P_r = \begin{cases} C & \text{if } E < E_r < E + \delta E \\ 0 & \text{else} \end{cases} . \tag{1.1}$$

Therein, C is a constant that can be determined from the normalization condition $\sum_r P_r = 1$.

A more practical ensemble can be constructed by considering a system that is fixed in size V and composition N, but allowed to exchange heat with a heat reservoir. This system shall be called A, the heat reservoir A'. The only assumption that we make is that A is small in comparison to A' ($A \ll A'$), which typically implies that it has significantly less degrees of freedom. The total system $A^{(0)} = A + A'$ shall be isolated, so that its total energy has a value between $E^{(0)}$ and $E^{(0)} + \delta E$. If the system A is in one of its quantum states r, with associated energy E_r, we know from the conservation of energy that

$$E_r + E' = E^{(0)} , \tag{1.2}$$

where E' is the energy of system A'. If A is in a well-defined quantum

state r, the number of accessible states of the combined system $A^{(0)}$ is equal to the number of accessible states of A', which shall be denoted as $\Omega'(E') = \Omega'\left(E^{(0)} - E_r\right)$. According to the fundamental postulate, the probability P_r of finding the system A in one of its microstates r is proportional to the number of states accessible to the combined system $A^{(0)}$. Thus,

$$P_r = C'\Omega'\left(E^{(0)} - E_r\right), \tag{1.3}$$

where C' is a proportionality constant that can be determined from the normalization condition $\sum_r P_r = 1$. We shall now exploit our assumption that A is much smaller than A' and thus $E_r \ll E^{(0)}$. In this case, eq. (1.3) can be approximated by expansion of the slowly varying logarithm of $\Omega'(E')$ around $E' = E^{(0)}$:

$$\ln \Omega'\left(E^{(0)} - E_r\right) = \ln \Omega'\left(E^{(0)}\right) - \left.\frac{\partial \ln \Omega'}{\partial E'}\right|_{E^{(0)}} E_r + \dots . \tag{1.4}$$

The partial derivative

$$\left.\frac{\partial \ln \Omega'}{\partial E'}\right|_{E^{(0)}} \stackrel{!}{=} \beta \tag{1.5}$$

evaluated at $E' = E^{(0)}$ is a constant that is independent of the system state E_r. This constant is a measure of the heat reservoir's temperature. It is related to the thermodynamic temperature T by

$$\frac{1}{\beta} = kT, \tag{1.6}$$

where $k = 1.380\,658 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ is Boltzmann's constant. The physical interpretation of eq. (1.5) is that the temperature of the heat reservoir is constant and unaffected by energy exchange with system A. An ensemble that is characterized by constant particle number N, volume V, and temperature T is called canonical ensemble. This is the most common ensemble used within this thesis and shall therefore be treated in more detail.[a]

[a] In particular, this is the ensemble underlying QCE theory (Chapter 2). The only other ensembles in use are the microcanonical ensemble, where kinetic properties of systems

Finally, we can truncate expression (1.4), arriving at

$$\ln \Omega'\left(E^{(0)} - E_r\right) \approx \ln \Omega'\left(E^{(0)}\right) - \beta E_r \,, \tag{1.7}$$

and rearrange, to obtain (assuming equality from now on)

$$\Omega'\left(E^{(0)} - E_r\right) = \Omega'\left(E^{(0)}\right) e^{-\beta E_r} \,. \tag{1.8}$$

Since $\Omega'\left(E^{(0)}\right)$ is independent of r, eq. (1.3) turns into

$$P_r = C e^{-\beta E_r} \,, \tag{1.9}$$

where the proportionality constant C can be obtained through the normalization condition $\sum_r P_r = 1$. Its inverse

$$Q = C^{-1} = \sum_r e^{-\beta E_r} \tag{1.10}$$

is called (canonical) partition function. Probabilities and partition functions are the central properties of statistical thermodynamics, allowing the calculation of any desired thermodynamic quantity.

1.4 Properties of Partition Functions

If a system is in contact with a heat reservoir, the probability of finding it in any of its states is given by the canonical distribution

$$P_r = \frac{e^{-\beta E_r}}{\sum\limits_r e^{-\beta E_r}} \,, \tag{1.11}$$

where the sum runs over all accessible states r. Calculating averages in such a situation is relatively easy. Consider, for example, the average system energy $\langle E \rangle$, where the chevron notation $\langle \ldots \rangle$ denotes an ensemble

can be calculated from molecular dynamics simulation without bias by commonly employed thermostating techniques (Chapter 3), and the grand canonical ensemble, which permits the treatment of particle number fluctuations (Chapter 6).

average,

$$\langle E \rangle = \sum_r P_r E_r = \frac{\sum_r e^{-\beta E_r} E_r}{\sum_r e^{-\beta E_r}} . \tag{1.12}$$

Eq. (1.12) can be simplified by exploiting

$$\sum_r e^{-\beta E_r} E_r = -\sum_r \frac{\partial}{\partial \beta}\left(e^{-\beta E_r}\right) = -\frac{\partial}{\partial \beta} Q , \tag{1.13}$$

which leads to

$$\langle E \rangle = -\frac{1}{Q}\frac{\partial Q}{\partial \beta} = -\frac{\partial \ln Q}{\partial \beta} . \tag{1.14}$$

Ensemble averages, such as eq. (1.14), can be related to macroscopically observable quantities. In the preceding case, the average ensemble energy can be identified as the internal energy of the system U. Relations that are similar to eq. (1.14) can be derived for other thermodynamic quantities, as well. Some examples are:

$$\text{internal energy} \quad U = -\frac{\partial \ln Q}{\partial \beta} \tag{1.15}$$

$$\text{Helmholtz energy} \quad A = -\frac{\ln Q}{\beta} \tag{1.16}$$

$$\text{Gibbs energy} \quad G = -\frac{\ln Q}{\beta} + PV \tag{1.17}$$

$$\text{entropy} \quad S = k\left(\ln Q + \beta \frac{\partial \ln Q}{\partial \beta}\right) \tag{1.18}$$

$$\text{pressure} \quad P = kT\frac{\partial \ln Q}{\partial V} . \tag{1.19}$$

For many-particle systems, knowledge of the complete set of energy eigenvalues E_r is often impossible. However, a common and often good approximation is to express the system energy as a sum of individual energies. For a system of N independent, distinguishable particles (e.g.,

atoms in a perfect crystal),

$$E_r = \underbrace{\varepsilon_i^a + \varepsilon_j^b + \varepsilon_k^c + \dots}_{N \text{ terms}}, \tag{1.20}$$

where ε_i^a denotes the i'th energy state of the particle labeled a (e.g., at lattice position a). In such a scenario, it is easy to see that the system partition function

$$Q = \sum_r e^{-\beta E_r} = \sum_{i,j,k,\dots} e^{-\beta\left(\varepsilon_i^a + \varepsilon_j^b + \varepsilon_k^c + \dots\right)} \tag{1.21}$$

becomes a product of individual partition functions q, by splitting the summation into individual sums for each particle:

$$Q = q^a q^b q^c \cdots, \tag{1.22}$$

where

$$q^a = \sum_i e^{-\beta \varepsilon_i^a} \tag{1.23}$$

and so on. Furthermore, if all energy states of the system are equal, as it is the case in a perfect crystal,

$$Q = q^N. \tag{1.24}$$

In general, however, particles are indistinguishable, thus

$$E_r = \underbrace{\varepsilon_i + \varepsilon_j + \varepsilon_k + \dots}_{N \text{ terms}} \tag{1.25}$$

and

$$Q = \sum_{i,j,k,\dots} e^{-\beta\left(\varepsilon_i + \varepsilon_j + \varepsilon_k + \dots\right)}. \tag{1.26}$$

Please note the lack of distinguishing superscripts. Unlike above, the preceding sum cannot be split into separate terms, because of restrictions

imposed on the allowed energy states by the Pauli exclusion principle and by the indistinguishability of particles.

In the case of fermions, for example, no two particles can occupy the same single-particle energy state, thus the summation over all energy states $i, j, k, \ldots$ is not independent anymore. For bosons, such a restriction does not exist, but they cause another problem. Consider, for example, the case, where one particle is in the energy state ε_i and all $N-1$ others are in another state $\varepsilon_j \neq \varepsilon_i$. The system energy is given by

$$E_r = \varepsilon_i + \underbrace{\varepsilon_j + \varepsilon_j + \ldots}_{N-1 \text{ terms}} . \tag{1.27}$$

Unrestricted evaluation of eq. (1.26) would produce N such sums, which should only be counted once, though, because they all correspond to the same system energy state. In the extreme case where all individual energies are different, $N!$ permutations exist for each energy state that fulfills this condition.

Exact evaluation of eq. (1.26) would lead to two sets of quantum statistics, called Bose–Einstein (for bosons) and Fermi–Dirac statistics (for fermions), which shall not be covered here. In the high-temperature limit, they both converge to the classical Boltzmann statistics, which can be rationalized as follows. Suppose that the number of accessible quantum states available to any of the particles is much larger than the number of particles. In such a case, it is highly unlikely that two particles are in the same quantum state. Thus, a good approximation to eq. (1.26) would be to sum over all indices $i, j, k, \ldots$ independently and to divide by $N!$ to correct for the overcounting. If all energy states of the particles are equal, this leads to the partition function of N indistinguishable particles according to Boltzmann statistics:

$$Q = \frac{1}{N!} q^N . \tag{1.28}$$

The number of translational quantum states alone is usually sufficient to guarantee the validity of this assumption. A criterion that can be used to

judge whether this is truly the case is

$$\left(\frac{N}{V}\right)^{1/3} \ll \Lambda \tag{1.29}$$

where Λ is the thermal de Broglie wavelength of a particle with mass m, given by

$$\Lambda = \frac{h}{\sqrt{2\pi m k T}} \,. \tag{1.30}$$

Therein, $h = 6.626\,070\,04 \times 10^{-34}$ J s is Planck's constant.

One further approximation is often made: The energy of a particle is decomposed into contributions arising from individual degrees of freedom which are assumed to be independent. For a molecule, for example,

$$\varepsilon_r = \varepsilon_i^{\text{trans}} + \varepsilon_j^{\text{elec}} + \varepsilon_k^{\text{rot}} + \varepsilon_l^{\text{vib}} \,, \tag{1.31}$$

where $\varepsilon_i^{\text{trans}}$ denotes the i'th energy state of translation (and so on for electronic motions, rotations, and vibrations). In these cases, individual partition functions can be further simplified. Since these energy states are distinguishable, the same reasoning as above can be applied to obtain

$$q = q^{\text{trans}} q^{\text{elec}} q^{\text{rot}} q^{\text{vib}} \,. \tag{1.32}$$

In the following section, expressions for individual partition functions are given.

1.5 Individual Partition Functions

1.5.1 Translational Partition Function

Translational motions are typically modeled by particles of mass m in a cubic container with side length a and volume $V = a^3$. The corresponding energy states are given by

$$\varepsilon_{n_x n_y n_z}^{\text{trans}} = \frac{h}{8ma^2}\left(n_x^2 + n_y^2 + n_z^2\right) \quad n_x, n_y, n_z = 1, 2, \ldots \,. \tag{1.33}$$

These can be shown to yield the translational partition function

$$q^{\text{trans}} = \frac{V}{\Lambda^3}\,, \tag{1.34}$$

where Λ is the thermal de Broglie wave length (1.30).

If the particles are non-punctiform and have an impenetrable volume v, the volume $V^{\text{ex}} = Nv$ must be subtracted from the phase volume, since it is inaccessible to any of the particles. In this case,

$$q^{\text{trans}} = \frac{V - V^{\text{ex}}}{\Lambda^3}\,. \tag{1.35}$$

1.5.2 Electronic Partition Function

At 'conventional' temperatures, that is, at temperatures for which the first excited state energy $\varepsilon_{e1} \gg kT$, only the electronic ground state of a particle is significantly occupied. This leads to an easy expression of the electronic partition function:

$$q^{\text{elec}} = e^{-\beta\varepsilon_{e0}}\,, \tag{1.36}$$

where ε_{e0} denotes the electronic ground state energy.

If the electronic ground state is degenerate, the following expression must be considered instead:

$$q^{\text{elec}} = g_{e0}\,e^{-\beta\varepsilon_{e0}}\,, \tag{1.37}$$

where g_{e0} is the degeneracy of the electronic ground state.

1.5.3 Rotational Partition Function

Rotational motions are often described by the model of a rigid rotor. The energy levels of a linear rigid rotor (e.g., a diatomic molecule), characterized by its moment of inertia I, are given by

$$\varepsilon_n^{\text{rot}} = \frac{\hbar^2 n(n+1)}{2I} \qquad n = 0, 1, 2, \ldots\,, \tag{1.38}$$

where $\hbar = h/2\pi$. Each level has a degeneracy $g_n = 2n + 1$. At high temperatures $T \gg \Theta^{\mathrm{rot}}$, where

$$\Theta^{\mathrm{rot}} = \frac{\hbar^2}{2Ik} \tag{1.39}$$

is called rotational temperature, the rotational partition function can be shown to be

$$q^{\mathrm{rot}} = \frac{T}{\sigma\Theta^{\mathrm{rot}}}\,. \tag{1.40}$$

Therein, the factor σ is called rotational symmetry number. It counts the number of indistinguishable orientations of a molecule.

For non-linear rotors, the rotational partition function can be shown to take on the form

$$q^{\mathrm{rot}} = \frac{1}{\sigma}\sqrt{\frac{\pi T^3}{\Theta_1^{\mathrm{rot}}\Theta_2^{\mathrm{rot}}\Theta_3^{\mathrm{rot}}}}\,, \tag{1.41}$$

where $\Theta_{1/2/3}^{\mathrm{rot}}$ are the rotational temperatures corresponding to the three principal moments of inertia, according to eq. (1.39).

1.5.4 Vibrational Partition Function

Molecular vibrations are often considered in the harmonic approximation. The energy states of a single harmonic oscillator (e.g., a diatomic molecule) are given by:

$$\varepsilon_n^{\mathrm{vib}} = \left(n + \frac{1}{2}\right)h\nu \quad n = 0, 1, 2, \ldots\,, \tag{1.42}$$

with ν being the vibrational frequency of the oscillator. The partition function resulting from summation of these energy states can be shown to be

$$q^{\mathrm{vib}} = \frac{e^{-\Theta^{\mathrm{vib}}/2T}}{1 - e^{-\Theta^{\mathrm{vib}}/T}}\,, \tag{1.43}$$

where

$$\Theta^{\mathrm{vib}} = \frac{h\nu}{k} \tag{1.44}$$

is called vibrational temperature, in analogy to eq. (1.39).

The vibrational energy of polyatomic molecules is given by the sum over all modes of vibration:

$$\varepsilon_r^{\text{vib}} = \sum_{m=1}^{f^{\text{vib}}} \left(n_m + \frac{1}{2} \right) h\nu_m \quad \{n_m = 0, 1, 2, \ldots\}, \tag{1.45}$$

where ν_m is the vibrational frequency of the m'th mode of vibration, f^{vib} is the number of vibrational degrees of freedom ($3N - 5$ for a linear molecule and $3N - 6$ for a non-linear one), and r counts all vibrational state variations. Since all these modes are distinguishable, the vibrational partition function of a polyatomic molecule simply becomes a product:

$$q^{\text{vib}} = \prod_{m=1}^{f^{\text{vib}}} \frac{e^{-\Theta_m^{\text{vib}}/2T}}{1 - e^{-\Theta_m^{\text{vib}}/T}}, \tag{1.46}$$

where Θ_m^{vib} are the characteristic vibrational temperatures according to eq. (1.44).

1.6 The van der Waals Equation

In this section, the famous van der Waals equation shall be derived from statistical thermodynamics, because QCE systems can be considered to be multi-component van der Waals fluids and a comparison with van der Waals gases is beneficial.

In the van der Waals picture, particles are assumed to be hard spheres which interact due to an attractive potential of the type

$$u(r) = \begin{cases} \infty & r < \sigma \\ -\epsilon \left(\frac{\sigma}{r} \right)^6 & r \geq \sigma \end{cases}, \tag{1.47}$$

where σ is the particle separation at contact (twice the van der Waals radius) and ϵ is the depth of the attractive well. Instead of explicitly considering all possible interactions between particles, a mean field approach shall be chosen, in which each particle moves independently in

the average potential field of all other particles. By doing so, the system energy can be expressed as sum of individual energies and the system partition function is given by

$$Q = \frac{1}{N!} q^N \,, \tag{1.48}$$

according to eq. (1.28).

To calculate the mean field energy per particle $\varepsilon^{\mathrm{mf}}$, an integration of the type

$$\varepsilon^{\mathrm{mf}} = \frac{1}{2} \frac{N}{V} \int_{\sigma}^{\infty} g(r) u(r)\, 4\pi r^2 dr \tag{1.49}$$

must be performed. Therein, N/V denotes the average particle density (assuming a homogeneous distribution), $g(r)$ is the density and temperature dependent radial pair distribution function of a hard sphere fluid, $u(r)$ is the potential given by eq. (1.47), and the factor of $\frac{1}{2}$ arises because this energy is shared between two particles. To facilitate the integral, we shall make use of the low-density approximation

$$g(r) = \theta(r - \sigma) = \begin{cases} 0 & r < \sigma \\ 1 & r \geq \sigma \end{cases} \,, \tag{1.50}$$

where $\theta(x)$ denotes the Heaviside step function, arriving at

$$\varepsilon^{\mathrm{mf}} = -a \frac{N}{V} \,, \quad a = \epsilon \frac{2\pi\sigma^3}{3} = \epsilon b \,. \tag{1.51}$$

The constant

$$b = \frac{2}{3} \pi \sigma^3 \tag{1.52}$$

is equal to half the volume of a sphere of radius σ. Such a sphere represents the volume excluded from translation for a hard sphere particle of radius $\frac{\sigma}{2}$ (Figure 1.2). The factor of $\frac{1}{2}$ arises to avoid counting each excluded volume twice.

With these considerations, the electronic and translational partition

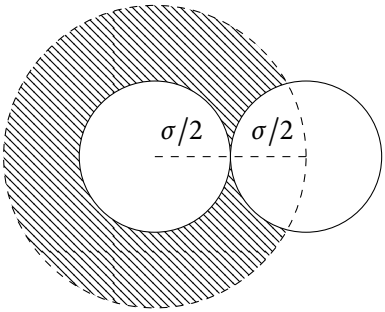


Figure 1.2: Two hard spheres of radius $\frac{\sigma}{2}$ in close contact. The center of a particle cannot be in the shaded area, making the excluded volume a sphere of size $\frac{4}{3}\pi\sigma^3$. This volume is shared by both particles.

functions of a van der Waals gas can be given by:

$$q^{\text{elec}} = e^{-\beta\varepsilon^{\text{mf}}} \tag{1.53}$$

$$q^{\text{trans}} = \frac{V - V^{\text{ex}}}{\Lambda^3}\,. \tag{1.54}$$

Therein, the models of a particle in a box with volume V and excluded volume $V^{\text{ex}} = Nb$, as well as electronic ground state energy $\varepsilon_{e0} = \varepsilon^{\text{mf}}$ were used (see Section 1.5.1).

Van der Waals gases are not limited to atoms, though, but may have internal (vibrational and rotational) degrees of freedom, as well. If the models of a harmonic oscillator and rigid rotor are used (Section 1.5.4), the partition function of a single particle reads

$$q(V,\beta) = q^{\text{trans}}(V,\beta)\, q^{\text{elec}}(V,\beta)\, q^{\text{rot}}(\beta)\, q^{\text{vib}}(\beta)\,, \tag{1.55}$$

where the functional arguments are included to emphasize the dependence of the individual partition functions on temperature and volume. In the following, we shall derive the equation of state, which is obtained from the relation for pressure

$$P = kT\frac{\partial \ln Q}{\partial V} = kTN\frac{\partial \ln q}{\partial V}\,, \tag{1.56}$$

where eq. (1.48) was inserted to obtain the rightmost part. It is obvious that contributions of eq. (1.55) arising from internal degrees of freedom vanish, since they do not depend on the volume. Thus, in the following,

it is sufficient to study the atomic part of eq. (1.55):

$$\begin{aligned} q^{\text{atomic}} &= q^{\text{trans}} q^{\text{elec}} \\ \ln q^{\text{atomic}} &= \ln\left(V - Nb\right) + \frac{Na}{VkT} - 3\ln\Lambda\,. \end{aligned} \tag{1.57}$$

Simple insertion into eq. (1.56) leads to the famous van der Waals equation of state:

$$P = \frac{NkT}{V - Nb} - \frac{N^2 a}{V^2}\,. \tag{1.58}$$

Before continuing, we shall now consider the energy of a cluster containing n van der Waals particles. This expression is needed for comparison with QCE theory, discussed in Chapter 2. For such a cluster, one would have to consider the interactions between the n particles constituting the cluster and all $N - n$ particles outside, that is,

$$\varepsilon_{\text{clust}}^{\text{mf}} = -na\frac{N - n}{V}\,. \tag{1.59}$$

However, if $n \ll N$, the following approximation holds:

$$\varepsilon_{\text{clust}}^{\text{mf}} \approx n\varepsilon^{\text{mf}} = -na\frac{N}{V}\,. \tag{1.60}$$

1.7 Binary van der Waals Mixtures

Since QCE theory will be presented for binary mixtures, it is useful to derive the equations describing a binary van der Waals system, as well. Binary mixtures of independent particles are easy to treat within the framework presented in Section 1.4. Since components of a mixture are distinguishable, the canonical partition function of a binary system is given by the product of two single-component systems (1.28):

$$Q = \frac{1}{N_1!} q_1^{N_1} \frac{1}{N_2!} q_2^{N_2}\,. \tag{1.61}$$

Therein, N_1 and N_2 are the number of particles of components 1 and 2, respectively, and q_1 and q_2 are the individual partition functions of these

components.

By inserting eq. (1.57) twice into eq. (1.61), the partition function of a binary van der Waals gas (once more omitting internal degrees of freedom) is given by

$$Q^{\text{atomic}} = \frac{1}{N_1!N_2!}\left(\frac{V - N_1 b_1 - N_2 b_2}{\Lambda_1^3 \Lambda_2^3}\right)^{N_1+N_2} e^{-\beta\left(N_1\varepsilon_1^{\text{mf}} + N_2\varepsilon_2^{\text{mf}}\right)}, \quad (1.62)$$

where the indices 1 and 2 refer to the components 1 and 2, respectively.

The mean field energies $\varepsilon_{1/2}^{\text{mf}}$ have contributions from both components:

$$\varepsilon_1^{\text{mf}} = -\frac{1}{V}\left(N_1 a_1 + N_2 a_{12}\right) \quad (1.63)$$

$$\varepsilon_2^{\text{mf}} = -\frac{1}{V}\left(N_2 a_2 + N_1 a_{12}\right). \quad (1.64)$$

This is because each particle of component 1 interacts with all other particles of component 1 (described by the term $\frac{N_1}{V} a_1$) and all particles of component 2 (leading to the term $\frac{N_2}{V} a_{12}$). The same is true for the particles of component 2. The mean field parameter a_{12} describes the mixed interactions between components 1 and 2. It is an additional parameter in the binary equation of state that cannot be derived from the properties of the pure components. Nevertheless, lack of reference data frequently motivates the use of mixing rules, such as

$$a_{12} = \sqrt{a_1 a_2}(1 - k_{12}), \quad (1.65)$$

where k_{12} is called binary interaction parameter. In this case, k_{12} is an empirical measure of the ideality of the binary system that can be assumed to be zero as a first approximation. Finally, note that the excluded volume of a binary system is simply equal to the sum $N_1 b_1 + N_2 b_2$ of the single-component terms.

Differentiation of eq. (1.62) with respect to volume would lead to the equation of state of a binary van der Waals mixture. This shall not be presented here, because it is not required to gain a better understanding of QCE theory anymore (instead, see for example, ref. [29]). However, for future reference, we shall now give the mean field energy of a cluster

containing n_1 particles of species 1 and n_2 particles of species 2. Following the lines of thought that led to eq. (1.60),

$$\begin{aligned}\varepsilon_{\text{clust}}^{\text{mf}} &\approx n_1\varepsilon_1^{\text{mf}} + n_2\varepsilon_2^{\text{mf}} \\ &= -\frac{1}{V}\left(n_1N_1a_1 + n_1N_2a_{12} + n_2N_2a_2 + n_2N_1a_{12}\right).\end{aligned} \tag{1.66}$$

From the considerations above, it is clear that a generalization to mixtures with more than two components is straightforward, albeit hardly ever presented in the literature. Keeping in mind a possible future extension of QCE theory to more complex systems, this generalization shall be presented in the following.

Consider a mixture of K van der Waals components. Its partition functions reads

$$Q = \prod_{i=1}^{K} \frac{1}{N_i!} {q_i}^{N_i}, \tag{1.67}$$

where N_i denotes the number of particles of component i and q_i is their individual partition function. In the following, we shall once more omit internal degrees of freedom, arriving at

$$Q^{\text{atomic}} = \prod_{i=1}^{K} \frac{1}{N_i!} \left(\frac{V - \sum_{j=1}^{K} N_j b_j}{\Lambda_i^3} \right)^{N_i} \exp\left(-\beta N_i \varepsilon_i^{\text{mf}}\right), \tag{1.68}$$

where $\varepsilon_i^{\text{mf}}$ is the mean field energy given by

$$\varepsilon_i^{\text{mf}} = -\frac{1}{V} \sum_{j=1}^{K} N_j a_{ji}. \tag{1.69}$$

After taking the logarithm,

$$\ln Q^{\text{atomic}} = \sum_{i=1}^{K} N_i \ln\left(V - \sum_{j=1}^{K} N_j b_j \right) - 3N_i \ln \Lambda_i - \beta_i N_i \varepsilon_i^{\text{mf}} - \ln N_i!, \tag{1.70}$$

and insertion into eq. (1.56), the van der Waals equation of state for a multi-component system emerges:

$$P = \sum_{i=1}^{K} \frac{N_i kT}{V - \sum_{j=1}^{K} N_j b_j} - \frac{N_i}{V^2} \sum_{j=1}^{K} N_j a_{ji} \,. \tag{1.71}$$

The expression reduces nicely to eq. (1.58) for $K = 1$.

2 Quantum Cluster Equilibrium Theory

In the following chapter, QCE theory and its extension to binary systems (bQCE) shall be described. QCE theory dates back to early work of Ludwig et al.,[30] but was fully presented for the first time by Weinhold in 1998.[17] The extension to binary systems was described by Brüssel et al. in 2011 and the following presentation initially follows their derivation.[31] Afterwards, several theoretical improvements are proposed in Section 2.7. Algorithmic and methodological aspects are presented in Chapter 4.

In his seminal work, Weinhold argued[17] that molecular clusters can be considered as intermediates between monomeric vapor phases and fully aggregated condensed systems. Their finite size permits a quantum chemical treatment, opening the path to an accurate description of liquid phases. This is possible because of the essential continuity of gases and liquids,[32] which demands that a cluster picture that is valid in dense vapor phases should be transferable to liquids by simple variation of the cluster distribution to reflect the changes in pressure and temperature. Thus, in the extreme limit that one would have a complete set of clusters available, a quantitative structural and thermodynamic description of the liquid phase should be possible. Naturally, such a complete cluster set is beyond reach, but a representative set of dominant clusters in a certain pressure and temperature range should still provide an adequate description of liquid properties.

These lines of thought establish the foundation of QCE theory, which can be considered as a mixture theory (a mixture of clusters), as opposed to one-component or continuum theories.[7] Since its original publication, QCE has been applied to a myriad of neat and binary systems, including water,[23,33–43] alcohols,[30,36,44–50] hydrogen halides,[51–53] simple amides,[54–57] organic acids,[58,59] ammonia,[60,61] liquid sulfur,[62] dimethyl

sulfoxide/water,[31] methanol/water,[63–65] and various other mixtures of common solvents.[66] For excellent reviews of these works, the reader is referred to refs. [67, 68]

2.1 Cluster Equilibrium

The central idea of Quantum Cluster Equilibrium theory is the chemical equilibrium between clusters; in the binary case between clusters of two components *a* and *b*. In this context, clusters are aggregates of atoms or molecules. Their size is defined as the number of subunits that they are built from. A cluster's size may be as small as one, in which case it is called a monomer. In this section, the monomers of components *a* and *b* are labeled A and B, respectively. Clusters may be neat (of type A_n or B_n) or mixed (of type A_nB_m).

The general cluster equilibrium reaction equation reads

$$n_i^a \, \mathrm{A} + n_i^b \, \mathrm{B} \rightleftharpoons \mathrm{C}_i \,, \tag{2.1}$$

where n_i^a and n_i^b denote the number of monomers of components *a* and *b* present in cluster C_i, and i is an index running from 1 to K, the total number of clusters. Two notes about this equilibrium reaction shall be made. First, the monomers are part of the cluster set, as well, and must be assigned a label C_i. Typically, A is called C_1 and B is called C_2. Thus, we can rewrite eq. (2.1) without loss of generality:

$$n_i^a \, \mathrm{C}_1 + n_i^b \, \mathrm{C}_2 \rightleftharpoons \mathrm{C}_i \,. \tag{2.2}$$

Second, this formulation permits the presence of multiple clusters of a given composition, representing different conformations or motifs (for example, two monomers in *cis* and *trans* configuration, or a cluster with different coordination patterns, such as chain and ring).

In order to derive the QCE equations, two neat clusters of each component must be chosen for reference's sake. Typically, this role falls to the monomers C_1 and C_2, and we will proceed accordingly. If there are multiple monomers of a given component, the choice is arbitrary.

2.2 The System Partition Function

The QCE equations are derived in the canonical ensemble, with constant number of particles N^{tot}, volume V, and temperature T. Note that the number of particles N^{tot} refers to the total number of monomers $N_1^{\mathrm{tot}} + N_2^{\mathrm{tot}}$ of components a and b (labeled 1 and 2, as discussed above). This fixed number of monomers may react according to eq. (2.2).

It is one of the main tasks of the QCE method to find the set of equilibrium populations $\{N_i\}$ of each cluster C_i. Once these are known, the calculation of the system partition function Q is straightforward. Given the individual atomic or molecular partition functions of each cluster, q_i, and assuming Boltzmann statistics, the system partition function reads

$$Q = \prod_{i=1}^{K} \frac{1}{N_i!} q_i^{N_i} , \tag{2.3}$$

which is the standard expression for a mixture of K components of independent, indistinguishable particles whose numbers are N_i.

2.3 Conserved Quantities

Several quantities are conserved in a QCE calculation, which are important constraints on the QCE solutions. These shall be briefly summarized here for easy reference in later sections. The total number of particles $N^{\mathrm{tot}} = N_1^{\mathrm{tot}} + N_2^{\mathrm{tot}}$ is fixed by choice of the canonical ensemble:

$$N^{\mathrm{tot}} = \sum_{i=1}^{K} \left(n_i^a + n_i^b \right) N_i = N_1^{\mathrm{tot}} + N_2^{\mathrm{tot}} = \mathrm{const.} \tag{2.4}$$

Populations are typically reported as so-called monomer-normalized populations $\tilde{N}_i$, which represent the fraction of monomers populating one particular cluster C_i. By rearranging eq. (2.4), their definition follows:

$$1 = \sum_{i=1}^{K} \frac{n_i^a + n_i^b}{N^{\mathrm{tot}}} N_i = \sum_{i=1}^{K} \tilde{N}_i . \tag{2.5}$$

A similar conservation equation exists for the total mass M^{tot} of the system, since a cluster's mass

$$m_i = n_i^a m_1 + n_i^b m_2 \tag{2.6}$$

depends linearly on the monomer masses m_1 and m_2:

$$\begin{aligned} M^{\text{tot}} &= \sum_{i=1}^{K} m_i N_i = \sum_{i=1}^{K} \left(n_i^a m_1 + n_i^b m_2 \right) N_i \\ &= m_1 N_1^{\text{tot}} + m_2 N_2^{\text{tot}} = \text{const.} \end{aligned} \tag{2.7}$$

The mass analog of eq. (2.5) is

$$1 = \sum_{i=1}^{K} \frac{m_i}{M^{\text{tot}}} N_i \,. \tag{2.8}$$

Finally, the total cluster volume V^{tot} is conserved. In QCE theory, all clusters are treated as hard particles with an impenetrable volume v_i. These cluster volumes are assumed to be multiple integers of the monomer volumes v_1 and v_2

$$v_i = n_i^a v_1 + n_i^b v_2 \,. \tag{2.9}$$

Because of this linear dependence, equations analogous to eq. (2.7) exist for the volume, as well:

$$\begin{aligned} V^{\text{tot}} &= \sum_{i=1}^{K} v_i N_i = \sum_{i=1}^{K} \left(n_i^a v_1 + n_i^b v_2 \right) N_i \\ &= v_1 N_1^{\text{tot}} + v_2 N_2^{\text{tot}} = \text{const.} \end{aligned} \tag{2.10}$$

We shall see in the following sections, that two independent constraints are required in order to solve the QCE equations of a binary system, for which conservation of mass and conservation of particle numbers are chosen. Further constraints will be needed in a possible future extension of QCE theory to multi-component systems. Here, one particular constraint shall be highlighted: conservation of charge. Such a constraint could open the path for the investigation of charged ternary mixtures, with promising applications for ionic liquids or deep eutectic solvents.

2.4 Cluster Partition Functions

The individual cluster partition functions q_i required in the calculation of the system partition function (2.3) are readily available from the models presented in Section 1.5. Each cluster's energy ε_i may be decomposed into a sum of contributions arising from translational, electronic, vibrational, and rotational degrees of freedom:

$$\varepsilon_i = \varepsilon_i^{\text{trans}} + \varepsilon_i^{\text{elec}} + \varepsilon_i^{\text{vib}} + \varepsilon_i^{\text{rot}} , \tag{2.11}$$

which leads to a product of the corresponding partition functions q_i:

$$q_i = q_i^{\text{trans}}\, q_i^{\text{elec}}\, q_i^{\text{vib}}\, q_i^{\text{rot}} . \tag{2.12}$$

The translational motions of a cluster are modeled by a particle in a box with the accessible volume $V - V^{\text{ex}}$ according to eq. (1.35):

$$q_i^{\text{trans}} = \frac{V - V^{\text{ex}}}{\Lambda_i^3} . \tag{2.13}$$

In QCE theory, V^{ex} has traditionally been a scaled version of the total cluster volume V^{tot} introduced in the previous section:

$$V^{\text{ex}} = b_{\text{xv}} V^{\text{tot}} = b_{\text{xv}} \left(\nu_1 N_1^{\text{tot}} + \nu_2 N_2^{\text{tot}} \right) . \tag{2.14}$$

In principle, V^{tot} could be used directly in the preceding equation, since it represents the volume excluded from translation due to the impenetrable nature of the clusters. However, molecular volumes are no quantum-chemical observables and their definition is somewhat arbitrary. Furthermore, calculations typically depend on atomic radii which may be subject to errors that enter cubically into the volume. For these reasons, the dimensionless scaling parameter b_{xv} was introduced, which corrects any errors made in the calculation of the total cluster volume. The resulting quantity V^{ex} is called excluded volume and b_{xv} is the excluded volume scaling parameter. It is one of only two empirical correction parameters that enter into the QCE equations. Its value is typically determined by fixing the density at ambient conditions to the experiment. Note that

after the considerations that led to the derivation of the binary van der Waals equation, one would have expected a different definition of V^{ex}, which is presented in Section 2.7.

The electronic partition function is governed by the electronic ground state energy of the system $\varepsilon_i^{\mathrm{elec}}$, which is typically the only populated electronic energy level at ambient and close-to-ambient conditions. The corresponding cluster partition function according to eq. (1.36) is given by

$$q_i^{\mathrm{elec}} = e^{-\beta\varepsilon_i^{\mathrm{elec}}} . \tag{2.15}$$

The electronic energy $\varepsilon_i^{\mathrm{elec}}$ is equal to the sum of the adiabatic binding energy $\Delta\varepsilon_i^{\mathrm{bind}}$ and the mean field energy $\varepsilon_i^{\mathrm{mf}}$,

$$\varepsilon_i^{\mathrm{elec}} = \Delta\varepsilon_i^{\mathrm{bind}} + \varepsilon_i^{\mathrm{mf}} , \tag{2.16}$$

both of which shall be explained in the following.

In QCE theory, the arbitrary zero point of energy is set to the energy of infinitely separated monomers in their relaxed geometries. Thus, instead of directly using total electronic energies obtained from quantum chemical calculations E_i^{QC}, the adiabatic binding energy

$$\Delta\varepsilon_i^{\mathrm{bind}} = E_i^{\mathrm{QC}} - n_i^a E_1^{\mathrm{QC}} - n_i^b E_2^{\mathrm{QC}} \tag{2.17}$$

is used to account for the change in electronic energy upon cluster formation. Additional energy contributions arise from inter-cluster interactions upon approach of the infinitely separated clusters. These have traditionally been accounted for by a density and cluster size proportional mean field energy

$$\varepsilon_i^{\mathrm{mf}} = -a_{\mathrm{mf}}\left(n_i^a + n_i^b\right)\frac{N^{\mathrm{tot}}}{V} . \tag{2.18}$$

The proportionality constant a_{mf} is the second empirical correction parameter that enters into QCE theory. It is typically determined by comparing experimentally and theoretically predicted boiling points at ambient pressure. Note that eq. (2.18) is not the same as the mean field energy of a binary van der Waals cluster (1.66) introduced earlier. This discrepancy

will be discussed in Section 2.7.

Vibrations are considered to be harmonic in their nature and the vibrational partition function is that of a harmonic oscillator. It is given by eq. (1.46) or eq. (1.43) for diatomic molecules. No additional parameters or modifications are made for this degree of freedom.

Rotational motions are treated by the model of a rigid rotor, with the rotational symmetry number σ and the principal moments of inertia I_1, I_2, and I_3. The corresponding cluster partition function is given by eq. (1.41) or eq. (1.40) for linear molecules. Again, no modifications are made and no additional parameters are introduced.

2.5 Determining Populations

As already mentioned in the beginning of this chapter, one of the main tasks of the QCE method is the determination of the set of equilibrium cluster populations $\{N_i\}$. In the canonical ensemble, this is the cluster distribution that minimizes the system's Helmholtz energy

$$\begin{aligned} A &= -kT \ln Q = -kT \sum_{i=1}^{K} N_i \ln q_i - \ln N_i! \\ &= -kT \sum_{i=1}^{K} N_i \ln q_i - N_i \ln N_i + N_i \,, \end{aligned} \tag{2.19}$$

where Stirling's approximation $\ln N! \approx N \ln N - N$ was used to obtain the bottom line. This function shall be minimized under the constraints of a fixed total particle number (2.4) and fixed total mass (2.7). The reason for choosing two constraints instead of only one is related to the binary nature of the system and will be apparent shortly.

In this section, the constraints shall be written as

$$g = N^{\text{tot}} - \sum_{i=1}^{\text{K}} \left(n_i^a + n_i^b \right) N_i = 0 \quad \text{and} \tag{2.20}$$

$$h = M^{\text{tot}} - \sum_{i=1}^{\text{K}} m_i N_i = 0 \,. \tag{2.21}$$

After introducing the Lagrange multipliers λ_g and λ_h, the Lagrange function

$$\mathcal{L} = A + \lambda_g g + \lambda_h h \tag{2.22}$$

can be minimized to solve this constrained optimization problem. For the gradient of this function to be zero, all partial derivatives must be zero:

$$\left\{ \frac{\partial \mathcal{L}}{\partial N_i} \stackrel{!}{=} 0 \right\}, \quad \frac{\partial \mathcal{L}}{\partial \lambda_g} \stackrel{!}{=} 0\,, \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \lambda_h} \stackrel{!}{=} 0\,. \tag{2.23}$$

Partial derivatives with respect to the Lagrange multipliers simply return the constraints (2.20) and (2.21)

$$\frac{\partial \mathcal{L}}{\partial \lambda_g} = g \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \lambda_h} = h\,. \tag{2.24}$$

Differentiation with respect to the populations N_i, on the other hand, yields

$$\frac{\partial \mathcal{L}}{\partial N_i} = -kT(\ln q_i - \ln N_i) - \lambda_g\left(n_i^a + n_i^b\right) - \lambda_h m_i \stackrel{!}{=} 0\,, \tag{2.25}$$

which must hold for each N_i, and which can be rearranged to yield

$$-kT \ln\left(\frac{q_i}{N_i}\right) = \lambda_g n_i^a + \lambda_g n_i^b + \lambda_h m_i\,. \tag{2.26}$$

The derivatives with respect to the monomer populations are special, because in this case, n_i^a and n_i^b are known to be either one or zero. Thus, suitable substitutions can be made in eq. (2.26) and the expression can be rearranged to obtain:

$$-kT \ln\left(\frac{q_1}{N_1}\right) = \lambda_g + \lambda_h m_1 \quad \Rightarrow \quad \lambda_g = -kT \ln\left(\frac{q_1}{N_1}\right) - \lambda_h m_1\,, \tag{2.27}$$

$$-kT \ln\left(\frac{q_2}{N_2}\right) = \lambda_g + \lambda_h m_2 \quad \Rightarrow \quad \lambda_g = -kT \ln\left(\frac{q_2}{N_2}\right) - \lambda_h m_2\,. \tag{2.28}$$

In order to solve the system of eq. (2.26), we shall now substitute the

first λ_g in eq. (2.26) by (2.27) and the second one by (2.28):

$$\begin{aligned} -kT\ln\left(\frac{q_i}{N_i}\right) = &\left[-kT\ln\left(\frac{q_1}{N_1}\right) - \lambda_h m_1\right] n_i^a + \\ &\left[-kT\ln\left(\frac{q_2}{N_2}\right) - \lambda_h m_2\right] n_i^b + \lambda_h m_i \,. \end{aligned} \tag{2.29}$$

Since $m_i = n_i^a m_1 + n_i^b m_2$, the second Lagrange multiplier λ_h can be eliminated after some trivial rearrangements:

$$\begin{aligned} -kT\ln\left(\frac{q_i}{N_i}\right) &= -kT\ln\left(\frac{q_1}{N_1}\right)^{n_i^a} - kT\ln\left(\frac{q_2}{N_2}\right)^{n_i^b} - \\ &\quad \lambda_h\left(n_i^a m_1 + n_i^b m_2 - m_i\right) \\ &= -kT\ln\left(\frac{q_1}{N_1}\right)^{n_i^a} - kT\ln\left(\frac{q_2}{N_2}\right)^{n_i^b} . \end{aligned} \tag{2.30}$$

Finally, an expression for each cluster population N_i with respect to the monomer populations N_1 and N_2 can be obtained:

$$N_i = q_i\left(\frac{N_1}{q_1}\right)^{n_i^a}\left(\frac{N_2}{q_2}\right)^{n_i^b} . \tag{2.31}$$

Eq. (2.31) represents a system of polynomial equations with K unknown variables that can be solved exactly if the two variables N_1 and N_2 are fixed. Thus, two more independent equations are needed to obtain the final solutions. These are given by the constraints g and h.

Substitution of the populations in eq. (2.20) leads to the so-called population polynomial

$$0 = N^{\text{tot}} - \sum_{i=1}^{K}\left(n_i^a + n_i^b\right) q_i\left(\frac{N_1}{q_1}\right)^{n_i^a}\left(\frac{N_2}{q_2}\right)^{n_i^b} . \tag{2.32}$$

Similarly, substitution of the populations in eq. (2.21) leads to the so-

called mass polynomial

$$0 = M^{\text{tot}} - \sum_{i=1}^{K} m_i q_i \left(\frac{N_1}{q_1}\right)^{n_i^a} \left(\frac{N_2}{q_2}\right)^{n_i^b} . \tag{2.33}$$

The eqns. (2.32) and (2.33) represent a system of two non-linear equations with two unknowns. Such a system can be solved numerically, which will be discussed in Section 4.2.2. The solutions are the monomer populations N_1 and N_2. If these are known, the remaining populations are readily available by evaluation of eq. (2.31).

As indicated in Section 2.3, the preceding formalism can be easily extended to ternary mixtures and beyond, by introduction of further suitable constraints.

2.6 Determining the Volume

Since the volume is an independent variable in the canonical ensemble, in principle, any physically reasonable volume could be chosen. In practice, however, it is often more interesting to choose a volume that is in accord with an externally applied pressure

$$P = kT \frac{\partial \ln Q}{\partial V} . \tag{2.34}$$

This requirement leads to a polynomial, called volume polynomial, which is demonstrated in the following.

We shall start by inserting the system partition function (2.3) into eq. (2.34):

$$\begin{aligned} 0 &= -P + kT \frac{\partial \ln Q}{\partial V} \\ &= -P + kT \sum_{i=1}^{K} N_i \frac{\partial \ln q_i}{\partial V} . \end{aligned} \tag{2.35}$$

The cluster partition function q_i is a product of partition functions per degree of freedom (2.12). Out of those, only the translational and elec-

tronic partition functions depend explicitly on the volume. Their partial derivatives with respect to the volume are

$$\frac{\partial \ln q_i^{\text{trans}}}{\partial V} = \frac{1}{V - V^{\text{ex}}} \tag{2.36}$$

and

$$\frac{\partial \ln q_i^{\text{elec}}}{\partial V} = -a_{\text{mf}}\left(n_i^a + n_i^b\right)\frac{N^{\text{tot}}}{kTV^2}\,. \tag{2.37}$$

Insertion into eq. (2.35) and some elementary rearrangements lead to:

$$\begin{aligned} 0 &= -P + kT\sum_{i=1}^{K} N_i\left[\frac{1}{V - V^{\text{ex}}} - a_{\text{mf}}\left(n_i^a + n_i^b\right)\frac{N^{\text{tot}}}{kTV^2}\right] \\ &= -PV^2\left(V - V^{\text{ex}}\right) + kT\sum_{i=1}^{K} N_i\left[V^2 - a_{\text{mf}}\left(n_i^a + n_i^b\right)\frac{N^{\text{tot}}\left(V - V^{\text{ex}}\right)}{kT}\right] \\ &= -PV^3 + \left(PV^{\text{ex}} + kT\sum_{i=1}^{K} N_i\right)V^2 - a_{\text{mf}}\left(N^{\text{tot}}\right)^2 V + a_{\text{mf}}\left(N^{\text{tot}}\right)^2 V^{\text{ex}}\,, \end{aligned} \tag{2.38}$$

which is a simple cubic polynomial that can be solved analytically. Its solution will be discussed in Section 4.2.2.

2.7 Improvements to QCE Theory

2.7.1 The Binary Mean Field

This section's purpose is to present an improved description of the binary mean field energy, as well as to clarify and to relate the different expressions that can be found in the literature and within this thesis.

In Weinhold's seminal work,[17] the mean field energy of a neat system was introduced as

$$\varepsilon_i^{\text{mf}} = -\tilde{a}_{\text{mf}} n_i \frac{1}{V}\,, \tag{2.39}$$

which he described as an energy proportional to cluster size n_i and

density. This expression is essentially equal to the one of a van der Waals cluster introduced in Section 1.6, after realizing that the total number of particles N^{tot} is fixed in QCE theory and can be included in the mean field parameter. However, since inclusion of N^{tot} makes the parameter $\tilde{a}_{\mathrm{mf}}$ dependent on that number, a more desirable expression for the QCE mean field energy of a neat substance would be

$$\varepsilon_i^{\mathrm{mf}} = -a_{\mathrm{mf}} n_i \frac{N^{\mathrm{tot}}}{V} . \tag{2.40}$$

This expression is used throughout this thesis for neat substances. The relation between Weinhold's original formulation and the present one is given by

$$\tilde{a}_{\mathrm{mf}} = a_{\mathrm{mf}} N^{\mathrm{tot}} . \tag{2.41}$$

The mean field energy of a binary van der Waals gas was given in Section 1.7. Transferred to the nomenclature used in QCE theory, it would read

$$\varepsilon_i^{\mathrm{mf}} = -\frac{1}{V}\left(n_i^a N_1^{\mathrm{tot}} a_{\mathrm{mf}}^a + n_i^a N_2^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + n_i^b N_1^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + n_i^b N_2^{\mathrm{tot}} a_{\mathrm{mf}}^b \right) . \tag{2.42}$$

Therein, a_{mf}^a and a_{mf}^b correspond to the a_{mf} values of the neat substances and the mixed mean field parameter a_{mf}^{ab} quantifies the average interaction energy between unlike species.

In the original presentation of bQCE theory,[31] a simplified expression was introduced:

$$\varepsilon_i^{\mathrm{mf}} = -a_{\mathrm{mf}}\left(n_i^a + n_i^b \right) \frac{N_1^{\mathrm{tot}} + N_2^{\mathrm{tot}}}{V} . \tag{2.43}$$

It is easy to see that both expressions match if $a_{\mathrm{mf}}^a = a_{\mathrm{mf}}^b = a_{\mathrm{mf}}^{ab} = a_{\mathrm{mf}}$, that is, if average interactions within and between the components a and b are equal. Such a scenario can be assumed in first approximation for substances exhibiting similar types of interactions (for example, acetonitrile/acetone mixtures were found to be well-described by the conventional QCE mean field).[66] The approach may also work adequately for dissimilar substances, if most of the differences in interaction strength

are accounted for by the quantum chemical treatment of the cluster itself. In such a case, the conventional mean field should be fit to account for the residual interactions. For other mixtures of unlike species (for example, water/benzene or water/acetone), this approximation will most likely fail and eq. (2.42) should be used instead. Indications of such shortcomings have already been observed in a systematic study of small-cluster QCE calculations.[66]

The mixed mean field parameter a_{mf}^{ab} is an additional parameter in the equation of state of a binary QCE mixture that must be adjusted to experimental reference data. However, the amount of reference data required for a bQCE calculation actually reduces in most cases, if the improved binary mean field is used. In general, the pure mean field parameters a_{mf}^{a} and a_{mf}^{b} have already been obtained from separate parameter optimizations of the neat components. The mixed mean field parameter must be determined only once for any particular system composition, where experimental reference data is available. Once found, it is valid over the entire composition range. This is different from the previous formulation (eq. (2.40)), where a_{mf} had to be determined for each system composition subject to investigation. Since one is typically interested in obtaining mixture properties over a broad composition range, significantly more reference data had previously been necessary.

If desired, the mixed mean field parameter can also be obtained from mixing rules, such as

$$a_{\mathrm{mf}}^{ab} = \sqrt{a_{\mathrm{mf}}^{a} a_{\mathrm{mf}}^{b}}\left(1 - k^{ab}\right), \tag{2.44}$$

Therein, k^{ab} is called binary interaction parameter, which is typically employed as an empirical measure for the ideality of a mixture, with values close to zero corresponding to ideal mixing behavior. Note, however, that such interaction parameters and their interpretation cannot be transferred between different equations of states.[69] Binary QCE theory is capable of treating non-ideal mixing effects, even if k^{ab} happens to be zero for a particular mixture. That is because mixed-type interactions are modeled accurately on the intra-cluster level. Thus, in combination with the parameter-free description of the binary exclusion volume that

is presented in the next section, assuming $k^{ab} = 0$ could be a reasonable approximation in bQCE theory that would completely eliminate the need for hard-to-obtain binary reference data. This approximation will be referred to as geometric averaging of a_{mf} and its validity will be investigated in Section 5.2.

In conclusion, eq. (2.42) is physically sound, converts into the simplified expression if appropriate, and usually reduces the number of empirical model parameters. It has become the new default in Peacemaker and is used exclusively throughout this thesis.[a] Note that changes to $\varepsilon_i^{\mathrm{mf}}$ also require adjustments of the QCE polynomials. The modified equations are given in Appendix S1.

2.7.2 The Binary Exclusion Volume

After the considerations that led to the derivation of the binary van der Waals equation and those that led to the introduction of b_{xv}, one would have expected the following definition of V^{ex}:

$$V^{\mathrm{ex}} = b_{\mathrm{xv}}^{a} v_1 N_1^{\mathrm{tot}} + b_{\mathrm{xv}}^{b} v_2 N_2^{\mathrm{tot}} . \tag{2.45}$$

Therein, b_{xv}^{a} and b_{xv}^{b} are the excluded volume scaling parameters of the neat components a and b. For equal values $b_{\mathrm{xv}} = b_{\mathrm{xv}}^{a} = b_{\mathrm{xv}}^{b}$, this expression reduces to the one introduced in ref. [31]:

$$V^{\mathrm{ex}} = b_{\mathrm{xv}} V^{\mathrm{tot}} = b_{\mathrm{xv}} \left(v_1 N_1^{\mathrm{tot}} + v_2 N_2^{\mathrm{tot}} \right) . \tag{2.46}$$

However, such an equality cannot be expected in general, thus eq. (2.45) should be used instead. Much like the improved binary mean field, eq. (2.45) is theoretically sound and reduces the number of empirical parameters that enter into the QCE calculation. It has become the new default in Peacemaker and is used exclusively throughout this thesis. Since V^{ex} is a conserved quantity, changes to its definition do not affect the QCE equations.

[a] The simplified model (2.18) can still be enforced for comparisons with older calculations.

3 Computer Simulations

The theory of statistical thermodynamics presented in Chapter 1 is generally valid, but for many systems, it is impossible to find an analytical expression for the partition function. This is especially true if real systems, which cannot be reduced to simple models, are the subject of the investigation. Prototype systems which typically elude a complete theoretical description are complex liquids.[70] Even QCE theory introduced in Chapter 2, which is an attempt to describe real liquids while still adhering to a set of models that can be treated analytically, requires numerical methods to obtain solutions. Further complications arise if dynamic processes shall be investigated or if external perturbations are present. These are out of the scope of what has been presented so far. They may be incorporated into the statistical mechanical framework, but at some point there is always a system that is too complex to be described analytically. In such cases, computer simulations are typically performed. The following presentation computer simulation techniques is largely based on the work of Allen.[71]

3.1 Classification and Historical Background

Computer simulations are manifold but have one thing in common: They all need a model of the system that shall be described, that is, a representation of the system and a set of rules that describe its behavior. The model should be adequate for the question at hand, which means that a relativistic quantum chemical description of air flow through a jet engine is most certainly preposterous, whereas fluid dynamics, for example, is incapable of treating chemical reactions. An excellent classification and hierarchy of common computer simulation methods can be found in ref. [72]. In computational chemistry, molecular dynamics (MD) and Monte Carlo (MC) simulation are the most commonly employed methods. The idea

behind both of them is to generate a set of representative microstates in such a way that accurate values of structural and thermodynamic properties can be calculated.

Historically, MC simulations came first. In 1953, the first computer simulations of liquids were performed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller.[73] They devised a scheme to sample microstates from Boltzmann statistics using random numbers, thus, opening the path for so-called Monte Carlo simulations. Later, in 1957, Alder and Wainwright performed the first numerical integration of the equations of motion for a real system, that is, the first molecular dynamics simulation.[74,75]

The theory behind both types of simulations and the methods that have evolved around them are broad. Excellent textbooks that cover some aspects are refs. [18, 76]. Once more, it is not the purpose of this thesis to explain every imaginable aspect of both methods. Instead, the focus is set on some general aspects of MD simulation, as this technique is (besides QCE) one of the two main methods employed later.

3.2 The Ergodic Assumption

In the description of dynamic processes, two types of averages are of interest. The first is the statistical mechanical average of some quantity $A(t)$ at a given time t over all systems of the ensemble. This ensemble average, which has already been introduced in Section 1.3, shall be denoted by chevrons:

$$\langle A(t) \rangle = \frac{1}{N} \sum_{k=1}^{N} A^k(t) \,, \tag{3.1}$$

where $A^k(t)$ is the value of $A(t)$ in the k'th system of the ensemble and N is the very large total number of systems in the ensemble. The other average is that of $A^k(t)$ in a given system k of the ensemble over a very large time interval T (with $T \to \infty$). This time average shall be denoted by an overline:

$$\overline{A^k(t)} = \frac{1}{T} \int_0^T A^k(t+\tau)\, d\tau \,. \tag{3.2}$$

From a mathematical point of view, it is easy to see that both operations are commutative:

$$\begin{aligned}\left\langle \overline{A^k(t)} \right\rangle &= \frac{1}{N}\sum_{k=1}^{N}\left[\frac{1}{T}\int_0^T A^k(t+\tau)\,d\tau\right] \\ &= \frac{1}{T}\int_0^T\left[\frac{1}{N}\sum_{k=1}^{N} A^k(t+\tau)\right]d\tau = \overline{\left\langle A^k(t)\right\rangle}\,.\end{aligned} \tag{3.3}$$

Let us now consider a stationary situation, that is, one where there is no preferred origin of time for the statistical description of A. In such a case, each member of the ensemble $A^k(t)$ can be created by shifting the origin of time in any of the member systems by an arbitrary number. This is the case in any equilibrium system. In stationary ensembles, there is a close relation between ensemble and time averages, assuming that $A^k(t)$ takes on any of its accessible values in a sufficiently long time interval. This is called the ergodic assumption.

By definition of the stationary ensemble, the time average of A taken over some very long time interval T must be independent of the time t. Furthermore, the ergodic assumption demands that the time average of A must be the same in each system of the ensemble, that is,

$$\overline{A^k} = \overline{A} \quad \text{independent of } k\,. \tag{3.4}$$

Similarly, the ensemble average of A must be independent of time:

$$\langle A(t)\rangle = \langle A\rangle \quad \text{independent of } t\,. \tag{3.5}$$

Since both operations commute, the important relationship

$$\langle A\rangle = \overline{A} \tag{3.6}$$

follows. Or, if put into words, the ensemble average and time average in a stationary ergodic ensemble are equal. This equivalence lays the foundation for MD simulations, in which the dynamic trajectory of a single system is followed for a sufficiently long time.

3.3 Molecular Interactions

In MD simulations, the classical equations of motion are integrated numerically. For a simple atomic system, they read

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i \,, \tag{3.7}$$

where m_i, $\ddot{\mathbf{r}}_i$, and $\mathbf{F}_i$ are mass of, acceleration of, and force acting upon particle i, respectively.

The forces

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i} \tag{3.8}$$

are derived from a potential energy surface $U\left(\mathbf{r}^N\right) = U(\mathbf{r}_1, \mathbf{r}_2, \ldots)$. A typical form of this potential energy surface (PES) will be described in the following. The discussion is restricted to atomistic systems. Non-spherical or rigid units with rotational degrees of freedom require rotational equations of motion and interaction potentials.

If direct and accurate comparisons to experiments are sought, a realistic interaction model is essential. The aim of so-called ab initio (or better, first-principles) methods is to reduce the amount of fitting and empiricism in the construction of the PES to a minimum. On the other hand, often only general statements about certain phenomena or the assessment of different theories are aimed at. In such cases, it is not necessary to have a perfectly realistic model; one that captures the essential physics is adequate. In the following, different contributions to the PES will be explained.

3.3.1 Non-bonded Interactions

The part of the potential energy surface representing non-bonded interactions $U_{\text{non-bonded}}\left(\mathbf{r}^N\right)$ is typically split into 1-body, 2-body, 3-body, … terms:

$$U_{\text{non-bonded}}\left(\mathbf{r}^N\right) = \sum_i u(\mathbf{r}_i) + \sum_i \sum_{j>i} v\left(\mathbf{r}_i, \mathbf{r}_j\right) + \sum_i \sum_{j>i} \sum_{k>j} w\left(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k\right) + \ldots . \tag{3.9}$$

Therein, $u(\mathbf{r}_i)$ represents externally applied potential energy fields. This term is typically ignored in computer simulations of bulk liquids and so are the 3-body terms $w(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k)$ and higher ones. Thus, focus is set on the pair potential $v(\mathbf{r}_i, \mathbf{r}_j) = v(r_{ij})$. There is a wealth of literature available on the experimental determination and theoretical modeling of these interactions.[77–80]

The most commonly employed continuously differentiable pair potential is the Lennard-Jones potential

$$v^{\mathrm{LJ}}(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{3.10}$$

where the two parameters ε and σ represent potential well-depth and particle separation at close contact, respectively. This potential has a rich history, dating back to the earliest simulations on liquid Argon.[81]

If electrostatic charges are present, the Coulomb potential

$$v^{\mathrm{Coulomb}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{3.11}$$

must be added, with q_i and q_j being the charges of particles i and j and ϵ_0 the permittivity of free space. Coulomb forces are long-ranged and must be treated specially in periodic simulations. For a discussion of long-range solvers, the reader is referred to other sources.[18,76]

3.3.2 Bonding Potentials

Interactions between molecular systems are typically treated by site-site potentials of the type of eqns. (3.10) and (3.11). The necessary partial charges can be obtained from quantum chemical calculations of the electron density, which can be partitioned over all sites of the molecule.

Furthermore, intramolecular interactions must be considered. Typically, these include terms of the following kind:

$$U_{\mathrm{intramolecular}} = \frac{1}{2}\sum_{\mathrm{bonds}} k_{ij}^{r}\left(r_{ij} - r_{\mathrm{eq}}\right)^2 \tag{3.12}$$

$$+\frac{1}{2}\sum_{\text{angles}} k^{\theta}_{ijk}\left(\theta_{\text{ijk}}-\theta_{\text{eq}}\right)^2 \tag{3.13}$$

$$+\frac{1}{2}\sum_{\text{dihedrals}}\sum_{m} k^{\phi,m}_{ijkl}\left[1+\cos\left(m\phi_{ijkl}-\gamma_m\right)\right]. \tag{3.14}$$

In this context, bonds are interactions between neighboring atoms within the same molecule, described by their separation $r_{ij} = \left|\mathbf{r}_i - \mathbf{r}_j\right|$. The harmonic approximation of the bonding potential (3.12), with force constant k^r_{ij} and equilibrium bond length r_{eq}, is the most common form. Angle potentials exist between neighboring bond vectors, in this case $\mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{r}_j - \mathbf{r}_k$. They are characterized by the angle

$$\cos\theta_{\text{ijk}} = \frac{\mathbf{r}_{ij}\cdot\mathbf{r}_{jk}}{r_{ij}r_{jk}}, \tag{3.15}$$

which typically enters into a quadratic bending term, such as (3.13), where k^{θ}_{ijk} is a force constant and θ_{eq} the equilibrium angle. Finally, three connected bonds form a dihedral angle

$$\cos\phi_{ijkl} = -\mathbf{n}_{ijk}\cdot\mathbf{n}_{jkl}, \tag{3.16}$$

where $\mathbf{n}_{ijk} = \mathbf{r}_{ij}\times\mathbf{r}_{jk}$ and $\mathbf{n}_{jkl} = \mathbf{r}_{jk}\times\mathbf{r}_{kl}$. The dihedral potential energy function is typically expanded in periodic functions of order $m = 1, 2, 3, \ldots$, phase γ_m, and associated force constant $k^{\phi,m}_{ijkl}$, such as (3.14).

The complete set of equations and parameters describing the PES is called a force field. Eqns. (3.10) to (3.14) are an example of a simple force field that can be encountered in MD simulations, though certainly not one of the best. Accurate force fields often include different functional forms and cross terms. Force field parameters are typically determined by quantum chemical calculations and comparison of simulation results with physicochemical data. Common force fields employed in computational chemistry that can be applied to a larger class of substances are the molecular mechanics (MM) family of force fields (MM3, MM4),[82,83] the generalized Amber force field (GAFF),[84] or the optimized potentials for liquid simulations (OPLS).[85,86]

3.4 The MD Algorithm

If we continue to discuss an atomic system (for simplicity's sake), with position coordinates $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$, potential energy surface $U(\mathbf{r}^N)$, atomic momenta $\mathbf{p}^N = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N)$, and kinetic energy

$$K(\mathbf{p}^N) = \sum_{i=1}^{N} \frac{|\mathbf{p}_i|^2}{2m_i}\,, \tag{3.17}$$

the Hamiltonian of the system becomes the sum of potential and kinetic energy $H = U + K$ and the classical equations of motion read

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} \qquad \text{and} \qquad \dot{\mathbf{p}}_i = \mathbf{F}_i\,. \tag{3.18}$$

This is a system of coupled differential equations that can be solved numerically with step-by-step integration.

The requirements on the integrator are manifold. First, it must be able to cope with both short and long time scales. Second, it should keep the number of force evaluations at a minimum, since those involve an expensive double sum over pairs of atoms. Third, it should accurately follow the constant energy surface for times that are much larger than typical correlation times τ_A of the dynamic quantities A under investigation. This is to ensure the sampling of the correct ensemble. Finally, the time step should be as large as possible in order to rapidly sample the phase space. Practical aspects, such as simplicity of implementation may also be imposed.

3.4.1 The Verlet Algorithm

All requirements given above favor the Störmer–Verlet family of integration algorithms,[87] which over the years have become the standard in MD simulation.[76] Other algorithms are typically implemented only for historical reasons or to fill niche applications. The most common variant of the family, the velocity Verlet algorithm, shall be described in the following.

The velocity Verlet algorithm may be written as

$$\mathbf{p}_i(t + \tfrac{1}{2}\,\delta t) = \mathbf{p}_i(t) + \tfrac{1}{2}\,\delta t\,\mathbf{F}_i(t) \tag{3.19}$$

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \delta t\,\mathbf{p}_i(t + \tfrac{1}{2}\,\delta t)/m_i \tag{3.20}$$

$$\mathbf{p}_i(t + \delta t) = \mathbf{p}_i(t + \tfrac{1}{2}\,\delta t) + \tfrac{1}{2}\,\delta t\,\mathbf{F}_i(t + \delta t)\,, \tag{3.21}$$

where δt denotes the time step. A single force calculation per integration step is performed between eqns. (3.20) and (3.21), making this algorithm very cheap. Furthermore, it is low-order, permitting large time steps, exactly time reversible, and symplectic, that is, it conserves the phase space volume. For a deeper discussion of such subtleties, the reader is referred to textbooks.[18,76]

3.4.2 Periodic Boundary Conditions

Consider a system of 1000 particles arranged on a $10 \times 10 \times 10$ cubic lattice. Unless surface effects are of explicit interest, periodic boundary conditions must be applied, or 488 atoms have surface contact. Such conditions can be realized by surrounding the simulation cell with replicas of itself. To avoid artificial interactions between periodic images, the minimum image convention must be applied, demanding that each atom interacts only once with the closest neighbor or that neighbor's periodic image. If an atom leaves the simulation cell during the course of the simulation, a replica enters from the other side of the cell. If the potentials employed during the simulation are short ranged and if no properties with large correlation lengths are studied, the influence of this artificial periodicity can be safely neglected. Otherwise, the system size must be increased so that it is larger than the correlation length of the property of interest. As mentioned earlier, long ranged potentials, such as charge–charge interactions, must be treated with special care in periodic simulations.

3.5 MD Simulation in Different Ensembles

Ensembles have already been introduced in Chapter 1. Molecular dynamics simulations sample from the microcanonical (NVE) ensemble,

which is said to be their natural ensemble. For comparisons with experiments or if quantities that require fluctuations of the particle number, volume, or total energy shall be calculated, simulations in different ensembles are sometimes desirable. In this section, constant temperature simulations in the canonical (NVT) ensemble shall be shortly explained. Analogous approaches exist for other ensembles, in particular in the isobaric-isothermal (NPT) or in the isoenthalpic (NPH) ensemble.

One of the simplest methods to implement constant temperature in MD simulations is to randomly reassign atomic velocities with values sampled from the Maxwell–Boltzmann distribution of velocities at the targeted temperature.[88] Either individual atoms or the entire system may be chosen in this process. For guidance on the sampling frequency and other matters, see ref. [88]. This procedure is called Andersen thermostat.

In another approach, the Nosé–Hoover thermostat,[89,90] an additional degree of freedom (corresponding to a thermal reservoir) is added to the dynamic equations:

$$\dot{\mathbf{r}}_i = \mathbf{p}_i / m_i \tag{3.22}$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \zeta \mathbf{p}_i \tag{3.23}$$

$$\dot{\zeta} = \nu_T \left(\frac{T^{\text{inst}}}{T} - 1 \right). \tag{3.24}$$

Here, ζ is a friction coefficient that varies in time, ν_T the thermal rate of relaxation, T^{inst} the instantaneous 'mechanical' temperature, and T the thermodynamic temperature of the system. Thus, if if the system is temporarily too hot ($T^{\text{inst}} > T$), the friction coefficient ζ increases, and vice versa, if the system is too cold ($T^{\text{inst}} < T$). A simple Nosé–Hoover thermostat may lead to non-ergodic behavior, which can be alleviated by chaining multiple thermostats together.[91] It can be shown that a Nosé–Hoover thermostat samples from a canonical distribution function. Furthermore, it conveniently yields a conserved quantity (the Nosé–Hoover Hamiltonian) that can be used to verify the accuracy of the time integration.

Thermostating is a popular approach in MD simulations for the reasons given in the beginning of this section. However, it should be stressed

that any thermostat artificially modifies the true equations of motion, which may falsify dynamic properties. The Andersen thermostat is an extreme example, which randomly breaks dynamic correlations. Other thermostats are more subtle, but should still be carefully tested before dynamic properties are evaluated. On the other hand, there is no reason to refrain from their use if only structural or thermodynamic properties are of interest, which explains their wide spread. For the reasons given above, no thermostats were used in the MD simulations presented in this work, if accurate dynamic quantities were sought.

Results and Discussion

4 The Peacemaker QCE Code

The original Quantum Cluster Equilibrium code "Peacemaker" was first developed by Barbara Kirchner in the early 2000s.[37] Its development continued until 2011, with contributions from the following authors (in alphabetical order): Marc Brüssel, Sebastian B. C. Lehmann, Eva Perlt, Christian Spickermann, and Michael von Domaros. At that time, a detailed overview of Peacemaker's development status was given: The code was written in Fortran 95, it was capable of treating neat substances, and it allowed the calculation of numerous thermodynamic quantities (Helmholtz energy A, Gibbs energy G, internal energy U, enthalpy H, entropy S, and heat capacities at constant volume and pressure C_V and C_P). The code also featured a so-called sampling procedure, which was designed to choose optimal a_{mf} and b_{xv} parameters, and a search routine, whose purpose was to select low-populated clusters and drop them from the cluster set if this led to an improved reproduction of experimental reference data.

In 2011, QCE theory was extended to binary systems (bQCE).[31] Unfortunately, the Peacemaker code was never designed to be expanded to more than one substance. For this reason and because Peacemaker's historic development led to a rather convoluted structure, these changes could not be easily implemented in the existing program. Instead, a new program called 'Mixtures in Cluster Equilibrium' (Mice)[31] was written to give the proof of concept for bQCE. The first goal of this PhD work was to unify both programs, in order to prevent a splitting of the user base, to address some methodological issues (such as the convergence to metastable phases), to accelerate the code, and to implement the theoretical improvements introduced in Section 2.7. The new program was written completely from scratch, but named Peacemaker 2 to honor its historical origins.

All features of the original code but the search procedure were repro-

duced and improved if appropriate. The search routine was dropped intentionally, as there is no theoretical justification to discard clusters, even if this leads to an apparently better description of experimental properties. Clusters which are 'unimportant' should not be populated and thus do not contribute to thermodynamic properties.[a] The new Peacemaker code is numerically robust enough to handle such low-populated clusters. Furthermore, clusters which are unphysical should not have been included in the first place. Careful analysis of the physics of a system is a common task in any computational model and cannot be automated by a software.

The new code was parallelized and significantly sped up by sophisticated design. It is properly documented, both in form of a user manual and within the source code. The software is published online[92] and licensed under the GNU General Public License in its third version (GPLv3),[93] since it is the author's firm believe that scientific software, which was used to generate published data, should be freely available in source form to ensure full documentation and reproducibility of all calculations carried out.

In the following chapter, the Peacemaker code is presented in full detail. This includes a discussion of the programming language, the structure of Peacemaker, its algorithms, and the parallelization strategy. The chapter is written both from a developer's point of view (to serve as a reference for future developers and to document the work performed for this thesis) and from a user's perspective (serving as usage instructions). Note that for the latter case, a separate manual was written, as well, which is available online.[92] Certain parts of this chapter may be quite technical. Unfortunately, this is unavoidable because of the nature of the content. An attempt was made to keep these parts at a minimum.

[a] Note that low-populated clusters may still contribute to the physics of the system. This is the case, for example, when using QCE to calculate the ionic product of water,[23] which relies on summing up the populations of very low-populated dissociated clusters. This project is described in Chapter 5 and is another reason, why the search routine was not reimplemented.

4.1 Choice of Programming Language and Style

When writing a new code, several decisions about the design of the program must be made. The most obvious and most fundamental one is the choice of the programming language, which shall be discussed in this section.

Peacemaker is a scientific software, involving a considerable amount of computationally demanding tasks. These have traditionally been tackled with the Fortran programming language, because of its strict set of rules, which permit easy optimizations by compilers. As a result, high quality compilers and a tremendous code base have evolved around Fortran. Unfortunately, the inclusion of modern programming techniques has been slow in Fortran and over the past decades, the C family of languages has gained increasing popularity in the scientific community, thanks to improving compiler quality, vast availability of libraries, and powerful programming constructs and paradigms inherent to these languages. Finally, interpreted languages such as Python have started to enjoy a considerable user base, as well. The ease of programming that comes with these languages makes them very attractive to scientists, for whom the time spent developing a program is often much larger than the time gained from executing a program written in a more optimized, compiled language.

Apart from the pros and cons of each programming language, which shall not be explicitly listed here, several requirements were imposed on the language, the code, and its usage. These are:

1. The code should be easy to read and understand, permitting future modifications by other developers.

2. The code should also be optimized and run fast, but not at the cost of sacrificing the ease of reading.

3. The language should be easy to learn for future PhD students and understandable to seasoned researchers wishing to read or modify the code.

4. The code should be parallelized and the language should promote a programming style that allows easy parallelization.

Of course, none of these requirements exclude any language per se, and any choice would be suitable, but the author believes that they are best covered by the Fortran language, which still enjoys large popularity, is easy to learn, and has alleviated many of the weaknesses that led to its decline in popularity. Unlike in interpreted languages, code written in Fortran is usually fast (unless poorly designed), and the strict set of rules (which may seem as a burden to programmers of other languages) helps avoiding common types of errors and permits easy parallelization using OpenMP. Thus, Peacemaker was eventually rewritten in modern Fortran, which at the time of writing was Fortran 2008. No restrictions were imposed on the Fortran standard, except for the implementation status in modern compilers.

Since new PhD students entering the field of theoretical chemistry often do not have a programming background, advanced programming paradigms were avoided in the core part of Peacemaker. This includes the calculation of the cluster and system partition functions, the solution of the QCE polynomials, and the evaluation of thermodynamic quantities. The routines performing these operations and the data structures that they work on were deliberately kept simple. Thus, future generations of PhD students working on the code should be able to modify or extend them easily, without having to learn about advanced programming concepts, such as object-oriented programming, first. Other parts, which are more technical in nature and less likely to be modified, such as the input parser, were not affected by this decision and written in modern, object-oriented Fortran, which facilitates complex tasks and ultimately benefits both users and developers. No external libraries were used and Peacemaker was written to be platform-independent, but has not been tested on other systems than GNU/Linux.

4.2 Program Structure and Algorithms

In this section, the program flow of Peacemaker is discussed. Peacemaker is a non-interactive program that essentially has a simple linear structure (read input, perform calculations, and report results). Details on each part are given in the following. In the process, usage instructions are

Listing 4.1: General structure of Peacemaker input and cluster set files.

```
[section1]
    keyword1 argument1 argument2 ... # comment
    keyword2 argument1 argument2 ...
    ...
[section2]
    keyword1 argument1 argument2 ...
    keyword2 argument1 argument2 ...
    ...
...
```

given, design decisions are explained, and algorithms are presented.

4.2.1 Program and Cluster Set Input

Peacemaker reads two files, the input and the cluster set file. Both share the same general structure, which is illustrated in Listing 4.1. Each file holds an arbitrary number of sections which are embraced in brackets and each section contains an arbitrary number of keywords after whom various arguments can be appended. Section labels must be unique and so must be keywords within a section. Unlike in most other Fortran codes, there are no limitations on the width of labels or line lengths. If desired, comments can be introduced after the number sign (#). This powerful and flexible input structure was implemented and hidden behind abstract objects. Simple inquiry functions exist to return a list of all sections, keywords, and arguments, or to query the presence of a certain entry. This should allow easy extension of the input files by future developers.

As an example, a typical input file for a QCE calculation of a neat substance is given in Listing 4.2. Care was taken to make the input file as self-explanatory as possible. In the example, a sampling calculation is performed, that is, the pair of parameters $(a_{\mathrm{mf}}, b_{\mathrm{xv}})$ that best matches experimental reference quantities is determined. The mean field parameter a_{mf} is distributed evenly on a range from 0.0–0.5 J m^3 mol^{-2} with 101 data points, including the boundaries. An analogous range specification

Listing 4.2: Typical input file for a QCE calculation of a neat substance.

```
[qce]
    amf 0.0 0.5 101 # Jm^3/mol^2
    bxv 0.8 1.3 101
[ensemble]
    temperature 200.0 400.0 201 # K
    pressure 1.01325 # bar
[reference]
    density 298.15 1.023 # K; g/cm^3
    phase_transition 373.15 # K
```

was used for b_{xv} and the temperature. As requested by the presence of the reference section, the temperature of phase transition and density at 298.15 K are calculated for each a_{mf} and b_{xv} pair and compared to the experimental reference data. These are the density at 298.15 K and the boiling point specified to be at 373.15 K. The exact nature of this sampling procedure will be explained shortly.

In the case of the cluster set file, each section corresponds to one cluster. The section label is used to designate the cluster and keywords give the composition n_i, paths to coordinate and frequency files, rotational symmetry numbers σ_i, adiabatic binding energies $\Delta\varepsilon_i^{\mathrm{bind}}$, and further, optional parameters. An example for a neat substance is given in Listing 4.3. Coordinate files must be in the standard xyz file format and frequencies must be listed one by one in a separate text file (as wavenumbers, in units of cm^{-1}). Thus, unlike its predecessor, Peacemaker 2 is not coupled to any quantum chemistry code in particular. The specification of all frequencies in a simple text file permits easy use of any frequency scaling function (e.g., a single scaling factor, or a wavenumber dependent scaling function) before Peacemaker is invoked. Such functions were not implemented on purpose, since the number of possible choices is very large, but the code was meant to be as general as possible. Finally, the reference monomers must be indicated by the monomer keyword and their volume must be specified as well.

Listing 4.3: Extract of a typical cluster set file for a QCE calculation of a neat substance.

```
[Cluster 1]
    monomer
    composition 1
    sigma 2
    coordinates /path/to/cluster1.xyz
    frequencies /path/to/cluster1.flist
    energy 0.0 # kJ/mol
    volume 60.0 # Angstrom^3
[Cluster 2]
    composition 3
    sigma 3
    coordinates /path/to/cluster2.xyz
    frequencies /path/to/cluster2.flist
    energy -30.0 # kJ/mol
...
```

Cluster volume calculations are not performed by Peacemaker itself. Instead, external programs such as Gepol93[94] have to be used. Simple-to-use Python programs is distributed with Peacemaker to facilitate the interaction with common external tools. These include programs that convert output from Turbomole[95,96] or Orca[97,98] to the respective Peacemaker files, a program that calculates molecular volumes by summing up atomic van der Waals volumes (which was the default in older versions of Peacemaker and is thus needed for reproducibility), and a program that generates input for Gepol93 in a modern, user-friendly fashion.

A complete list of required and optional keywords, both for the cluster set file and for the input file, can be found in Appendix S2.

4.2.2 Solution of the QCE Polynomials

The core of Peacemaker deals with the solution of the volume polynomial (2.38), as well as with the solution of the population and mass poly-

nomials (2.32) and (2.33). In this section, the algorithms implemented to solve the QCE polynomials shall be explained in more detail. They were mostly inspired by ref. [99] or implemented according to common knowledge.

The volume polynomial is a cubic equation with all-real coefficients a_i:

$$0 = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \,. \tag{4.1}$$

Such a polynomial has at least one real root and is easy to solve. In Peacemaker, the polynomial equation is solved in complex space. For this purpose, the following coefficients are calculated first:

$$A_1 = a_2^2 - 3a_3 a_1 \tag{4.2}$$

$$A_2 = 2a_2^3 - 9a_3 a_2 a_1 + 27a_3^2 a_0 \tag{4.3}$$

$$A_3 = \sqrt[3]{\frac{A_2 + \sqrt{A_2^2 - 4A_1^3}}{2}} \,. \tag{4.4}$$

With their aid, all three roots can be obtained as follows:

$$x_k = -\frac{1}{3a_3}\left(a_2 + \sigma^k A_3 + \frac{A_1}{\sigma^k A_3}\right), \tag{4.5}$$

with $k = 1, 2, 3$ and $\sigma = -\frac{1}{2} + \frac{\sqrt{3}}{2} i$. These equations return three complex numbers, with either one, or all of them being real (except for numerical noise). If there are multiple solutions to the volume polynomial, either the largest or the smallest volume is chosen. The exact choice depends on the mode of the temperature loop, which is explained in Section 4.2.4.

The population and mass polynomials are more advanced. In the binary case, they span a set of two non-linear, polynomial equations with two unknowns (the monomer populations N_1 and N_2). For a neat substance, the solution of one non-linear equation with a single unknown variable (N_1) is sufficient. In this section, the solution of the two-dimensional set of polynomial equations are discussed. The one-dimensional case follows from reduction of the more general two-dimensional case.

In order to solve the system of equations, the Newton–Raphson algo-

rithm was implemented in Peacemaker. The set of equations that must be solved can be written as

$$\begin{aligned} P_1(x_1, x_2) &= 0\,, \\ P_2(x_1, x_2) &= 0\,, \end{aligned} \tag{4.6}$$

or in vector form:

$$\mathbf{P}(\mathbf{x}) = \mathbf{0}\,. \tag{4.7}$$

Each of the functions P_i can be expanded in Taylor series:

$$\mathbf{P}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{P}(\mathbf{x}) + \mathbf{J} \cdot \delta\mathbf{x} + \mathcal{O}\left(\delta\mathbf{x}^2\right), \tag{4.8}$$

wherein $\mathbf{J}$ denotes the Jacobian matrix of partial derivatives

$$J_{ij} = \frac{\partial P_i}{\partial x_j}\,. \tag{4.9}$$

By neglecting terms of order $\delta\mathbf{x}^2$ and higher and by setting $\mathbf{P}(\mathbf{x} + \delta\mathbf{x}) = 0$, a set of linear equations is obtained that can be solved analytically for the correction $\delta\mathbf{x}$ that moves all function values closer to zero. Thus, an iterative procedure can be constructed that systematically refines an initial guess $\mathbf{x}_0$ by means of the following recurrence relation

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \delta\mathbf{x}\,. \tag{4.10}$$

where $i = 1, 2, 3, \ldots$ denotes the number of the current iteration.

In Peacemaker, the initial guess for the populations N_1 and N_2 is set to N_1^{tot} and N_2^{tot}, respectively. This means that initially, only monomers are populated. If a Newton–Raphson step $\delta\mathbf{x}$ would move either of the populations outside of the physical meaningful range of values (i.e., below zero or above $N_{1/2}^{\mathrm{tot}}$), the step's magnitude is halved until the condition is no longer violated.

The Newton–Raphson method requires the frequent evaluation of polynomials and their derivatives. These operations are the numerical bottleneck of Peacemaker and their optimization is thus of utmost im-

portance. A naive evaluation of a polynomial $P(x)$, according to its definition

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_N x^N \tag{4.11}$$

is an extremely expensive operation, since each power of x must be evaluated separately. At this point, a quotation from ref. [99] seems appropriate: "Come the (computer) revolution, all persons found guilty of such criminal behavior will be summarily executed, and their programs won't be!" A much more efficient strategy, called Horner's method, transforms the polynomial into a computationally efficient form first:

$$P(x) = a_0 + x\big(a_1 + x\big(a_2 + \cdots + x\big(a_{N-1} + x a_N\big)\big)\big)\,, \tag{4.12}$$

which can be evaluated iteratively in computer codes by starting from the innermost parenthesis and moving outwards. A similar scheme can be devised for the derivative of a polynomial, and both methods can be efficiently combined into a single algorithm. Horner's method can also be expanded to two dimensions, albeit losing some of its efficiency. Nevertheless, the speed-up gained in comparison to a naive evaluation of the polynomials and their derivatives is tremendous, thus the method is being consequently used in Peacemaker.

4.2.3 The QCE Iteration

Given a temperature T and an initial volume guess V_0, as well as all other required parameters (e.g., pressure, total particle number, total mass, or mean field and excluded volume parameters), Peacemaker solves the QCE equations in an iterative procedure. The need for that is easy to see, after considering the QCE equations once more. To obtain a volume that is consistent with the specified external pressure, the volume polynomial (2.38) must be solved. This polynomial depends on knowledge of the populations, that is, $V = V(\{N_i\}, T)$. Determination of the populations, however, requires the solution of the population and mass polynomials (2.32) and (2.33), which are functions of the single-particle partition functions q_i (eq. (2.12)), that is, $\{N_i = N_i(q_1, q_2, q_i)\}$. These partition functions are easy to evaluate, but depend on the volume, because of the translational partition function (1.34) and the mean field energy (2.42).

However, the volume is the quantity that was to be determined in the first place.

For these reasons, the following iterative cycle was implemented into Peacemaker, which is depicted schematically in Figure 4.1. Given an initial volume guess V_0, the partition functions, populations, and the volume are subsequently calculated. This cycle is repeated until populations and volume do not change anymore, that is, until self-consistence is reached. A quantity, that is ideal to check for this purpose is the Gibbs energy $G = -kT \ln Q + PV$, as it depends both on populations (by means of the system partition function) and volume. Thus, the relative deviation of G between two successive cycles $|\Delta G|/G$ is calculated and compared against a maximally allowed threshold value ε_G, which by default is set to $\varepsilon_G = 1.0 \times 10^{-9}$.

The initial volume guess is a critical quantity in the QCE iteration. For thermodynamic states that permit metastable solutions, the initial volume guess may decide whether the stable or metastable solution is chosen. If a gas-like, low-density volume is given, the QCE iteration favorably converges to the gas phase solution, whereas a liquid-like, high-density guess is more likely to converge to the liquid phase solution. Note that these are observations based on experience. An exact numerical analysis of the convergence behavior would be very tedious, if possible at all, and reach well beyond the scope of this thesis.

Usually, the stable solution is sought (i.e., the one with lower Gibbs energy G), thus great care must be taken in choosing V_0. An automated procedure that yields the stable solution is presented in the next section. It essentially involves running the complete QCE iteration twice, once with a gas-like initial volume guess and once with a liquid-like one, and choosing the solution that leads to the smaller Gibbs energy.

4.2.4 The Temperature Loops

Peacemaker loops twice over all specified temperature values, once from the highest temperature $T_{\max}$ to the lowest one $T_{\min}$, and once vice versa. These loops are referred to as gas phase loop and liquid phase loop, respectively. Running both is a necessary procedure to prevent metastable solutions. The reasoning behind this approach shall now be explained.

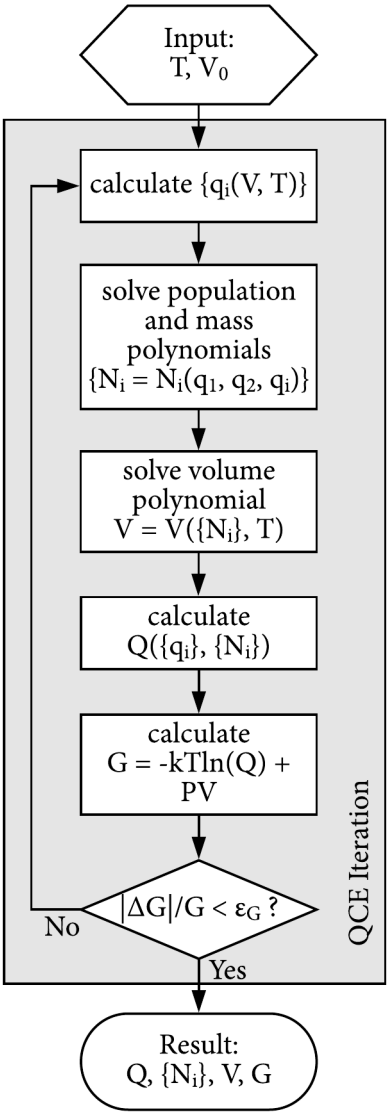


Figure 4.1: Flow diagram of the QCE iteration. Given a temperature T and an initial volume guess V_0, Peacemaker proceeds by iteratively solving the QCE equations. Volume V and populations $\{N_i\}$ are systematically refined, until the Gibbs energy G does not change anymore. Apart from volume and temperature, a variety of other parameters (such as pressure, total particle number, or mean field and excluded volume parameters) also enter into the equations, but are not listed here for clarity's sake.

The exact procedure is illustrated in Figure 4.2.

Traditionally, QCE calculations were run from the highest temperature $T_{\max}$ to the lowest one $T_{\min}$. The initial volume guess for $T_{\max}$ was set to the ideal gas volume $V_0(T_{\max}) = V^{\mathrm{id}}(T_{\max}) = N^{\mathrm{tot}} k T_{\max}/P$, as this volume provides a natural upper bound to the range of possible volumes. Initial volumes for subsequent temperatures were set to the converged volume of the previous QCE iteration.

Unfortunately, the former procedure tends to favor the largest volume in temperature regions, where multiple solutions coexist, even if this corresponds to a metastable state. Because of this tendency to yield the gas phase solution, this loop is called gas phase loop. However, given multiple volumes to choose from, the one that minimizes the Gibbs energy G for a fixed pressure and temperature should be taken.

The problem can be illustrated with the aid of Figure 4.3. Without further action, the gas phase loop leads to the blue line (corresponding to the gas phase solution), even if there is a more stable liquid phase at temperatures below the boiling point. Once this metastable solution cannot be maintained anymore, the Gibbs energy drops artificially (indicated by the arrows in Figure 4.3), and calculations continue in the liquid phase.

Tests show that the tendency of the QCE iteration to favor metastable solutions cannot be suppressed by reversing the temperature loop to run from $T_{\min}$ to $T_{\max}$. In this case, metastable liquid solutions would emerge. Such a loop is called liquid phase loop. Therein, the initial volume at the lowest temperature is set to a fraction of the ideal gas volume $V_0(T_{\min}) = \lambda V^{\mathrm{id}}(T_{\min})$, with λ being set to 0.01 by default. Initial volumes for subsequent temperatures are set to the converged solution of the previous QCE iteration.

In complete analogy to the gas phase loop, liquid phase solutions emerge for as long as they can be maintained (Figure 4.3).

The obvious solution to prevent metastable phases is to run each QCE iteration twice: once with a gas-like initial volume and once with a liquid-like one. In Peacemaker, this is realized by running two temperature loops as described above. The loops are run one after another, and both solutions are stored for each temperature. At the end, the Gibbs energies of all solutions are compared, and the lowest ones are chosen for each temperature. The metastable solutions are discarded.

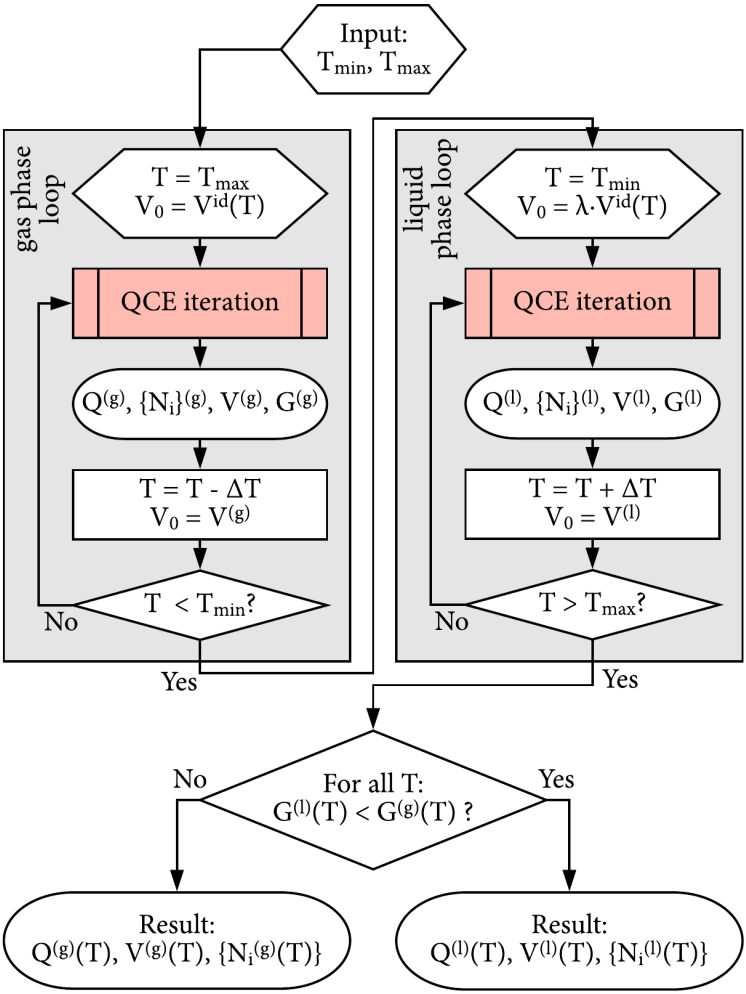


Figure 4.2: Flow diagram of the QCE temperature loops. Peacemaker loops twice over each specified temperature in order to avoid metastable solutions.

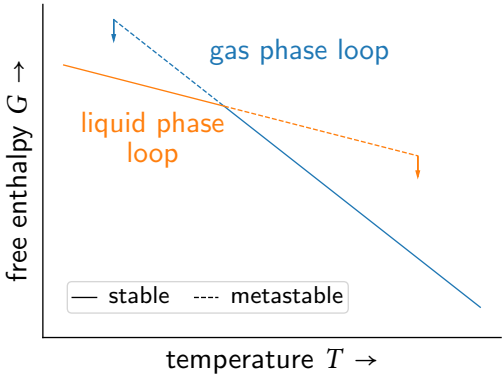


Figure 4.3: Solutions of the gas phase loop and of the liquid phase loop in a temperature region where both coexist. Solid lines correspond to the thermodynamically stable state, dashed lines correspond to metastable ones. After leaving the coexistence range, solutions jump as indicated by the arrows.

Although the proposed procedure effectively doubles the computational efforts of a QCE calculation, the slow-down was minimized by making use of the direction of the temperature loops (from high to low and vice versa). In its current implementation, previously converged volume solutions are used as initial guesses in subsequent QCE iterations and typically lead to rapid convergence.

Finally, note that the mode of the temperature loop also determines the choice of the root of the volume polynomial, if there are multiple solutions (see Section 4.2.2). Within the gas-phase loop, the largest root is chosen and within the liquid-phase loop, the smallest one.

4.2.5 Parameter Sampling

Peacemaker can perform the preceding calculations for a single pair of $(a_{\mathrm{mf}}, b_{\mathrm{xv}})$ values, or over a specified range of these parameters.[a] The latter is called parameter sampling and its purpose is to determine the a_{mf} and b_{xv} values that lead to the best agreement with experimental reference data.

[a] Note that in this section, a_{mf} shall refer to either the mixed mean field parameter a_{mf}^{ab} (for binary systems) or to the neat mean field parameter a_{mf} (in QCE calculations of neat substances). Furthermore, note that for binary systems, sampling of b_{xv} is not necessary anymore (see Section 2.7), but possible for comparisons with older versions of the theory.

Therefore, two nested loops are performed in Peacemaker: one over a_{mf} and one over b_{xv}. After running calculations for each pair of parameters, results are compared to experimental quantities and the pair of parameters that best matches is chosen. All other results are discarded. The following experimental properties can be specified as reference in the current version of Peacemaker: a single density ρ^{ref} at a specified temperature, an isobar $\left\{\left(T_i^{\mathrm{ref}}, V_i^{\mathrm{ref}}\right)\right\}$, and the temperature of phase transition $T_{\mathrm{pt}}^{\mathrm{ref}}$. In principle, any other thermodynamic property could be taken, as well, and would be easy to implement, if desired. These reference values are compared to their calculated counterparts by means of the following error function

$$E = W_\rho \left(\frac{\rho^{\mathrm{calc}} - \rho^{\mathrm{ref}}}{\rho^{\mathrm{exp}}} \right)^2 + W_{\mathrm{pt}} \left(\frac{T_{\mathrm{pt}}^{\mathrm{calc}} - T_{\mathrm{pt}}^{\mathrm{ref}}}{T_{\mathrm{pt}}^{\mathrm{exp}}} \right)^2 + \frac{W_V}{N} \sum_{i=1}^{N} \left(\frac{V_i^{\mathrm{calc}} - V_i^{\mathrm{ref}}}{V_i^{\mathrm{exp}}} \right)^2 . \tag{4.13}$$

Therein, the weights W_ρ, W_{pt}, and W_{V} can be used to specify the relative importance of the different reference values. By default, equal weights are assigned to all specified quantities.

At this point, a few notes on phase transitions and the temperature at which they occur shall be given. If the temperature range is chosen accordingly, QCE isobars show first order phase transitions. Peacemaker can determine these temperatures by looking for the largest volume jump between two converged solutions. Note that QCE calculations yield a sharp temperature of phase transition, even in binary mixtures. That is because QCE theory is derived in the canonical ensemble, where mixtures cannot stabilize by partially evaporating (with associated change in composition). When specifying a reference boiling point for a binary mixture, the average between bubble point and dew point of the mixture should be chosen.

4.3 Parallelization Strategy

A single QCE iteration for a given temperature is fast on a standard workstation ($\ll 1\,\mathrm{sec}$) and would not require any parallelization. However, Peacemaker performs multiple of these iterations, since it typically cal-

culates a range of temperatures and covers multiple $(a_{\mathrm{mf}}, b_{\mathrm{xv}})$ values as part of its sampling procedure. Computational time can thus accumulate easily, which necessitates parallelization. Consider, for example, a sampling calculation with 100 a_{mf} and 100 b_{xv} data points over a temperature interval spanning 250 K, incremented by 1 K steps. If we assume that, on average, one QCE iteration takes 0.02 s, then the total calculation time would sum up to $100 \times 100 \times (2 \times 250) \times 0.02\,\mathrm{s} = 27.8\,\mathrm{h}$. Depending on one's perception of a 'long' calculation, this is probably enough to warrant parallelization on a typical workstation or single compute node (say 8–32 CPUs), but not long enough to warrant more complicated parallelization across multiple compute nodes or workstations.

There are two common parallelization mechanisms, by which parallel processes can communicate with each other. In shared-memory parallelization, processes share a single global address space that they can read from and write to. This binds them to the physical memory location that they work on. The most common alternative is message passing, where processes exchange data by sending messages to each other. These messages can be sent over a network, thus eliminating the restriction to a single compute node.

As explained above, typical usage scenarios in Peacemaker hardly need more than one compute node, so that a shared memory parallelization strategy is an appropriate choice. OpenMP parallelization was selected for this purpose, since it is included in any modern compiler, integrates easily into the code, and reduces to a serialized code, if the compiler is unable to handle OpenMP directives, or if needed for debugging purposes. The code was written in such a way that each QCE iteration is independent, so that the final piece of code that handles parallelization is extremely simple to read and understand. It is shown in Listing 4.4.

This code unrolls the double loop, which runs over all $n_{a_{\mathrm{mf}}}$ and $n_{b_{\mathrm{xv}}}$ values of a_{mf} and b_{xv} and distributes each loop element over all available CPUs. The subroutine qce_main performs the QCE iteration for the given $(a_{\mathrm{mf}}, b_{\mathrm{xv}})$ pair and for each temperature specified in the input file. Arguments to qce_main were not specified in the listing and neither were other bookkeeping operations. Compared to other parallelization techniques (such as OpenMPI) or different approaches with OpenMP, this piece of code is extremely simple, efficient, and scalable. This simplicity

Listing 4.4: The OpenMP "parallel do" construct chosen for Peacemaker's parallelization.

```
!$OMP PARALLEL
!$OMP DO COLLAPSE(2)
      do iamf = 1, namf
           do ibxv = 1, nbxv
                ...
                call qce_main( ... )
                ...
           end do
      end do
!$OMP END DO
!$OMP END PARALLEL
```

was made possible by careful design of all data structures and procedures and most importantly by exploitation of Fortran's strict rules for variable declaration and parameter passing.

5 Applications of QCE Theory

QCE calculations yield two important system properties: the partition function and the equilibrium cluster populations; both over any desired range of thermodynamic states. While partition functions can be readily used to extract thermodynamic potentials of interest, exciting opportunities for making use of the equilibrium cluster populations have emerged recently.[23,100] Simply adding up populations of carefully chosen clusters gives access to equilibrium constants of acid-base reactions. In the following, this procedure is demonstrated for liquid water. Afterwards, a study on thermochemical data of a selection of binary amide/water mixtures follows.

5.1 Predicting the Ionic Product of Water

5.1.1 Motivation

The chemistry of water is largely determined by acid-base reactions.[21] Indeed, one of the earliest taught and best known properties to new chemistry students is the self-dissociation reaction of water

$$2\,H_2O_{(aq)} \rightleftharpoons H_3O^+_{(aq)} + OH^-_{(aq)}\,, \tag{5.1}$$

the corresponding law of mass action (brackets denote concentrations)

$$K = \frac{[H_3O^+][OH^-]}{[H_2O]^2}\,, \tag{5.2}$$

and the value of the associated dissociation constant at standard state conditions

$$K_W = K[H_2O]^2 = 1 \times 10^{-14}\,\mathrm{mol}^2\,\mathrm{l}^{-2}\,, \tag{5.3}$$

also called ionic product of water. This seemingly inconspicuous number is proof of water's remarkable ability to catalyze its own dissociation. Consider the equivalent gas phase reaction

$$2\,H_2O_{(g)} \rightleftharpoons H_3O^+_{(g)} + OH^-_{(g)}; \quad \Delta G^0 \approx 1628\,\mathrm{kJ\,mol^{-1}}, \tag{5.4}$$

where the standard state Gibbs energy ΔG^0 was estimated by a simple density functional theory calculation.[a] The corresponding equilibrium constant

$$K_W^{(g)} = [H_2O]^2 \exp\left(-\frac{\Delta G^0}{RT}\right) \approx 1 \times 10^{-283}\,\mathrm{mol^2\,l^{-2}} \tag{5.5}$$

is several hundred orders of magnitude lower than eq. (5.3) (for saturated water vapor at 373.15 K and standard state pressure).

From a theoretical point of view, however, K_W is still low. The large number of particles required to describe self-dissociation and the subtle interplay between all of them make its prediction hard, both for theory and simulation. Few classical and semi-empirical studies have accepted the challenge, attempting to describe, for example, K_W variations in the supercritical region or other aspects of K_W,[101–106] but were unable to provide absolute numerical evaluation or mechanistic insight into eq. (5.1).

In principle, quantum chemical methods should be employed, if accurate predictions of K_W are sought. While static calculations are routinely applied to estimate gas-phase acidities, the treatment of acid-base reactions in aqueous solutions is not straightforward.[107] First principles molecular dynamics (FPMD) simulations are better suited, but the rarity of dissociation events in liquid water forces the use of additional constraints and/or approximations. At the moment, such simulations are unsuited to deliver quantitative estimates of K_W, but have provided detailed mechanistic insight into the self-dissociation, ion transport, and ion recombination in liquid water.[108–114]

Dissociation reactions in liquid water are triggered by strong local fluc-

[a] Performed at the B3LYP-D3/6-311++G** level of theory. Computational details won't matter for the arguments that follow.

tuations in the electrostatic potential caused by subtle rearrangements of its hydrogen bond network.[113,114] Such fluctuations are very short lived (< 0.1 ps) but long enough for the destabilized chemical bonds to break and for the resulting ions to separate by roughly 1 nm. This rapid severance is caused by multiple small amplitude motions of several protons along hydrogen-bonded water wires, similar in its nature to the Grotthus mechanism of excess proton transport in liquid water.[115] The process is driven by collective electrostatic forces from the solvent environment. Note that the separated ion pair is far from being in a metastable state. Subtle compression of the water wire (i.e., shortening of the distances between adjacent water molecules) may lead to rapid recombination. Similarly, modest expansion is a way for the transient charge separated ion pair to become metastable and to diffuse apart. Such changes in the water wire occur on somewhat longer time scales (< 0.5 ps), but are well below times of complete hydrogen bond network reorganization.[116]

QCE calculations provide a simple and inexpensive, but powerful alternative to FPMD simulations and permit quantification of K_W, as demonstrated in the following.

5.1.2 Ionic Dissociation in the QCE Framework

On first sight, QCE treatment of ionic dissociation seems to require separate components for neutral, cationic and anionic species. While a possible extension to ternary mixtures was sketched in Chapter 2, the canonical ensemble in which QCE is set up would require fixing the total particle number of each component, defeating the purpose of this work.

Note, however, that the standard cluster equilibrium reaction (here, for water)

$$n\,H_2O \rightleftharpoons (H_2O)_n \tag{5.6}$$

makes no assumption about the structure of the product clusters $(H_2O)_n$. Internal dissociation and formation of ion pair (IP) clusters violates no assumption made during the derivation of the QCE equations, thus equilibria of the type

$$n\,H_2O \rightleftharpoons (H_3O^+)(H_2O)_{n-2}(OH^-) \quad (n > 2) \tag{5.7}$$

can be easily treated within the existing QCE framework, as long as the product clusters are electronically and vibrationally stable.

Once stable IP clusters have been found, calculation of K_W essentially reduces to summing up equilibrium populations:

$$\begin{aligned} K_W &= [\mathrm{H_3O^+}][\mathrm{OH^-}] \\ &= \left[\frac{1}{V}\sum_{i=1}^{K} n_i(\mathrm{H_3O^+})N_i\right]\left[\frac{1}{V}\sum_{i=1}^{K} n_i(\mathrm{OH^-})N_i\right]. \end{aligned} \tag{5.8}$$

Therein, the sums run over all K clusters, V is the phase volume, N_i denotes the equilibrium population of cluster i, and $n_i(\mathrm{X})$ is equal to the number of $\mathrm{H_3O^+}$ or $\mathrm{OH^-}$ ions present in cluster i.

5.1.3 Computational Details

Choice of the Cluster Set

Choosing a representative cluster set is the most challenging task in constructing a QCE model of a liquid, even without the added difficulty of obtaining stable IP clusters. QCE can be well compared with conventional Boltzmann sampling, where thermodynamic potentials of a molecule with different conformations are weighted according to the energy (or Helmholtz energy) of the conformers.[a] However, instead of conformers of a single molecule, configurations of complex clusters must be considered in a QCE calculation. Following this analogy, a cluster set suitable for an accurate thermodynamic description of a liquid should contain all configurations at a given maximum cluster size and below. The maximum cluster size that should be considered depends on the temperature range investigated, with larger clusters being more relevant at lower temperatures. Such a cluster set shall be called complete.

While highly desirable, complete cluster sets are often unattainable, in particular for cluster sizes that span the length scale on which ionic dissociation occurs.[113,114] However, past studies have shown that it is

[a] In fact, QCE can be shown to be essentially equal to Boltzmann weighting of a set of clusters according to their Helmholtz energy, calculated with aid of the models presented in Section 2.4 and with additional constraints on the permitted phase volume.

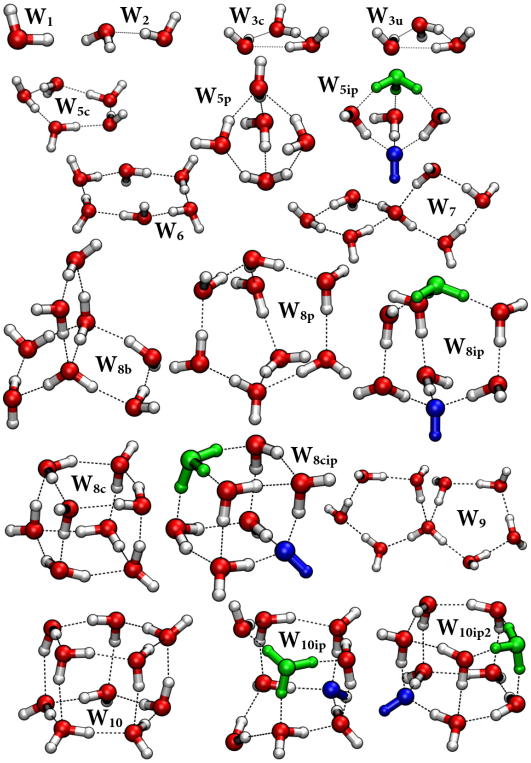


Figure 5.1:
Cluster set structures at the B3LYP-D3 level of theory. Computational details follow. For visual aid, hydronium and hydroxide moieties were colored in green and blue, respectively.

much more important to treat the main motifs that represent the liquid (like chains, cycles, or cubes), instead of striving for cluster set completeness.[37,38,51,68,117] These motifs are usually already present in small clusters. Motivated by these findings, small IP clusters were sought in this study.

Any attempt to internally dissociate two-fold coordinated clusters (such as $\mathbf{W_{3c}}$ or $\mathbf{W_{5c}}$, see Figure 5.1) failed due to recombination of the ions, even though such structures are generally featured in QCE calculations of neat water.[35,37,38,68,117,118] Electronically and vibrationally stable clusters could only be obtained by internal dissociation of three-fold coordinated clusters (such as $\mathbf{W_{8c}}$ or $\mathbf{W_{10}}$), leading to the IP structures $\mathbf{W_{8cip}}$, $\mathbf{W_{10ip}}$, and $\mathbf{W_{10ip2}}$. Similar structures were generated by placing the ions at three-coordinate bridgehead positions of polycyclic clusters, leading to

the propellane-like structure $\mathbf{W_{5ip}}$ and the (2.2.2)bicyclooctane-like cluster $\mathbf{W_{8ip}}$. The associated neutral structures $\mathbf{W_{5p}}$ and $\mathbf{W_{8p}}$ were included in the cluster set for the sake of completeness. All other clusters shown in Figure 5.1 are neutral and were found to be of importance in the QCE description of liquid water in earlier studies.[35,37,38,68,117,118] Larger clusters than those presented in Figure 5.1 may contribute to cluster distributions at low temperatures, but were not considered in this work.

Note that QCE calculations are not restricted to the often employed limiting cases of Zundel and Eigen structures. Nevertheless, clusters can still be classified based on their $H_2O\cdots H_3O^+$ hydrogen bond lengths. Clusters are Eigen-like if all three hydrogen bonds are shortened ($\mathbf{W_{8ip}}$ and $\mathbf{W_{8cip}}$), Zundel-like if one out of three hydrogen bonds is shortened ($\mathbf{W_{10ip2}}$), or intermediate if two out of three hydrogen bonds are shortened ($\mathbf{W_{5ip}}$ and $\mathbf{W_{10ip}}$). Cluster geometries were published online.[23]

Quantum Chemical Calculations

Energy calculations, geometry optimizations and harmonic frequency analyses were performed using the Orca 3.0 quantum chemistry code.[97] Various levels of theory were invoked. These are Hartree–Fock (HF), density functional theory (DFT), second-order Møller–Plesset perturbation theory (MP2), and a local variant of coupled cluster with single, double, and perturbative triple excitations (CCSD(T)). HF calculations were performed using the 6-311++G** basis set.[119,120] In DFT calculations, the hybrid functionals B3LYP[121] and PBE0[122] were used in combination with the def2-TZVP basis set.[123] Both functionals were employed standalone (labeled B3LYP and PBE0), in combination with D3(BJ) dispersion correction (labeled B3LYP-D3 and PBE0-D3),[124,125] as well as in combination with both D3(BJ) and a geometrical counterpoise correction scheme (labeled B3LYP-D3,gCP and PBE0-D3,gCP) to account for the basis set superposition error (BSSE).[126] Furthermore, the highly efficient composite DFT method PBEh-3c was used.[127] MP2 calculations were restricted to single point energy evaluations based on the B3LYP-D3,gCP geometries and performed with the aug-cc-pVTZ basis set.[128] Likewise, CCSD(T) calculations were single point energies, only, based on the same DFT geometries. CCSD(T) energy calculations were sped up by

Table 5.1: Methods, basis sets, employed corrections, and labels used in this study. The last column indicates whether geometry optimizations were performed or if structures are based on a different method. The same applies to frequency analyses.

label	method/basis set	corrections	geometry
HF	HF/6-311++G**		optimized
B3LYP	DFT-B3LYP/def2-TZVP		optimized
B3LYP-D3	DFT-B3LYP/def2-TZVP	D3	optimized
B3LYP-D3,gCP	DFT-B3LYP/def2-TZVP	D3, gCP	optimized
PBE0	DFT-PBE0/def2-TZVP		optimized
PBE0-D3	DFT-PBE0/def2-TZVP	D3	optimized
PBE0-D3,gCP	DFT-PBE0/def2-TZVP	D3, gCP	optimized
PBEh-3c	DFT, composite method	D3, gCP	optimized
MP2	MP2/aug-cc-pVTZ		B3LYP-D3,gCP
CCSD(T)	DLPNO-CCSD(T)/CBS		B3LYP-D3,gCP

employing the efficient domain based local pair natural orbital (DLPNO) method[129] and extrapolated[130] to the complete basis set (CBS) limit from results obtained using aug-cc-pVTZ and aug-cc-pVQZ basis sets.[128] All methods, basis sets, corrections, and labels employed in this study are summarized in Table 5.1. Tight SCF and geometry convergence criteria (as defined by Orca)[97] were applied in all calculations and fine numerical quadrature grids (grid5)[97] were used in all DFT calculations.

Adiabatic binding energies (defined in Section 2.4, eq. (2.17)) are listed in Appendix S3 for all methods. As a short reminder, these are adiabatic interaction energies with respect to infinitely separated monomers in their relaxed geometries. Unless counterpoise corrected, these energies are subject to the basis set superposition error (BSSE) if incomplete local basis sets are used. While common in previous QCE studies, the supramolecular counterpoise correction scheme[131] was not used in this study. This traditional approach works satisfyingly for conventional clusters in which the intramolecular part of the BSSE cancels to a large degree in the evaluation of the interaction energy. If the subunits of a cluster do not exactly match the reference species (e.g., because of internal dissociation in IP clusters), the intramolecular BSSE contributes

significantly to the interaction energy. These errors cannot and should not be corrected with the traditional approach, thus the geometrical counterpoise correction scheme (gCP) was used instead.[126]

QCE Calculations

QCE calculations were performed using the new Peacemaker code described in Chapter 4 in its published version 2.4.[92] Two experimental reference values were used to determine optimum a_{mf} and b_{xv} values: the density of water at 298.15 K and its boiling point, both at P = 101.325 kPa (the same pressure used in all QCE calculations).[132] Molecular volumes were calculated using Gepol93's Esurf algorithm[94] with van der Waals radii taken from Bondi's compilation[133] and a probe radius of 0.14 nm. Optimized values for a_{mf} and b_{xv} are given in Appendix S3.

5.1.4 Results

The temperature dependence of $pK_W = -\lg K_W$ is shown in Figure 5.2 for all investigated methods. Compared to the experiment,[134] significant inaccuracies can be observed for HF and the correlated methods MP2 and CCSD(T). While poor HF results can be easily justified with the insufficient quality of the method, the deviation of correlated methods, in particular that of CCSD(T), surprises. Note, however, that only electronic energies were obtained at these levels of theory and that equilibrium populations depend on the correct balance between enthalpic (primarily electronic) and entropic (primarily vibrational) contributions to the Helmholtz energy. This balance is distorted in mixed-type treatments of electronic energy and vibrational frequencies, explaining the mediocre quality of results.

All DFT methods show good-to-excellent agreement with the experimental reference. Combined application of dispersion and geometrical counterpoise correction performs generally favorable, bringing predicted pK_W values closer to the experiment. Their individual influence is opposite, though. While dispersion correction lowers pK_W values, gCP raises them again. In the case of PBE0, these two effects almost cancel out. In general, DFT methods suffer from a range of systematic errors, but

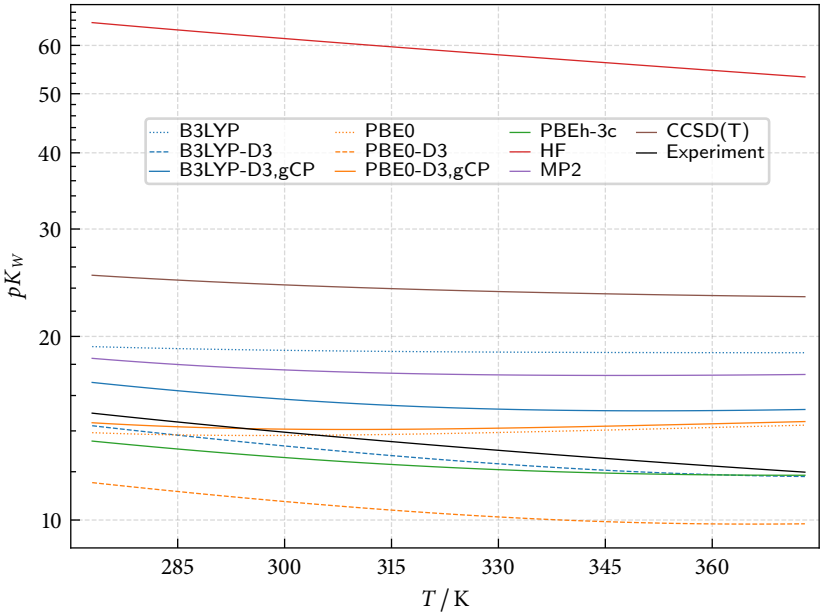


Figure 5.2: Ionic product pK_W as a function of temperature for all investigated methods and in comparison to the experiment.[134] Note the semi-logarithmic scale.

also profit from error cancellation. PBEh-3c is a composite method that was designed with systematic error compensation in mind and performs quite well in these delicate calculations. In terms of absolute deviation from the experiment, it is only beaten by B3LYP-D3. In anticipation of the conclusion, these results demonstrate that QCE theory is well capable of predicting the ionic product of water.

The ionic product of water is highly sensitive to the overall hydrogen bond network patterns, which can be seen from Figure 5.3. This figure shows monomer-normalized populations $\tilde{N}$ (defined in Section 2.3, eq. (2.5)) as a function of temperature. Please compare populations for $\mathbf{W_{10ip}}$ and $\mathbf{W_{10ip2}}$. The only difference between these two sandwich-like clusters is the lack of a single hydrogen bond connecting both pentamers (Figure 5.1). This seemingly inconspicuous loss of one out of 15 hydrogen

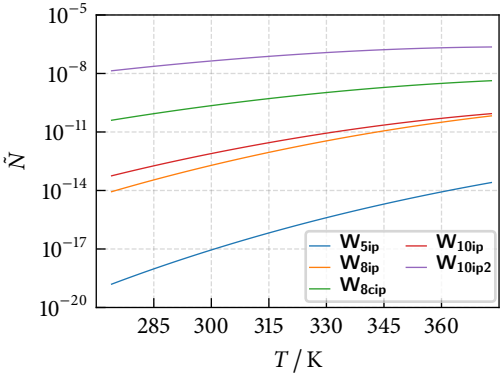


Figure 5.3: Monomer-normalized populations $\tilde{N}$ as a function of temperature for all clusters at the B3LYP-D3 level of theory.

bonds causes a decline in populations of around five orders of magnitude. Thus, special care must be taken when setting up QCE calculations to ensure that motifs containing maximum cooperative hydrogen bond patterns are present.

Finally, note that the distribution of the undissociated clusters, whose populations are several orders of magnitude larger, is essentially unaffected by these subtle effects contributing to pK_W and is comparable to what was reported previously.[35,37,38,68,118]

5.1.5 Conclusion

QCE theory was successfully applied to obtain semi-quantitative first principles estimates for the ionic product of water, as well as its temperature dependence. This was accomplished by inclusion of three-coordinate polycyclic cluster motifs, that are compliant with the Grotthus mechanism for excess ion transport and the concept of water wires observed from FPMD simulations.[113–115] Such calculations depend on the delicate balance between enthalpic and entropic contributions to the Helmholtz energy, requiring the use of fully consistent quantum-chemical treatment of all degrees of freedom (notably electronic and vibrational), but yield robust results if these premises hold. The theory is free from limiting concepts of Eigen and Zundel hydronium ions and can be easily expanded by inclusion of further IP clusters, offering a new perspective for the investigation of aqueous ionic properties. Finally, this study also opens the

path for a range of interesting and relevant studies on aqueous acid-base phenomena.

5.2 Predicting Thermochemical Data of Binary Amide/Water Mixtures

5.2.1 Motivation

In the previous study on the ionic product of water, the importance of choosing the QCE cluster set has already been highlighted. In principle, one should strive for a complete cluster set, both in terms of maximum cluster size and maximum cluster variation. Much like basis sets in quantum chemistry, however, practical QCE cluster sets are usually incomplete and one has to deal with errors originating therefrom.

In neat substances, the inclusion of important cluster motifs has been shown to be sufficient to ensure accurate predictions of most quantities of interest.[37,38,51,68,117] In binary systems, on the other hand, the delicate interplay between neat and mixed phases may lead to drastic errors in mixing-induced changes of thermodynamic potentials, if these phases are being described by cluster sets with different levels of completeness.[31]

As an example, consider the calculation of Gibbs energies of mixing, which are available from QCE calculations by evaluating

$$\Delta_{\mathrm{mix}} G(x_1) = G(x_1) - x_1 G_1 - x_2 G_2 \,. \tag{5.9}$$

Therein, x_1 and $x_2 = 1 - x_1$ refer to the mole fractions of components 1 and 2, respectively, and G_1 and G_2 are the corresponding Gibbs energies of the neat phases. If one phase is described by a more complete cluster set than the other, the calculation of Gibbs energies of mixing will be erroneous. In unfortunate cases, the demixing of two substances might be predicted, even if they mix completely in reality.

Since it is practically hard to obtain complete cluster sets for all phases, the concept of balanced cluster sets (i.e., cluster sets that describe all phases equally 'well') was introduced in the seminal work on bQCE theory.[31] To date, however, there is no recipe that defines such a balanced cluster set and many questions remain open.

The search for such a recipe is being severely hindered by the lack of a systematic approach to generate QCE cluster sets. As of now, all cluster sets have been obtained by educated guesswork and chemical intuition.

In an attempt to overcome empiricism, a systematic approach to generate cluster sets of a given maximum cluster size is presented for the first time in this work. In this approach, only global energy minimum structures are included in the cluster set, and their size is being systematically varied. Errors originating from cluster set incompleteness are expected to cancel systematically in this procedure. This error cancellation comes at the cost of neglecting less stable energy minima, which are known to contribute to the thermochemistry of the system, mostly for entropic reasons, or which may contribute to other system properties, such as acid constants.[23,135] Nevertheless, results presented here should be fit to answer questions about the size of a cluster set (measured in terms of maximum cluster size); a question that is of high relevance in light of recent work on small cluster set predictions.[66] Once this size has been determined, future studies may continue the search for an optimal cluster set by systematic inclusion of less stable energy minima.

Apart from the systematic cluster set generation scheme, the improved binary mean field energy and exclusion volume are applied for the first time in this study. Geometric averaging of a_{mf} values is employed, which eliminates the need for binary reference data (see Section 2.7).

As test cases, aqueous mixtures of three simple amides were chosen: formamide (FA), *N*-methylformamide (NMF), and *N,N*-dimethylformamide (DMF). The chosen amides differ in the number of methyl groups attached to the amide bond (Figure 5.4) and they are the simplest systems containing the peptide bonding system O=C=N. They serve as an ideal model system for protein linkage[136] and their hydration properties are of fundamental interest to establish an understanding of protein stability, folding, and function.[137] As a consequence, there is a wealth of thermophysical reference data available, which makes these systems attractive test cases for validating the proposed measures.

Note that NMF is subject to *cis-trans*-isomerism and both isomers are shown in Figure 5.4. While the *trans*-monomer is more stable, liquid NMF (both neat and in aqueous solutions) contains a significant amount of *cis*-isomers,[138–143] and global minimum clusters may also contain *cis*-NMF.

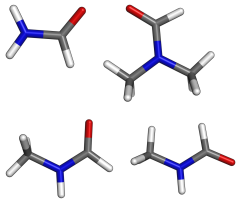


Figure 5.4: Amide monomers investigated in this study: FA (top left), DMF (top right), *trans*-NMF (bottom left), *cis*-NMF (bottom right). Note that the *cis-trans* nomenclature for amides refers to the relative position of the amide bond hydrogen and oxygen atoms.

5.2.2 Computational Details

Cluster Set Generation

In this work, clusters of the type $\mathbf{A}_i\mathbf{W}_j$ were considered, where **A** refers to an amide, **W** to water, and the indices correspond to the number of the respective monomer units constituting the cluster. Monomer amounts up to $i + j = 10$ were considered in this study, leading to a total of 175 cluster combinations. The number ten was chosen, because the binary properties investigated later were studied in mole fraction intervals of 0.1. The intent was to include mixed $(i = 1, j = 9)$ and $(i = 9, j = 1)$ clusters which reflect the smallest and largest system compositions at a molecular level. Furthermore, previous QCE studies have shown excellent results for smaller maximum cluster sizes.[68]

Given fixed numbers of amide and water molecules i and j, the global minimum structure of the cluster $\mathbf{A}_i\mathbf{W}_j$ was sought. Since an ab initio search for this minimum is computationally infeasible for the large numbers and sizes of the clusters investigated, initial structures were obtained by running a genetic-algorithm-based global structure optimization at a classical force field level. The classical force field employed in this work was the general Amber force field (GAFF).[84] Energy evaluations and local geometry optimizations were performed by the Amber 14 molecular dynamics code.[144] Settings were provided by Ogolem 2016-NG,[145] a genetic-algorithm-based global structure optimization framework. On a genetic pool of 100 species, 50 000 global geometry iterations were performed by Ogolem. A mixture of different crossover and mutation strategies was employed and a fitness-based diversity checker was used. For computational details with respect to the genetic algorithm, the reader is referred to ref. [146].

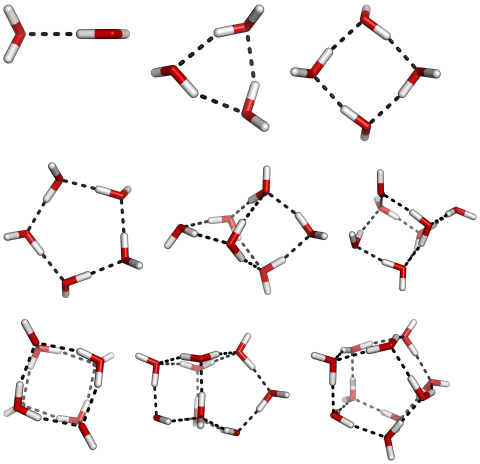


Figure 5.5:
Neat water clusters investigated in this study. Clusters are ordered as in Table 5.2 and arranged in rows first. Hydrogen bonds were determined by a loose geometric criterion and are for visualization only.

The minimum structures were subsequently refined by DFT-level geometry optimizations using the Orca 3.0 quantum chemistry code.[97] Therein, the low-cost PBEh-3c composite method was used, both for geometry optimizations and frequency analyses.[127] This method keeps the computational efforts reasonable, has shown good accuracy for thermophysical data, and performed well in previous QCE studies.[23,66] Computational details of the quantum chemical calculations were equal to those described in Section 5.1.3.

Of course, there is no guarantee that

1. the genetic algorithm has returned the global energy minimum at the force field level

2. the global energy minimum at the force field level matches with that of the DFT method

3. the global energy minimum of the DFT method is the true global energy minimum of the system.

Nevertheless, global minimum structures of small water clusters (Figure 5.5) are known and match with those obtained from the procedure described above, lending it credibility.[147,148]

At this point, one should keep in mind that the intent of the presented procedure is to attempt a systematic QCE cluster set generation. The proposed method is free from empiricism, even if a certain cluster is not a true global energy minimum. Previous approaches to cluster set generation relied on intuition and educated guesswork and are not viable for a systematic investigation of such a large number of clusters.

Note that monomer geometries were allowed to relax at all stages described above, but their internal geometries were not explicitly mutated. Unfortunately, local geometry optimizations were hardly ever found to change the *cis*-*trans* conformation of NMF, even though certain amounts of *cis*-NMF are part of most global minimum structures. In order to cover these structures, as well, all permutations of *cis*- and *trans*-NMF monomers were explicitly optimized for each number of NMF molecules i present in a cluster. This procedure increased the total number of clusters investigated to 395.

In the following, the letters **F**, **C**, **T**, and **D** are used to refer to FA, *cis*-NMF, *trans*-NMF, and DMF, respectively. Adiabatic binding energies of the neat amide and water clusters are listed in Table 5.2 and the corresponding structural representations are shown in Figures 5.5 to 5.8. Trying to visualize the mixed clusters is probably a fruitless attempt, nevertheless, pictures can be found in Figures S1 to S3 and adiabatic binding energies are listed in Table S3, both in Appendix S4. Cluster geometries and vibrational frequencies can be obtained upon request.

Given a maximum number of monomers $i = j = n_{\max}$, two types of cluster sets were generated: combination and permutation sets. Combination sets contain all possible combinations of $i = j = n_{\max}$ monomers and less. As an example, the largest neat clusters in a combination set of size three are trimers, and the largest mixed cluster is a hexamer (containing three monomers of each species). Combination sets of size two to five were considered in this study. Permutation sets contain all possible permutations of $i = j = n_{\max}$ monomers and less. Returning to the example above, the largest neat clusters in a permutation set of size three are trimers, and the largest mixed clusters are also trimers (two trimers with $(i = 1, j = 2)$ and $(i = 2, j = 1)$). Permutation sets of size two to ten were considered in this study. Differences between both sets are expected to diminish with increasing cluster set completeness, but should be very

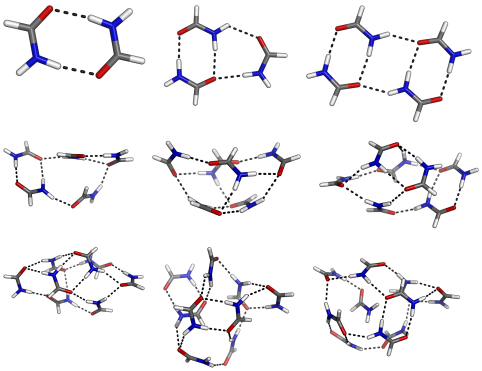


Figure 5.6:
Neat FA clusters investigated in this study. Clusters are ordered as in Table 5.2 and arranged in rows first. Hydrogen bonds were identified by a loose geometric criterion and are for visualization only.

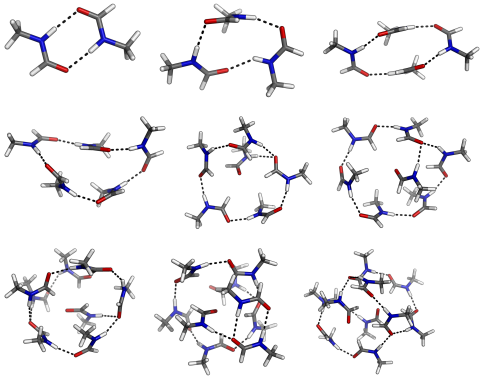


Figure 5.7:
Neat NMF clusters investigated in this study. Clusters are ordered as in Table 5.2 and arranged in rows first. Hydrogen bonds were identified by a loose geometric criterion and are for visualization only.

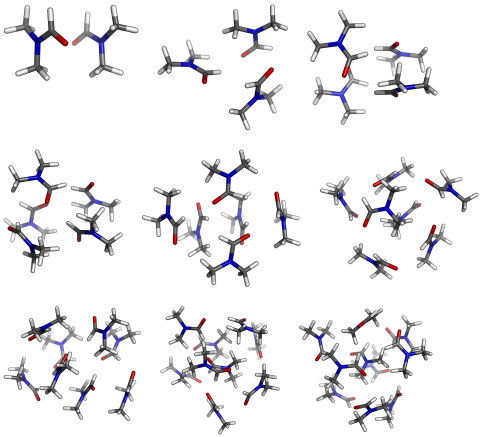


Figure 5.8:
Neat DMF clusters investigated in this study. Clusters are ordered as in Table 5.2 and arranged in rows first.

Table 5.2: Adiabatic binding energies of the neat amide and water clusters in kJ mol^{-1}. The cluster label of NMF represents the *cis-trans* ratio of the global minimum structure.

FA		NMF		DMF		water	
$\mathbf{F_1}$	0.00	$\mathbf{T_1}$	0.00	$\mathbf{D_1}$	0.00	$\mathbf{W_1}$	0.00
$\mathbf{F_2}$	−61.23	$\mathbf{C_2}$	−56.35	$\mathbf{D_2}$	−30.75	$\mathbf{W_2}$	−25.15
$\mathbf{F_3}$	−102.76	$\mathbf{C_2T_1}$	−93.24	$\mathbf{D_3}$	−61.94	$\mathbf{W_3}$	−78.30
$\mathbf{F_4}$	−163.73	$\mathbf{C_2T_2}$	−155.70	$\mathbf{D_4}$	−109.08	$\mathbf{W_4}$	−144.66
$\mathbf{F_5}$	−205.90	$\mathbf{C_2T_3}$	−201.12	$\mathbf{D_5}$	−139.15	$\mathbf{W_5}$	−190.34
$\mathbf{F_6}$	−261.06	$\mathbf{C_1T_5}$	−245.31	$\mathbf{D_6}$	−175.72	$\mathbf{W_6}$	−236.18
$\mathbf{F_7}$	−320.91	$\mathbf{C_1T_6}$	−299.19	$\mathbf{D_7}$	−220.20	$\mathbf{W_7}$	−295.40
$\mathbf{F_8}$	−378.03	$\mathbf{T_8}$	−350.17	$\mathbf{D_8}$	−253.11	$\mathbf{W_8}$	−375.00
$\mathbf{F_9}$	−437.00	$\mathbf{C_3T_6}$	−405.97	$\mathbf{D_9}$	−288.39	$\mathbf{W_9}$	−423.73
$\mathbf{F_{10}}$	−490.88	$\mathbf{C_2T_8}$	−463.54	$\mathbf{D_{10}}$	−333.26	$\mathbf{W_{10}}$	−482.40

relevant for small cluster sets.

QCE Calculations

QCE calculations were performed with a pre-release version of Peacemaker 2.6.[92] The pressure was set to 101.325 kPa in all calculations. Temperatures were evenly spaced in 1 K steps, and the intervals were chosen large enough to cover the complete liquid range of the neat substances. System compositions were varied in 0.1 mole fraction steps.

To obtain optimized parameters for the neat substances, calculations were compared against experimental reference densities and boiling points at the specified pressure (for FA,[149,150] NMF,[150,151] DMF,[150,152] and water[150,153]).

The a_{mf} and b_{xv} parameters of the neat substances were sampled on a grid between 0 J m^3 mol^{-2} and 2.5 J m^3 mol^{-2} (for a_{mf}) and between 1.0 and 2.0 (for b_{xv}), both in 0.001 unit steps. No sampling was performed for the binary systems, where the improved binary mean field and exclusion volume were applied. The required mixed mean field parameter a_{mf}^{ab} was obtained by geometric averaging of the neat parameters (see Section 2.7).

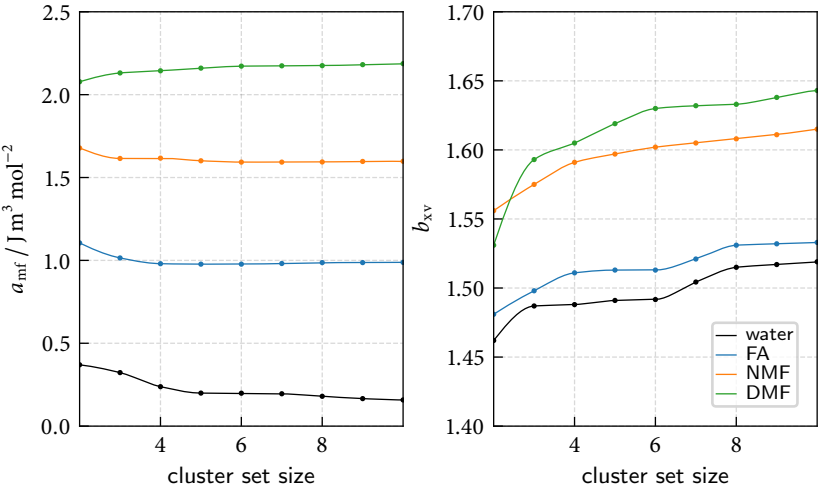


Figure 5.9: Optimized QCE parameters a_{mf} and b_{xv} of the neat components as a function of the cluster set size. Lines are meant to guide the eye.

5.2.3 Results

Parameter Sampling

Before starting the discussion of thermodynamic potentials, we shall look at the optimized QCE parameters a_{mf} and b_{xv} of the neat substances, which are shown in Figure 5.9. The theoretical foundations of QCE theory demand that large clusters are being populated at lower temperatures than small clusters. Thus, systematic addition of larger clusters should not affect properties in the investigated temperature range anymore, and QCE parameters should also converge. Vice versa, convergence of QCE parameters with respect to cluster size is an indicator of cluster set size convergence. Such convergence is the case for formamide, and can be assumed with some confidence for the other systems, though future studies should extend the maximum cluster size to be certain. Motivated by these findings, one should also expect convergence of the thermodynamic potentials, at least for the very largest cluster sets.

Energetics of Mixing

QCE calculations yield the system partition function and any derived thermodynamic potential at a specified temperature, pressure, and system composition. Consequently, mixing-induced changes of a thermodynamic potential, such as the Gibbs energy of mixing $\Delta_{\mathrm{mix}}G$ and the enthalpy of mixing $\Delta_{\mathrm{mix}}H$, are easily available by eq. (5.9) or analogous expressions.

Often, it is more insightful to look at excess quantities, where the contributions of an ideal mixture are subtracted. In this work, excess enthalpies H^{E} and Gibbs energies G^{E} were calculated according to

$$H^{\mathrm{E}}(x_W) = \Delta_{\mathrm{mix}}H(x_W) - \Delta_{\mathrm{mix}}H^{\mathrm{id}}(x_W) \quad \text{and} \tag{5.10}$$

$$G^{\mathrm{E}}(x_W) = \Delta_{\mathrm{mix}}G(x_W) - \Delta_{\mathrm{mix}}G^{\mathrm{id}}(x_W) \tag{5.11}$$

where the ideal contributions to enthalpy and entropy of mixing are given by

$$\Delta_{\mathrm{mix}}H^{\mathrm{id}}(x_W) = 0 \quad \text{and} \tag{5.12}$$

$$\Delta_{\mathrm{mix}}S^{\mathrm{id}}(x_W) = N^{\mathrm{tot}}k(x_W \ln x_W + x_A \ln x_A)\,. \tag{5.13}$$

In the preceding equations, the indices W and A refer to water and amide, respectively.

We shall now turn to the discussion of thermodynamic potentials. Excess Gibbs energies of mixing obtained from permutation sets are shown in Figure 5.10. To avoid cluttering of these figures, only results for the two smallest (dashed lines) and for the two largest cluster sets (solid lines) are included. Convergence with respect to the cluster set size can be observed in all systems, in the sense that deviations between the large cluster sets are small (within chemical accuracy) compared to deviations between the small cluster sets. Unfortunately, the experimental reference cannot be reproduced satisfactorily by any calculation. Predictions are qualitatively and quantitatively wrong at any cluster set size.

As argued above and in previous studies,[31] excess Gibbs energies that are too positive may originate from an imbalanced cluster set. One might expect better agreement with experiments if combination sets are used

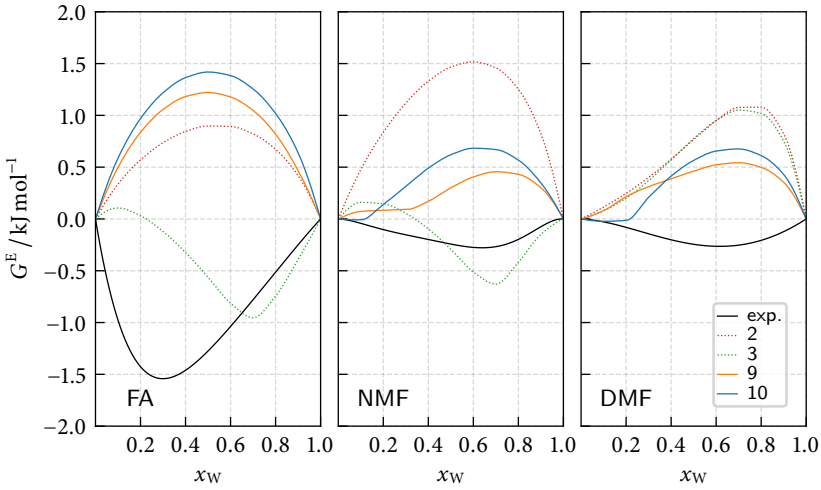


Figure 5.10: Excess Gibbs energy of mixing G^E as a function of the water mole fraction obtained from permutation sets. Line colors indicate different cluster set sizes. Experimental reference data was taken from ref. [154].

instead. These sets feature mixed phases that are characterized by a larger number of clusters (and therefore system states), which should stabilize the mixture with respect to the neat substances. This is indeed the case in Figure 5.11, where excess Gibbs energies obtained from combination sets are presented. Most excess Gibbs energy curves show the correct sign and are quantitatively in better agreement with the experiment. However, most curves are spiked towards $x_W \approx 0.5$ and show a significant degree of concave curvature to the left and right of this minimum. In open systems, these effects would lead to a spontaneous phase separation into neat and mixed phases. Such effects are in contrast to what is being observed experimentally, thus results for combination sets also fail to correctly describe the thermochemistry of the systems.

In order to understand why excess Gibbs energies $G^E = H^E - TS^E$ are in disagreement with experiment, it is useful to look at enthalpic (H^E) and entropic (S^E) contributions, separately. Excess enthalpies obtained from combination sets are shown in Figure 5.12. These curves demon-

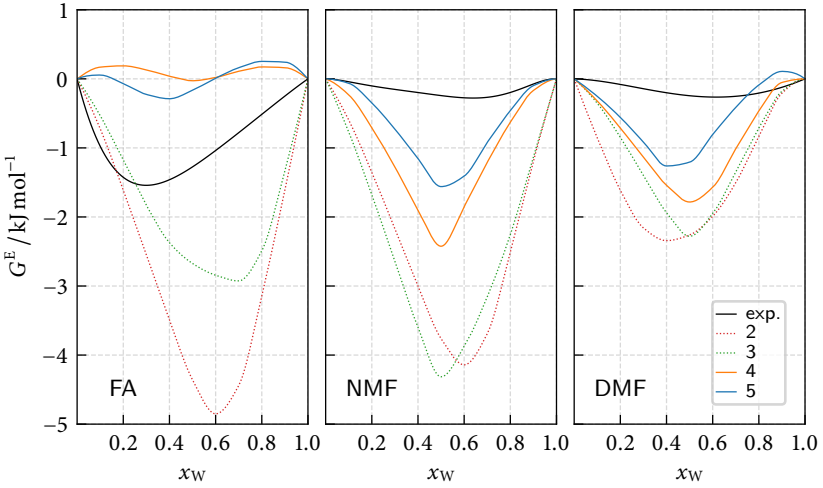


Figure 5.11: Excess Gibbs energy of mixing G^E as a function of the water mole fraction obtained from combination sets. Line colors indicate different cluster set sizes. Experimental reference data was taken from ref. [154].

strate remarkably, why combination sets fail to correctly describe the thermochemistry of the systems and should not be used, even if they lead to apparently better predictions of the Gibbs energy. Excess enthalpies obtained from combination sets are too negative and may deviate by more than 10 kJ mol^{-1} from experimental reference data. Like any other method that is capable of describing liquids and binary mixtures, QCE theory employs a number of approximations and is subject to multiple sources of errors, which will be discussed later. Deviations as those seen in Figure 5.12, however, can with great confidence be attributed to the combinatorial design of the cluster set.

Excess enthalpies obtained from permutation sets are presented in Figure 5.13. Unlike all other results presented so far, these curves are qualitatively correct for the largest cluster sets and also perform remarkably well on a quantitative level. While predictions for FA are essentially perfect, certain deviations between experiment and bQCE remain for NMF and DMF. However, these discrepancies are within chemical accuracy

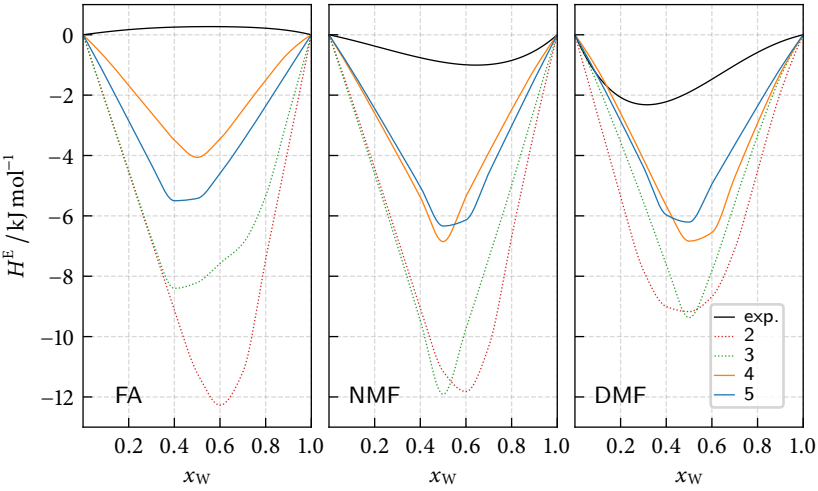


Figure 5.12: Excess enthalpy of mixing H^E as a function of the water mole fraction obtained from combination sets. Line colors indicate different cluster set sizes. Experimental reference data was taken from ref. [154].

and can easily be attributed to a range of other error sources, which will be discussed shortly.

Since predictions of H^E are accurate within the limits of the method, deviations between experiment and G^E must be entropic in their origin. This is not surprising, since only global energy minima are included in the cluster sets. These cluster sets lack less stable energy minima that contribute largely to the entropy of the system.[135]

Please note that all excess quantities presented so far, whether they are incorrect or correct, deviate from zero. In other words, all systems show non-ideal mixing behavior, even though the mixed mean field parameters a_{mf}^{ab} were obtained from mixing rules that are commonly associated with ideal mixing behavior. That is because bQCE theory is capable to include non-ideal mixing effects at the intra-cluster level.

In agreement with these considerations, mixed clusters are being significantly populated in all systems, which is shown in Figure 5.14 for the largest combination set. When comparing populations among all amides, NMF and DMF show a markedly different behavior than FA. In

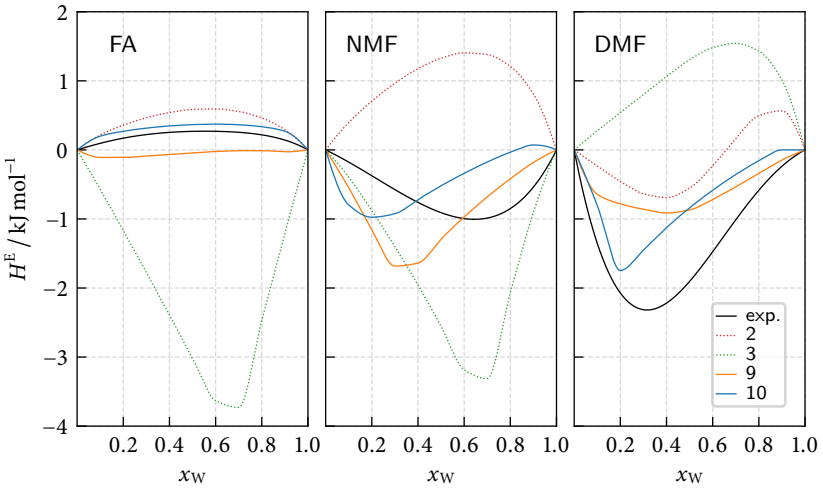


Figure 5.13: Excess enthalpy of mixing H^E as a function of the water mole fraction obtained from permutation sets. Line colors indicate different cluster set sizes. Experimental reference data was taken from ref. [154].

FA, neat and mixed clusters are roughly equally populated in 1:1 mixtures, whereas the mixed clusters dominate NMF and DMF mixtures. This behavior is somewhat correlated with the preferential solvation behavior of the amides,[154] but cannot be entirely explained thereby. At this point, one may only speculate about further origins of these differences. One possible explanation might be related to the presence of methyl groups in NMF and DMF. These groups introduce additional interactions to the system (dispersion interactions), which differ greatly from the primarily hydrogen bond dominated FA/water mixtures. Different types of interactions can be accurately described on the intra-cluster level, but less well by the mean field approach. Such effects would favor the population of mixed clusters.

5.2.4 Discussion and Conclusion

The preceding results seem sobering on first sight, but should be considered in light of their premises.

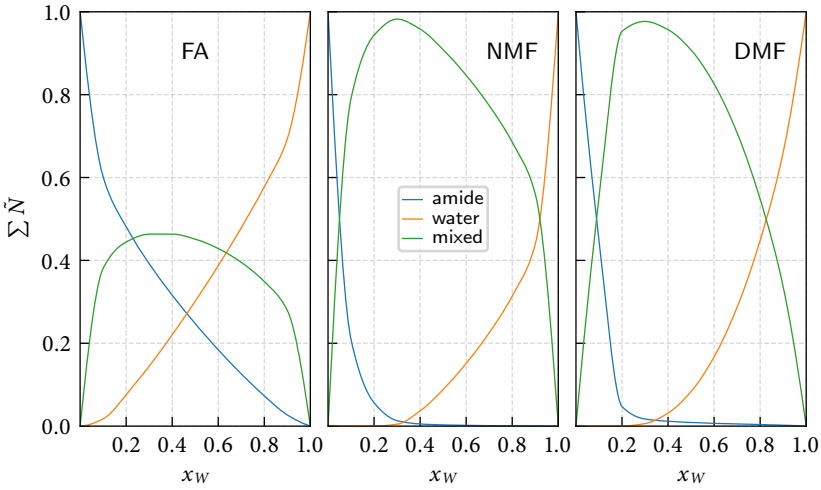


Figure 5.14: Sum of the neat amide, neat water, and mixed cluster populations $\tilde{N}$ as a function of the water mole fraction obtained from permutation sets of size ten.

From the point of view of bQCE theory, there are various sources of errors that contribute to deviations between bQCE theory and experiments.[a] These are incompleteness of, or inaccuracies in:

- the cluster set (completeness/balance),
- the quantum chemical method,
- the cluster partition functions (e.g., the harmonic treatment of vibrations or the neglect of rotational-vibrational coupling), and
- the underlying equation of state.

Despite of these error sources, chemically accurate excess enthalpies were obtained and as argued above, one should not expect more from a limited cluster set that employs only global minimum structures. Note

[a] The other side of the medal would be experimental errors, which are often not negligible.

that for the amide systems investigated in this work, small cluster sets that employ only dimers or trimers do not capture the thermochemistry of the system. Consistent accuracy across all systems could only be reached for the largest (size ten) permutation sets.

The approach presented here is the first attempt at a systematic cluster set generation and opens the path for a range of follow-up studies. After having determined the appropriate maximum cluster set size, less stable energy minima should be included in future studies. The genetic-algorithm-based solution is, in principle, fit for the task, since it keeps a pool of cluster structures that deviate from the global minimum. These conformations could easily be added to the cluster set. If fruitful, each step of the cluster set generation scheme could be automated and a black box scheme could be devised.

Methodological details are likely to change in the future. In particular, the force field will most likely be replaced by low-cost quantum chemical methods (i.e., suitably corrected minimum basis set or semiempirical methods).[155]

A second major step towards a mostly empiricism and parameter-free ab initio model for binary mixtures has been achieved in this study: the elimination of binary QCE parameters (Section 2.7). This was made possible by physically motivated improvements of the binary mean field energy and exclusion volume in combination with mixing rules for the mixed mean field parameter a_{mf}^{ab}. Unlike in traditional equations of state, such mixing rules do not imply ideal mixing behavior, since bQCE theory is capable of treating non-ideal mixing effects at the intra-cluster level. As a result, no binary reference data was needed to acquire the results presented above.

Finally, thanks to robust design of the new Peacemaker code, no bQCE calculation, not even one of those involving $\approx$ 100 clusters, did cause convergence problems.

In conclusion, the combination of systematically generated cluster set and improved bQCE methodology permits the chemically accurate prediction of excess enthalpies in amide/water mixtures without need for binary reference data. A systematic route to improvements of these results was sketched, and the new Peacemaker QCE code is numerically stable enough to handle the required cluster sets.

6 Interfacial Water Dynamics

A large number of phenomena in nature, science, and technology take place at interfaces of water. Protein folding, the operation of solar cells, or corrosion are few examples that demonstrate the ubiquity and importance of these interfaces. As a consequence, studies date back many years and a wealth of literature on the subject is available. For some excellent reviews, the reader is referred to refs. [156, 157].

While research performed before the last decade of the twentieth century was mostly focused on the measurement of macroscopic properties, such as surface tension and potential, focus has shifted to microscopic surface phenomena since then, thanks to the development of powerful experimental techniques, advances in theory, and most prominently the rise of computer simulations.[156,157] While initially used to test the validity of statistical mechanical approximations, computer simulations have grown independent and contributed largely to our current understanding of microscopic interfacial phenomena.[156]

In this chapter, two of these phenomena will be studied: the influence of electric fields on interfacial water dynamics and the impact of spherical confinement on the properties of water, each of which will be introduced and motivated separately in the following two sections.

6.1 Water Dynamics at a Janus Interface

6.1.1 Motivation

A promising technique to modify interfacial properties in modern nanotechnology is electrowetting,[158] which has already been exploited in numerous applications (e.g., in microfluidics, ink-jet printing, or electric control of optical properties).[159–165]

Electrowetting is the process of contact angle (θ_c) reduction and concomitant increase in pressure and density upon application of an electric field.[158,159,166] According to macroscopic thermodynamics, these effects depend only on the strength of the applied field. However, macroscopic theory is inadequate to describe new phenomena observed at the nanoscale and experiments become increasingly difficult, as well.[167] Computer simulations, on the other hand, are ideally suited to study such phenomena, and have been successfully applied to show that when surface-to-bulk fractions are high enough, electrowetting in a nanopore depends on field direction and polarity.[168] Studies have shown that in hydrophobic nanopores, a field-induced crossover from drying to wetting behavior occurs at lower field strengths when the field is aligned parallel to the interface rather than perpendicular.[168–170] Furthermore, in an electric field **E** perpendicular to the interface, spontaneous orientational polarization of water (caused by interfacial hydrogen bonding) either competes or cooperates with the field alignment, resulting in an asymmetric behavior at opposing surfaces. The contribution of spontaneous orientational polarization of water to the wetting free energy $\Delta\gamma$ is dominated by the term $-\mathbf{E}\cdot\mathbf{M}$, where **M** is the field-induced collective dipole moment. Since water is easier to polarize along the outgoing (E_{out}) than the incoming field (E_{in}), the field-induced reduction of the wetting free energy $\Delta_E\Delta\gamma$, is bigger for the field with outgoing direction, and $\Delta_{E_{\text{out}}}\Delta\gamma < \Delta_{E_{\text{in}}}\Delta\gamma$. By carefully tuning the field strength, a Janus interface[171] emerges,[168–170,172] where water is favorably attracted to one surface (hydrophilic), but repelled from the other (hydrophobic). These observations suggest a general way to control electrowetting in nanoporous materials, with the main determinants being not only field strength but also its direction and polarity.[168–170,172]

While recent studies have demonstrated novel ways to alter the macroscopic contact angle of water through modification of surface polarity or morphology,[173–175] the slow switching time (hours to days)[173] and hysteresis[174] pose current experimental challenges. Nanoelectrowetting, however, provides an exciting alternative,[176,177] and could promise fast actuation, as well as immediate and hysteresis-free response.

In principle, the field-induced response of a hydrophobic system can take place in two partially overlapping stages, characterized by distinct

time scales: The fast change in the interfacial free energy and the subsequent process of solvent transfer to a new equilibrium state. Depending on the system at hand, this could be a change in droplet spreading, nanopore wetting, electrostriction, or nanoparticle reorientation.[172,178]

Here, focus is set on the initial stage involving fast dynamics of interfacial polarization and local density adjustments in response to the imposition of the field. Although the dielectric response of water includes significant quadrupolar and higher multipole contributions to the overall polarization,[179] analyses were restricted to the dipolar response, since it provides the dominant contribution to the field-induced change in interfacial free energy.

A field-induced Janus interface can only be obtained when walls are initially hydrophobic and focus in the main text has been set on such systems. Nanoelectrowetting does generally depend on field direction, however, and hydrophilic confinement was studied as well. Results are shown in Appendix S5 and demonstrate that the pronounced asymmetry in polarization dynamics persists, even when static properties do not show Janus behavior.

6.1.2 Models and Methods

The model system employed in this study consists of a slab of 839 water molecules, confined between two parallel walls structured like a graphene bilayer and separated by a distance $D = 2.86\,\text{nm}$ (a so-called sandwich geometry, see Figure 6.1). This distance has been found to be sufficient to secure convergence to bulk behavior at the center of the bilayer for all properties investigated in the following.[168,170] The confinement walls consist of 767 carbon-like atoms, which were held at fixed positions during the simulation. The simulation box was shaped like a rectangular prism with edges $L_x = 3.19\,\text{nm}$, $L_y = 3.12\,\text{nm}$, and $L_z = 3.19\,\text{nm}$. For easy reference, the coordinate system shall be located as follows: The graphene bilayer is placed into the xy-plane. Thus, separations from the interface can be measured by specifying the z component of the position vector.

The confining bilayer is called graphene-like, because the chosen water–carbon interactions do not actually represent those of real graphene interfaces. Instead, simple Lennard-Jones interactions between the confine-

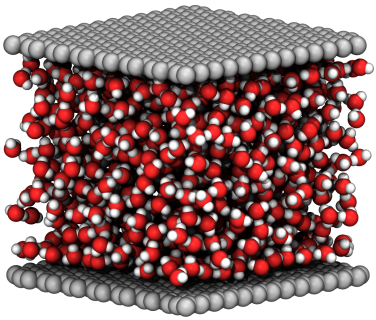


Figure 6.1: Snapshot of the system. Applying periodic boundary conditions renders confinement walls twice as thick (bilayer of graphene). Note that a periodic shift of the simulation cell in the direction of the surface normal results in a geometry where both walls are adjacent to each other, located in the center of the simulation cell. Figure taken from ref. [24].

Table 6.1: Lennard-Jones parameters σ_{CO} and ϵ_{CO} and the corresponding microscopic contact angles θ_m on a planar graphene-like surface, reported in ref. [180].

	σ_{CO}/nm	ϵ_{CO}/kJ mol^{-1}	θ_m/deg
hydrophilic	0.319	0.5016	69.9
hydrophobic	0.319	0.2508	127.8

ment atoms and the oxygen atoms of water were used, with parameters taken from Werder et al.[180] Distinctions were made between two water–confinement interaction parameter sets: an attractive one that shall be called hydrophilic and a less attractive one termed hydrophobic. These parameter sets are summarized in Table 6.1. The microscopic contact angle reported therein describes the curvature of tiny water droplets.[181–183] In analogy to their macroscopic counterparts, small droplets are described by a modified Young equation

$$\gamma_{sv} = \gamma_{sl} + \gamma_{lv} \cos\theta_m + \frac{\tau}{r_b}, \tag{6.1}$$

where γ_{ab} denotes the surface tension between phases a and b (with indices l, v, and s for liquid, vapor, and solid phase, respectively), τ is the line tension, and r_b the droplet base radius. This equation turns into Young's equation for macroscopic drops ($r_b \to \infty$). Note that initially, both walls are chemically equal. Asymmetries arise only after imposition

of the electric field.

Water molecules were described by the extended simple point charge (SPC/E) model,[184] which is an effective, rigid, three-site pair potential with Lennard-Jones and Coulomb interactions and implicit polarization-self-energy correction. It has already been used successfully in similar studies[168–170,172,178,185,186] and within the wide range of water potentials that share this setup, SPC/E is arguably one of the better ones.[187,188] It allows an acceptable description of structure, internal energy, density, and diffusion of ambient water[187–190] and gives a reasonable estimate of the critical temperature, the dielectric constant, and surface tension.[187,189,191] It has also been argued that SPC/E water satisfies the most important requirements for the description of hydrophobic effects and the distinction of hydrophobic and hydrophilic solvation.[192,193] Furthermore, a vast collection of simulation data on solid–liquid interfacial properties[194–203] makes the model attractive for comparison of this work's data with the literature. Moving away from ambient conditions reveals deficiencies of SPC/E (and of most other water models) and its predictive power has been questioned.[187,204] While promising (classical) alternatives are now emerging,[205,206] their availability in common simulation codes is not yet given. Please note that the choice of a particular force field may introduce some quantitative differences in orientational preferences of water near interfaces, but angle preferences due to anisotropic hydrogen bond interactions are a common and well-reproducible feature in a variety of models including ST2,[207] TIP4P,[208] and SPC/E.[183]

In the presence of electric fields, electronic polarization and dissociation must be considered, suggesting the use of electronically polarizable or even reactive force fields. However, with all components of the molecular polarizability of water being close to $1.5 \times 10^{-30}\,\mathrm{m}^3$,[209] the fields chosen in this study can induce a change in the dipole moment of water up to 0.2 %, only. Furthermore, the largest employed electric field of $1\,\mathrm{V\,nm^{-1}}$ is well below the dissociation threshold of $\approx 3\,\mathrm{V\,nm^{-1}}$ and is thus unable to significantly dissociate water molecules.[210,211]

An important quantity investigated in this work is the collective dipole

moment relaxation time

$$\tau_M = \int_0^\infty \frac{\langle \mathbf{M}(\mathbf{t}) \cdot \mathbf{M}(\mathbf{0}) \rangle}{\langle M^2(0) \rangle} \, dt \,, \tag{6.2}$$

which is closely related to the Debye relaxation time τ_D.[189,212,213] Equality between both quantities holds if Ewald summation with conducting boundary conditions is used; for non-conducting boundary conditions or a different treatment of long-range electrostatic interactions, however, τ_M needs to be scaled appropriately.[189,213] Literature values for τ_D range between 5 and 10 ps for bulk SPC/E water depending on the simulation methodology.[189,214–216]

Molecular dynamics simulations were performed in the canonical ensemble at $T = 300$ K using the Lammps molecular dynamics code.[217,218] The equations of motion were solved by Velocity-Verlet integration with a time step of 1 fs. Thermostating was achieved by a Nosé–Hoover thermostat with a time constant of 100 fs. Initial configurations were generated by randomly packing water molecules between the interfaces. Water molecules were kept rigid by the SHAKE[219] algorithm. Graphene atoms were not propagated, remaining rigid and frozen. Periodic boundary conditions were applied in all three dimensions. Lennard-Jones interactions and real-space Coulomb forces were cut off beyond 1.1 nm. Long-range Coulomb forces were treated in reciprocal space, using a particle-particle particle-mesh solver. Equilibrated systems were typically simulated for ≈ 17 ns with a time step of 1 fs.

All electric fields $\mathbf{E}_0$ were applied perpendicular to the interface, with strengths E_0 selected from the range 0.1–1.0 V nm^{-1}. Note that these fixed values correspond to a set of displacement fields $\mathbf{D}$, which are position independent.[179] The actual local field $\mathbf{E}(z)$ is subject to screening and depends on the distance to the interface. However, since conducting ("tin foil") boundary conditions were employed in this study, the actual fields in the bulk-like phase (beyond ≈ 0.5 nm away from the surfaces) are close to the input strengths E_0. The reader is referred to ref. [179] for an in-depth description of the theoretical framework for the dielectric response in terms of the above field functions, and the results for the

distance dependence of $\mathbf{E}(z)$ in the interfacial layer in closely related model systems. Dipole moments were calculated according to the IUPAC definition, where the vector points from negative to positive charges.

6.1.3 Orientational Preferences and Field Alignment

Due to its molecular asymmetry and anisotropic hydrogen bond interactions, water is orientationally biased near interfaces.[207,208,220–222] In order to optimize the hydrogen bond network, water dipoles align almost parallel to the surface, pointing just slightly away.[221] This causes spontaneous polarization, which can be quantified in terms of the collective interfacial dipole moment

$$\mathbf{M} = \sum_i \boldsymbol{\mu}_\mathbf{i} \quad (i \in \text{interface}), \tag{6.3}$$

where the sum runs over all water dipoles $\boldsymbol{\mu}_\mathbf{i}$ in the first hydration layer. Thanks to the rigid wall, this layer can be easily determined through the density layer profiles (Appendix S5) and is defined as the region between the wall and the first minimum in water density.

Consider now two identical, but opposing surfaces. In the absence of an applied field, the average polarization $\langle \mathbf{M} \rangle$ is of the same magnitude at each surface, but points in opposite directions. Consequently, average values of

$$\cos\phi = \frac{\mathbf{M} \cdot \hat{\mathbf{n}}}{M}, \tag{6.4}$$

the cosine of the dipole angle ϕ between $\mathbf{M}$ and the surface normal $\hat{\mathbf{n}}$, are equal in magnitude, but of opposite sign. These averages are shown in Figure 6.2. Once an electric field is applied perpendicular to the interface, water dipoles at both walls gradually align with the direction of the field. However, field alignment competes with the initial orientational preferences if the electric field is incoming (orange lines) and cooperates if the electric field is outgoing (blue lines). That is because water hydrogen atoms are pushed towards the interface in the former case and turned away therefrom in the latter scenario. This situation is shown schematically in Figure 6.3. As a result, polarization shifts in the direction of the field but remains asymmetric. At even higher fields than investigated in

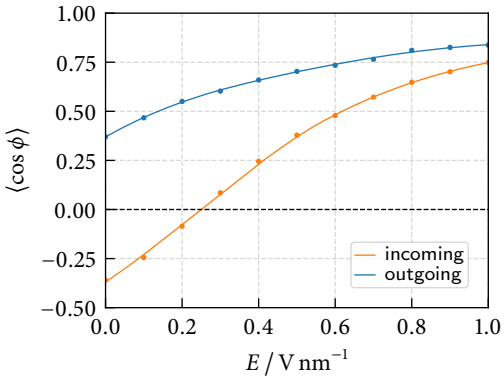


Figure 6.2: Field dependence of $\langle \cos \phi \rangle$, where ϕ is the angle between the collective interfacial dipole moment and the electric field **E** applied perpendicular to the interface. Lines are meant to guide the eye. Hydrophobic confinement.

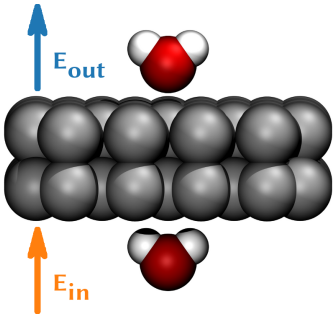


Figure 6.3: Schematic illustration of the asymmetric behavior of interfacial water molecules subject to incoming (E_{in}) and outgoing (E_{out}) electric fields.

this work, sensitivity to the field is diminishing and saturation seems to occur. Note that such a high alignment of water dipoles can be achieved without serious penalties in hydrogen bond number and free energies, and without distortion of tetrahedral coordination.[186,223–228]

For incoming electric fields, field alignment can compensate the angular bias of the hydrogen bonds at the interface (i.e., $\langle \cos \phi \rangle \approx 0$). The field strength at which that happens is slightly below 0.3 V nm^{-1}. Such fields are comparable to those in ion channels[229] and ionic colloids.[230] They are an order of magnitude weaker than local fluctuating fields present in liquid water and aqueous solutions.[231] Qualitatively similar results were obtained for hydrophilic walls (Appendix S5), however, cancellation of the orientational bias requires slightly higher field strengths, since stronger water–wall interactions must be overcome in these systems.

While such averages demonstrate the basic physics, dipole angle distri-

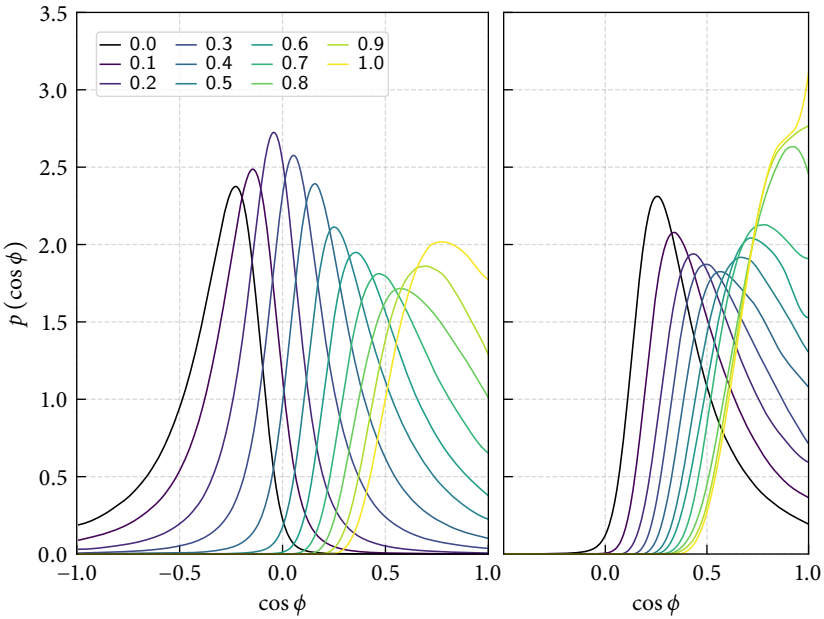


Figure 6.4: Probability density $p(\cos\phi)$ of the dipole angle ϕ for all investigated electric field strengths (given in V nm^{-1}). Left: incoming electric fields, and right: outgoing fields. The color gradient (blue → yellow) indicates growing field strengths. Note that the field-free distributions (black) are equal, apart from being mirrored. Hydrophobic confinement.

butions $p(\cos\phi)$ offer deeper insight into polarization near the interface. They are shown in Figure 6.4 for the hydrophobic system. The corresponding figures for the hydrophilic system can be found in Appendix S5. The color gradient (blue → yellow) indicates growing field strengths. Valuable information can be obtained from the widths of these distributions. For incoming fields, they decrease until they are narrowest around 0.3 V nm^{-1} and $\langle\cos\phi\rangle \approx 0$, whereas in stronger electric fields, they broaden again. Distributions in outgoing fields broaden monotonically. Since the restoring force which causes the decay of thermal fluctuations from the average orientation is strongest in systems with narrowest angle distributions, the same asymmetric and non-monotonic behavior can be observed in

the dynamic quantities that follow shortly.

Please note that the competition between field alignment and orientational preferences is only relevant on the nanoscale, where the surface-to-bulk ratio is significant. The range of the effects discussed above can be seen in Figure 6.5, which shows distance dependent probability densities $p(\cos\phi, z)$ in the field-free system (left) and under the influence of $E = 0.3\,\mathrm{V\,nm^{-1}}$ (right). Without the influence of an electric field (Figure 6.5, left), dipole angles outside of the solvation layers are symmetrically distributed around $\cos\phi = 0$ (i.e., distributions are bulk-like and unbiased). The orientational preferences discussed earlier manifest themselves only at small distances from the interface. Imposition of an electric field (Figure 6.5, right) affects the angle distributions at all positions, thus bringing the dipole orientations closer to the direction of the field (i.e., to larger values of $\cos\phi$).

The competition between spontaneous orientation and field alignment has a pronounced impact on interfacial densities, as well. The changes in the density maxima in the first hydration layer are illustrated in Figure 6.6. While macroscopic theory predicts an increase in water density (electrostriction) that is proportional to the compressibility of the liquid and the strength of the applied electric field squared,[172] an asymmetric wetting behavior can be observed at opposing surfaces on the nanoscale. For a more in-depth discussion, see Appendix S5.

6.1.4 Fluctuation Dynamics of Interfacial Water Polarization

In this section, polarization dynamics will be investigated, which should be subject to the same competition between field alignment and spontaneous interfacial polarization. For this purpose, dipole angle time correlation functions

$$C_{\cos\phi}(t) = \frac{\langle \delta\cos\phi(t)\,\delta\cos\phi(0)\rangle}{\left\langle [\delta\cos\phi(0)]^2 \right\rangle}, \tag{6.5}$$

were studied. Therein, $\delta\cos\phi(t) = \cos\phi(t) - \langle\cos\phi\rangle$ denotes fluctuations of $\cos\phi(t)$, as defined in eq. (6.4), around its ensemble average

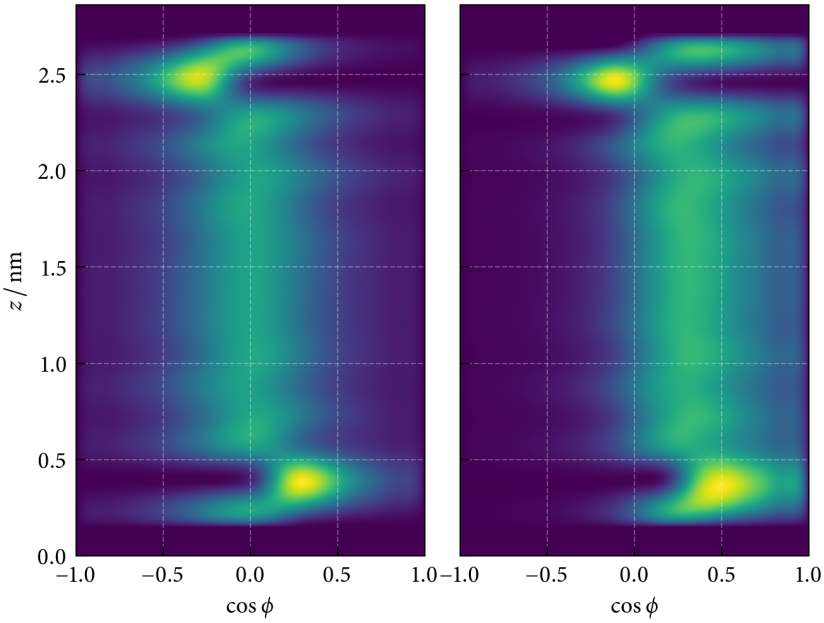


Figure 6.5: Distance dependent dipole angle probability densities $p(\cos\phi, z)$. Left: field-free; right: $E = 0.3\,\mathrm{V\,nm^{-1}}$. Electric field direction goes from bottom to top. Hydrophobic confinement.

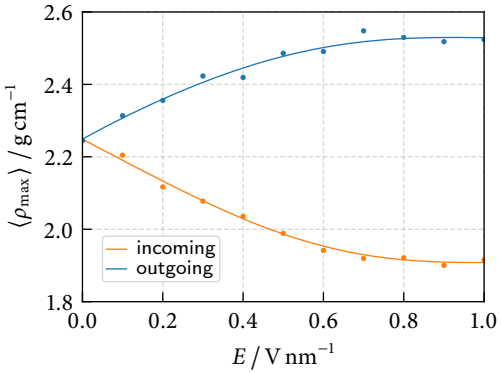


Figure 6.6: Maximum water density $\langle\rho_{\mathrm{max}}\rangle$ within the first hydration layer at various incoming fields E_{in} (orange) and outgoing fields E_{out} (blue). Lines are meant to guide the eye. Hydrophobic confinement.

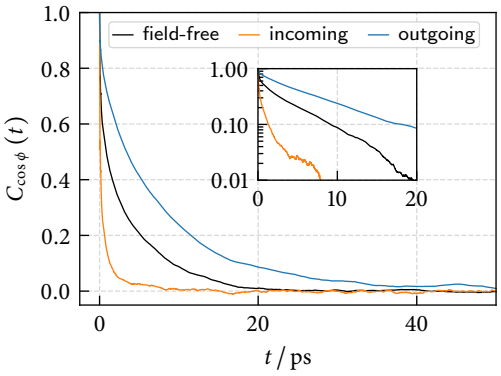


Figure 6.7:
Decay of dipole angle fluctuations without electric field (black), and in incoming (orange) and outgoing (blue) fields of strength 0.3 V nm^{-1}. The inset shows the correlation functions on a semi-logarithmic scale. Hydrophobic confinement.

$\langle \cos\phi \rangle$.

According to Onsager's regression hypothesis, these time correlation functions describe the dynamic response of the interfacial free energy to small changes in the electric field, since the dominant electric field contribution to the interfacial free energy is given by the relation

$$\Delta_E \Delta\gamma = -\frac{\langle EM\cos\phi\rangle}{A} + \dots , \tag{6.6}$$

with $\Delta_E\Delta\gamma$ being the field-induced change in wetting free energy $\Delta\gamma$ and A the surface area.

Results are shown in Figure 6.7 for an electric field $E_0 = 0.3\,\mathrm{V\,nm^{-1}}$. Note that no simple mono-exponential decay should be expected for $C_{\cos\phi}(t)$, which describes a collective phenomenon that involves multiple coupled molecular processes operating on similar time scales. This is indeed the case. All time correlation functions exhibit a sharp initial drop for times < 1 ps due to the coupling with fast molecular motions, such as vibrations, librations, and hydrogen bond dynamics.[196,197,232–234] While the decay seems to be roughly exponential for intermediate times in some fields, significant curvature on the semi-logarithmic plot is visible at larger times in all systems. In lieu of constructing a model that offers full mechanistic insight into the molecular processes contributing to the shape of these functions, characteristic decay times (τ_{int}, $\tau_{1/e}$) were

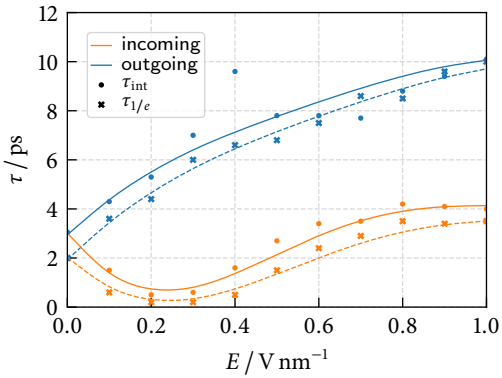


Figure 6.8:
Time scales $\tau_{1/e}$ (crosses, dashed lines) and τ_{int} (dots, solid lines) of the dipole angle fluctuation dynamics for incoming and outgoing fields. Lines are meant to guide the eye. Hydrophobic confinement.

extracted by a) analytic[a] continuation and integration of the time correlation functions and b) reporting times, where $C_{\cos\phi}(t)$ has decayed to $1/e$. The latter are subject to less statistical noise, nevertheless, correlation times from both procedures are compared in Figure 6.8. Observed results are essentially the same no matter how times are determined.

Without an electric field, interfacial fluctuations decay to $1/e$ within 2.0 ps near hydrophobic walls and within 3.4 ps (see Appendix S5) close to hydrophilic walls. Non-exponential behavior of orientational dynamics is expected close to interfaces and the extracted decay times fit well to previously reported literature data.[222,235–237] A more detailed discussion of the mechanisms causing this behavior will be given in Section 6.2.

Once more, imposition of an electric field results in appreciable asymmetries between incoming and outgoing electric fields. Fluctuations decay much faster (within ≈ 0.2 ps) when the electric field is incoming, but are slowed down by close to two orders of magnitude ($\approx$ 6.0 ps), when the field points away from the interface. These asymmetries persist even in the highest investigated field. The fastest decay of dipole angle fluctuations is observed at incoming field strengths around 0.3 V nm^{-1} (Figure 6.8). These are the same fields that led to narrowest dipole angle distributions and negligible water polarization. On the wall where the field is outgoing, the distribution over the angle broadens and decay times increase monotonically with strengthening the field.

[a] Functions were fit to a sum of three exponentials.

Similar findings were observed by Debenedetti and co-workers, who investigated water in nanoscale confinement between β-cristobalite.[238] By introducing a linear scaling factor $k \in [0,1]$ for surface charges (which is similar to increasing an applied electric field), they observed a non-monotonic dependence of both orientational and translational dynamics on k, with a minimum in single molecule reorientation times found at 60 % surface charge. It is plausible that both the behavior illustrated in Figure 6.8 and that reported in ref. [238] have a common explanation in the balance between the angular bias due to hydrogen bonding and alignment by local electric field: In the state where the field balances the angle preferences of water, dipoles appear to orient least freely as angular distributions are narrowest for parallel dipole–wall orientations.

6.1.5 Dynamic Response to the Imposition of the Field

Figure 6.2 demonstrates that the response of the system to the strong electric fields employed in this study is non-linear, thus the decay of equilibrium fluctuations cannot be used to extract surface tension actuation rates from eq. (6.6). Therefore, the relaxation of $\cos\phi(t)$ to its equilibrium value after switching the field from off to on was investigated by calculating

$$R(t) = \frac{\overline{\delta\cos\,\phi(t)}}{\delta\cos\,\phi(0)}\,, \tag{6.7}$$

where $\delta\cos\phi(t)$ denotes the deviation of $\cos\phi(t)$ from its equilibrium value under the influence of the field and overlines are averages over 1800 independent non-equilibrium trajectories. Switching started at $t = 0$. Once again, characteristic time scales were estimated in terms of the times where $R(t) = 1/e$ (Figure 6.9).

As expected, asymmetries arise once more between incoming and outgoing fields. Characteristic time scales decrease from 0.7 ps under the influence of an incoming electric field of 0.3 V nm^{-1} to 0.3 ps at 0.9 V nm^{-1}. Under the influence of outgoing fields, these times are of the order of 0.1 ps and do not vary significantly with field strength. Note that these times are well below hydrogen bond lifetimes,[196,197,232,233] thus orientational reordering occurs without significant breaking of interfacial hy-

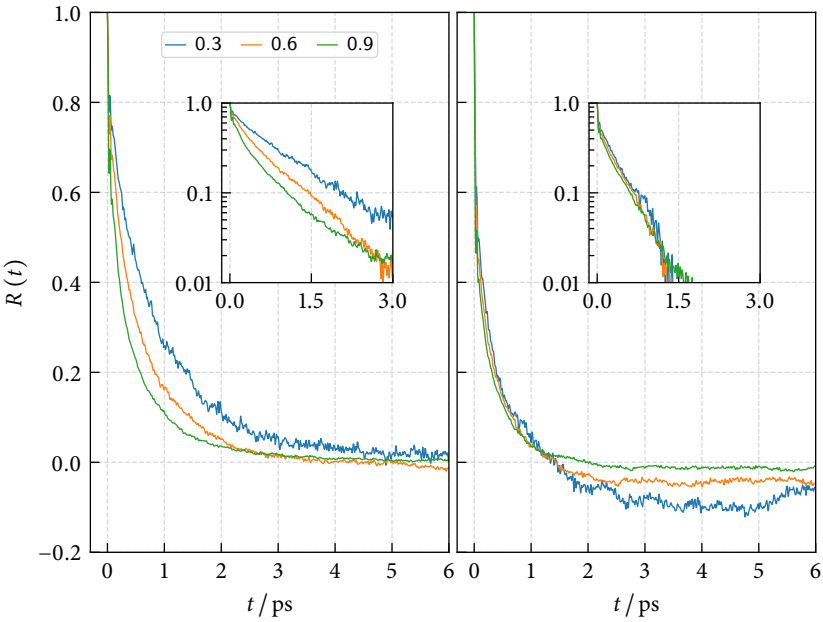


Figure 6.9: Relaxation of the polarization to its equilibrium value after switching the electric field from off to on. Lines correspond to field strengths of 0.3 V nm^{-1}, 0.6 V nm^{-1} and 0.9 V nm^{-1} Left: incoming fields; right: outgoing fields.

drogen bonds. This observation is consistent with the reported resilience of the hydrogen bond network, which has been found to be surprisingly robust with respect to dielectric polarization.[186,223–228]

What the above time scales do not reflect, however, is the long-time behavior of the relaxation functions. All curves for outgoing fields show a negative well starting at ≈ 1.5 ps and ending at ≈ 13 ps (data not shown). This behavior is likely related to the adjustment of local density to the change of the electric field, which is investigated in detail in Appendix S5. Here, only the main findings that are important to understand the effects in Figure 6.9 are summarized.

Density fluctuation dynamics in the first hydration layer was studied

by calculating the local density fluctuation functions

$$C_N(t) = \frac{\langle \delta N(t)\, \delta N(0) \rangle}{\langle [\delta N(0)]^2 \rangle}, \tag{6.8}$$

where $\delta N(t)$ denotes the deviation of the number of molecules within the first hydration layer N from its ensemble average $\langle N \rangle$. Density relaxes independently of the field direction within time scales of ≈ 1.2 ps, which is an order of magnitude slower than the response of polarization for outgoing fields (0.1 ps). When the electric field is switched on, polarization overshoots, as it initially relaxes at the lower density corresponding to the absence of the field. As the density increases toward its equilibrium value, polarization retracts more slowly. The density adjustment is almost complete after ≈ 10 ps, which coincides with the end of the negative well in Figure 6.9. The disparity between the time scales of orientational and density relaxations is most prominent in the relatively weak field $E = 0.3\,\mathrm{V\,nm^{-1}}$. Furthermore, saturation effects in both the polarization (Figure 6.2) and density (Figure 6.6) reduce the relative sensitivity in stronger fields. Hence the negative minimum is most pronounced in weaker fields.

The above mechanism has a smaller and opposite effect in incoming fields, which induce a density depletion that is somewhat smaller than the density rise in the outgoing field (Appendix S5). Since local density is decreased in this case, polarization adjusts first to the higher density corresponding to zero field. Once density decreases to its equilibrium value, polarization follows. Retardations of similar magnitudes as the negative wells in outgoing fields are expected, this time however, in the positive direction (consistent with the data in Figure 6.9).

6.1.6 Conclusion

On the nanoscale, electric fields can induce a Janus interface (i.e., opposing surfaces of different wetting behavior). Here, interfacial dynamics in such systems was investigated, which is an important criterion in the design of electro-switchable, nanofluidic devices and may be of relevance in understanding the function of biological voltage-gated channels.

By molecular dynamics simulation, a new and significant polarity dependence of interfacial polarization dynamics has been uncovered: In electric fields that balance the inherent angular preferences of water close to the interface, dipoles orient least freely and relaxation times go through a minimum. Such a competition exists only when the electric field is incoming. When the electric field is outgoing, both effects cooperate, resulting in monotonously increasing relaxation times; thus the asymmetric behavior at opposite confinement walls. Typical response times to the imposition of an electric field are in the range of 0.1–10.0 ps.

These results shed light on recent observations of a non-monotonic change in water reorientation rate as a function of increasing solute polarity.[222,238] The existing analogy suggests the possibility of tuning interfacial dynamics by optimizing the external field as an alternative to permanent modulation that can be achieved by controlled chemical modification.

6.2 Water Dynamics in Spherical Nanoconfinement

6.2.1 Motivation

In this chapter, water in spherical nanoconfinement will be investigated; a geometry that is of high relevance in nature and technology. Enzymatic reactions, for example, are controlled by few water molecules located in nanometer-sized protein cavities[239,240] and in material science, confined water can play an important role in metal-organic frameworks and many other systems.[241] As a consequence, a wealth of literature data is available, which shall be summarized shortly.

Rapid water flow has been observed in narrow carbon nanotubes and other hydrophobic pores, which form low-friction conduits for water transport.[240,242–244] Furthermore, prolonged hydrogen bond lifetime is associated with that flow in narrow pores, where only small threads of water can pass.[242]

Similar results were obtained close to planar hydrophobic interfaces, where parallel diffusion is exceptionally fast,[245] as well, but unlike in narrow confinement, interfacial hydrogen bond dynamics accelerates.[196] These results can be transferred to the air/water interface,[246,247] which may be considered as a special case of hydrophobic interfaces.

Drastically different results have been observed in hydrophilic confinement, notably inside silica nanopores,[233] polyoxomolybdate (POM) cages,[248,249] and reverse micelles.[201,250–254] The latter are particularly popular system choices in experimental studies, since reverse micelles encapsulate well-characterized nanoscopic pools of water, whose size can be easily controlled by the concentration ratio between water and surfactant. Common surfactants include Aerosol-OT (dioctyl sodium sulfosuccinate) and Igepal CO-520 (polyoxyethylene 5-nonylphenylether), which form ionic and non-ionic (but still highly polar) cavities, respectively.[251] Through computer simulations,[201,233,252] ultrafast infrared pump-probe spectroscopy and 2D vibrational echo experiments,[250,251,253,255] as well as small-angle X-ray scattering and dielectric relaxation spectroscopy,[254] the unanimous picture of hindered translational motions across the interface, decelerated reorientation dynamics, and prolonged hydrogen bond

lifetime has been established. These effects have been mostly attributed to the presence of hydrogen bonds between water and the interface.

Inside reverse micelles, a core/shell model model has been found to accurately describe many dynamic properties,[253,255,256] wherein the confined water molecules are divided into interfacial and bulk-like subensembles. Note that in these models, effects of the core ensemble diminish in smaller micelles ($\lesssim 1\,\mathrm{nm}$), where most water molecules are interfacial, and effects of the shell ensemble diminish in larger cavities ($\gtrsim 10\,\mathrm{nm}$), where most molecules are bulk-like.[256,a]

Kayal and Chandra performed a computational study similar to the present one.[257] The authors investigated water confined inside fullerenes of size C_{60} to C_{720}. From the structural point of view, they observed a transition from cage-like water structures in the smallest cavities to bulk-like conformations in the larger systems. Dynamic quantities such as orientational time correlation functions and hydrogen bond time correlation functions were found to depend on the fullerene diameter, with faster dynamics in smaller cavities. However, their hydrogen bond lifetimes should be interpreted with care, since the influence of diffusion was neglected in this study.[116,258] Furthermore, the occupation numbers used by the authors were significantly lower than those obtained from grand canonical Monte Carlo simulations (see computational details in the following section). The impact of confinement on system properties can thus be expected to be much weaker than in the present study.

From the above considerations it should have become clear that spherical confinement has many faces. Drastic differences can occur due to different surface composition (ionic, polar, hydrophobic), occupation number, and of course shape (tube, sphere) and diameter.

Here, MD simulations of water inside model fullerenes will be presented. These systems possess well-defined, almost spherical structures that are compatible with existing force fields for modified and unmodified graphene sheets and carbon nanotubes. They allow the construction of cavities of various sizes and surface characteristics. Focus was set on hydrogen bond kinetics, which is a major determinant of water's

[a] In terms of size, the fullerene model systems investigated in this work fall into the former category.

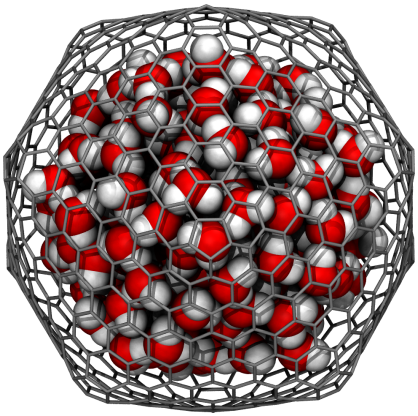


Figure 6.10: Snapshot of the C_{720}-I_h system filled with 145 water molecules (in hydrophilic confinement). Details on occupation numbers and interaction potentials are given in Tables 6.2 and 6.3.

properties.[21] In bulk water, the kinetics is known to couple to diffusion, since hydrogen bonds may only form after two molecules have approached each other, and may only break once they have diffused apart.[116,258] The correct treatment of diffusion in spherical symmetry is non-trivial, however, and is covered for the first time in this study. Two confinement–water interaction sets were chosen (a hydrophobic and a hydrophilic set), which shall shed light on the impact of the chemical nature of the confinement. Finally, realistic occupation numbers that lead to a vanishing pressure inside of the cavities were calculated in order to make meaningful comparisons with experimental setups. Computational details will be discussed in the next section.

6.2.2 Computational Details

The model systems employed in this study consist of SPC/E[184] water molecules (see Section 6.1.2) confined inside rigid structures that share their topology with the icosahedral fullerenes C_{320}-I_h, C_{500}-I_h, C_{720}-I_h, and C_{1500}-I_h (see Figure 6.10 for an example).

The confinement structures were generated by the program Fullerene with default parameters for all settings.[259–261] Initial structures were built from their canonical ring spiral pentagon indices[260] and subsequently geometry optimized using an extended harmonic oscillator force field,[259,261] which had been parameterized to reproduce the C_{60}-I_h ge-

Table 6.2: Maximum inner sphere radius r_{mis} and exact volume V_{exact} of the fullerenes, as well as water occupation numbers N_{water} and their standard deviations σ_N in differently parameterized (hydrophilic and hydrophobic, see text) systems. Geometrical parameters were calculated by the program Fullerene.[260,262]

Fullerene	r_{mis}/nm	V_{exact}/nm^3	$N_{\mathrm{water}}^{\mathrm{hydrophilic}}$	$N_{\mathrm{water}}^{\mathrm{hydrophobic}}$	σ_N
C_{320}-I_h	0.762	2.02	31	25	1
C_{500}-I_h	0.954	3.97	79	65	1
C_{720}-I_h	1.146	6.87	145	130	2
C_{1500}-I_h	1.641	20.73	501	473	3

ometry and thus may not lead to perfectly accurate structures of other fullerenes. However, it is not the purpose of this work to model real water–fullerene interactions. The generated structures serve only as a model for an atomistic, spherical nanoconfinement, thus deviations from the structure of real fullerenes are acceptable. For a more detailed discussion of fullerene structures and their construction, the reader is referred to the literature.[260,262] An overview of selected geometrical parameters and water occupation numbers in hydrophobically and hydrophilically parameterized model fullerenes (to be introduced shortly) is given in Table 6.2. Note that for notational simplicity, the suffix I_h will be dropped from now on.

The confinements shall be called fullerenes, even though the chosen water–confinement interactions do not actually represent those of real fullerene systems. A similar model system as described in Section 6.1.2 was chosen. The parameter sets are summarized in Table 6.3. Note that the hydrophilic confinements derive their hydrophilicity solely from enhanced Lennard-Jones interactions and not from the ability to form hydrogen bonds. In order to obtain the occupation numbers reported in Table 6.2, initially empty cavities were put and held in apparent equilibrium with a bulk water reservoir by means of grand canonical Monte Carlo simulations (GCMC). In these simulations, particle number fluctuations at constant volume V, temperature T, and chemical potential μ

Table 6.3: Lennard-Jones parameters σ_{CO} and ϵ_{CO} and the corresponding microscopic contact angles θ_m on a planar graphite-like surface, reported in ref. [180].

	σ_{CO}/nm	ϵ_{CO}/kJ mol^{-1}	θ_m/deg
hydrophilic	0.319	0.6270	29.4
hydrophobic	0.319	0.3762	101.2

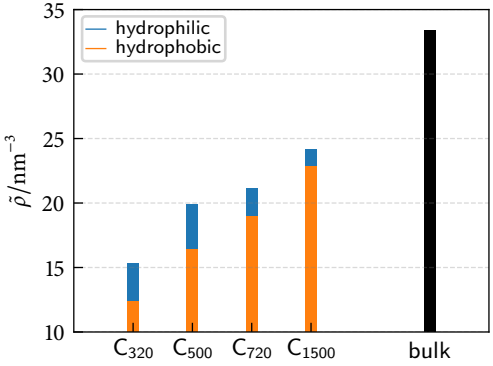


Figure 6.11: Average particle number density $\tilde{\rho} = N_{water}/V_{exact}$ in hydrophilic and hydrophobic confinement, as well as in a bulk water system at $T = 298$ K and $P = 101.325$ kPa.

are accounted for by trial insertions and deletions of molecules in such a way that the grand canonical probability density is sampled.[263] The method thus mimics the experimental setup of an adsorption experiment. The chemical potential of water was chosen to obtain a pressure $P = 101.325$ kPa at temperature $T = 298$ K in separate bulk phase simulations. All Monte Carlo simulations were performed by D. Bratko, using an in-house code. Methodological and algorithmic details were described elsewhere.[168,264] Note that this procedure differs from the one reported in the Master's thesis [24], where GCMC simulations were employed using a chemical potential that was too low.

To put occupation numbers into better perspective, average particle number densities $\tilde{\rho} = N_{water}/V_{exact}$ are visualized in Figure 6.11. The volume of the depletion layer between confinement boundary and water has not been subtracted from these densities, explaining the considerate reduction compared to bulk density. As expected, densities in hydrophilic

confinement are larger than in hydrophobic confinement. Both converge to the bulk limit with increasing fullerene size; in C_{500} (hydrophilic), however, an exceptionally large rise in density can be observed. This elevated density has a pronounced impact on the dynamic quantities discussed later.

Finally, MD simulations were performed in the microcanonical (NVE) ensemble, which permits a highly accurate description of dynamics, without bias by thermostats or barostats. This comes at the cost of sacrificing the convenience of setting up the simulation at a specified temperature, which allows easy comparisons between simulations of differently sized fullerenes and with experiments. To circumvent this problem, care was taken in setting up the initial velocities so that the average temperature after equilibration was within the range of (300 ± 3) K and that the total angular momentum of the system was zero. Since the angular momentum is a conserved quantity in non-periodic NVE simulations, it remains zero if the time integration is accurate enough.

The equations of motion were solved by Velocity-Verlet integration with a time step of 2 fs. Initial configurations were taken from the Monte Carlo simulations. Water molecules were kept rigid by the SHAKE[219] algorithm. Fullerene atoms were not propagated, remaining rigid and frozen.

All quantities except for diffusion were calculated from 5 ns trajectories with snapshots dumped every second time step. Transition counts between radial and angular bins needed for the calculation of diffusion constants were collected over 100 ns with snapshots dumped every 500th time step.[a] In these long simulations, an energy drift in the NVE ensemble is unavoidable (see Table 6.4). The chosen time step is a compromise between accurate time integration needed to keep the energy drift small and computational efficiency required to achieve such long simulation times. In other simulations of rigid water molecules, time steps of up to 4 fs have been used, while still yielding a sufficient accuracy.[184] Furthermore, the energy drift was minimized by refraining from cutting off Lennard-Jones interactions and from using long-range electrostatic solvers, such as the Ewald sum. Instead, the full Coulomb sum was calcu-

[a] Transition counts and their need will be introduced shortly.

Table 6.4: Average temperature $\langle T \rangle$ of the 100 ns trajectories, the corresponding standard deviations σ_T, and the energy drift ΔT, calculated from the difference between initial and final temperatures, which were determined by linear regression of the temperature time series.

		$\langle T \rangle$/K	σ_T/K	ΔT/K
hydrophilic	C_{320}	297.9	25.0	−3.1
	C_{500}	298.6	15.9	−3.7
	C_{720}	298.1	11.7	−3.5
	C_{1500}	297.7	6.4	−3.5
hydrophobic	C_{320}	300.6	28.4	−3.3
	C_{500}	300.7	17.8	−3.5
	C_{720}	298.5	12.3	−3.2
	C_{1500}	297.6	6.6	−3.6

lated, which is easily possible in non-periodic simulations. Unfortunately, this severely limits the efficiency of common techniques for speeding up MD simulations (such as neighbor lists) and puts an upper bound on the maximum size of the system, if reasonable computational times are aimed at. Thus, no systems larger than C_{1500} were chosen in this study.

All MD simulations were performed using the Lammps molecular dynamics software.[217]

6.2.3 Radial Structure

As expected, packing effects, geometrical restrictions, and specific interactions contribute to a non-uniform potential of mean force within the cavities, that leads to a distinct layering of molecules. The corresponding radial density profiles are shown in Figure 6.12. In principle, such effects are well known from planar interfaces.[156] However, some peculiarities arise due to the spherical nature of the cavities, as explained below. Most notably, one should be aware that the layers shown in the density profiles actually correspond to spherical shells, a geometric feature that is often overseen in one-dimensional representations, such as Figure 6.12. The following observations can be made:

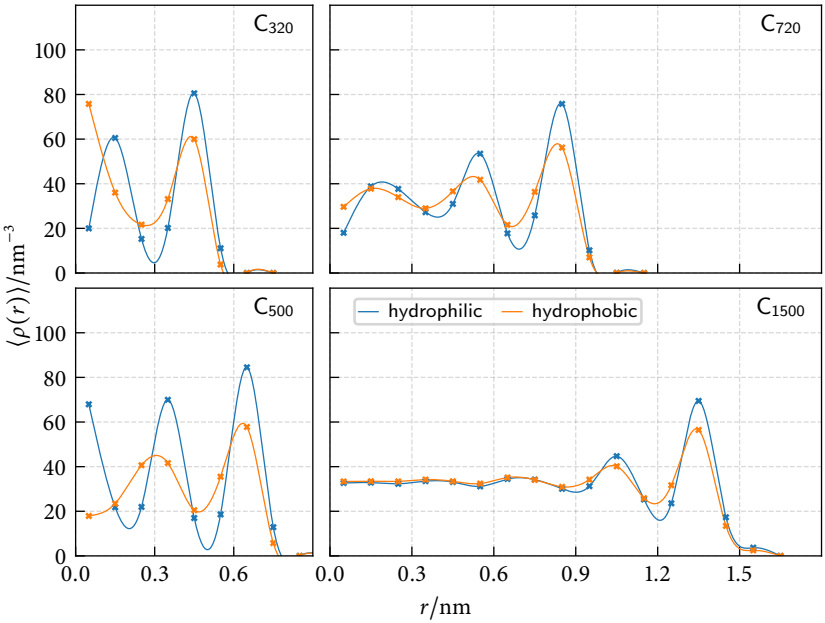


Figure 6.12: Average oxygen number density $\langle \rho(r) \rangle$ as a function of the radial distance r from the fullerene center. Lines are meant to guide the eye.

1. The interfacial density (outermost density peak) is consistently higher in hydrophilic confinement, in which the enhanced attractiveness of the cavity and the elevated occupation numbers induce a much more structured water droplet.

2. Within the core regions of the cavities, layering levels off to bulk density (33.4 nm^{-1}) with increasing fullerene size. However, due to the cubic dependence of the volume on radii, the vast majority of water molecules can still be considered to be interfacial. As an example, even in C_{1500} (hydrophobic), 54 % of all water molecules reside in the outermost density layer. Furthermore, slight oscillations in local density persist, even in the core region of C_{1500}.

3. Due to the slightly non-spherical nature of the fullerenes, density

profiles do not drop to zero abruptly at the confinement boundaries, but may exhibit a small shoulder at large radii.

4. Two systems (C_{320} hydrophobic and C_{500} hydrophilic) exhibit a large density peak in the center of the cavity. These peaks originate from the high binning resolution ($10\,\mathrm{nm}^{-1}$), which leads to bin volumes that are smaller than a typical water molecule. Thus, if a water molecules resides favorably in the center of the cavity, the calculated local density will be much higher than expected due to the small binning volume.

In turns of radial density profiles, these results do not differ much from those reported in earlier studies on water/fullerene systems.[257]

While the local density fluctuates quite remarkably, even in the center of the cavities, water's hydrogen bond network is hardly disturbed beyond the interfacial region. In order to study the network, local hydrogen bond numbers and tetrahedral order parameters were calculated. The latter were split into an angular (S_g) and distance (S_k) part, according to the definition[265] of Chau and Hardwick:

$$S_g = \frac{3}{32} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos \Psi_{j,k} + \frac{1}{3} \right)^2 \tag{6.9}$$

$$S_k = \frac{1}{3} \sum_{k=1}^{4} \frac{(r_k - \hat{r})^2}{4\hat{r}^2} \,. \tag{6.10}$$

Therein, the summation indices run over the four closest neighbors of a selected water molecule, $\Psi_{j,k}$ denotes the angle between the central molecule and its neighbors j and k, r_k is the distance to the closest neighbor k, and $\hat{r}$ is the arithmetic mean of the four radial distances. Both values are zero for a perfect tetrahedron, but approach somewhat larger values ($S_g \approx 0.1$, $S_k \approx 1.3 \times 10^{-3}$) in bulk water at ambient conditions.[265] The parameters are shown in Figures 6.13 and 6.14 and their discussion follows shortly.

Hydrogen bonds were detected by the following geometric criterion,

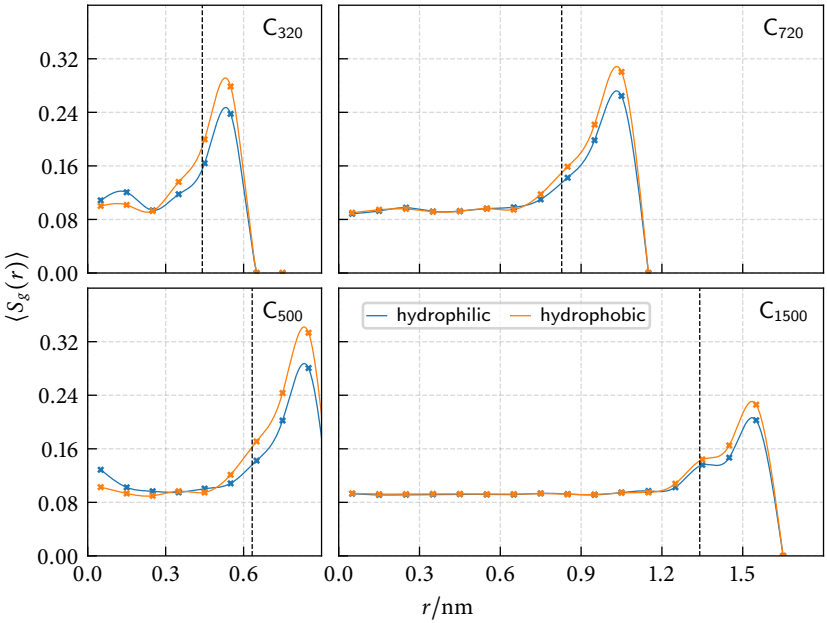


Figure 6.13: Average angular part $\langle S_g(r) \rangle$ of the tetrahedral order parameter[265] as a function of the radial distance r from the fullerene center. Lines are meant to guide the eye. The dashed vertical line indicates the position of the interfacial density maximum.

which is common in the discussion of hydrogen bond kinetics:[258]

$$h(R, r, \phi) = \theta(R_c - R)\theta(r_c - r)\theta(\phi_c - \phi)\,. \tag{6.11}$$

Therein, R denotes the O···O separation of the tagged pair, r is the O···H intermolecular separation, ϕ is the angle between the O···O vector and the covalent OH-bond, and the index c denotes the respective cutoff value. $\theta(x)$ is the Heaviside step function. The preceding equation defines a binary operator, the hydrogen bond population operator h, which equals one if a particular tagged pair of water molecules is hydrogen bonded and zero otherwise. Its average over all possible pairs of water molecules and time steps $\langle h \rangle$ is closely related to the average number of hydrogen

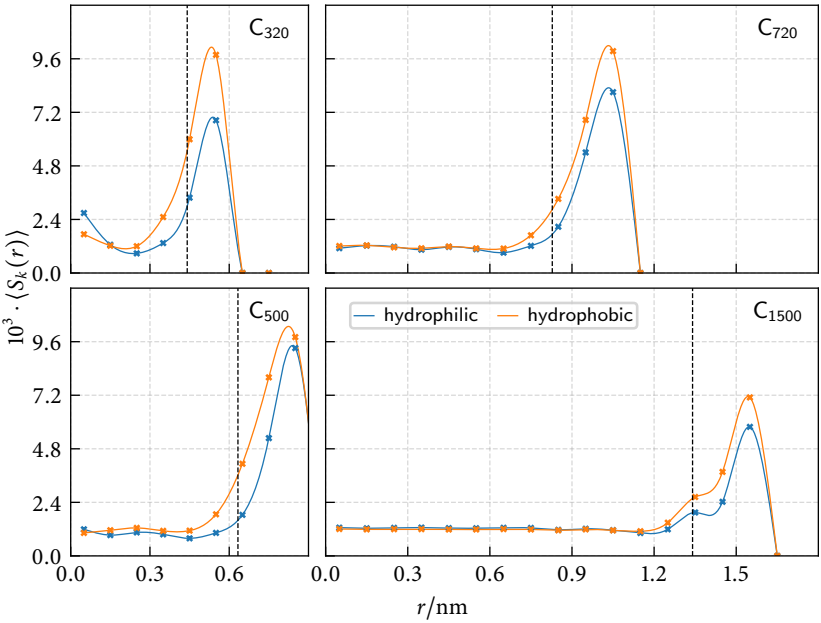


Figure 6.14: Average of the distance part $\langle S_k(r) \rangle$ of the tetrahedral order parameter[265] as a function of the radial distance r from the fullerene center. Lines are meant to guide the eye. The dashed vertical line indicates the position of the interfacial density maximum.

bonds per water molecule $\langle n_{\mathrm{hb}} \rangle$, according to

$$\langle n_{\mathrm{hb}} \rangle = \frac{N-1}{2} \langle h \rangle \,. \tag{6.12}$$

The required cutoff values were set to $R_c = 350\,\mathrm{pm}$, which corresponds to the first minimum in the oxygen–oxygen radial pair distribution function of water, $r_c = 2.45\,\mathrm{pm}$ which is the first minimum in the oxygen–hydrogen radial pair distribution function, and $\phi_c = 30°$, a value at which the average number of hydrogen bonds is close to its asymptotic value for large ϕ_c.[258] Distance dependent profiles of the average number of hydrogen bonds are shown in Figure 6.15.

All distance dependent profiles shown in Figures 6.13 to 6.15 probe

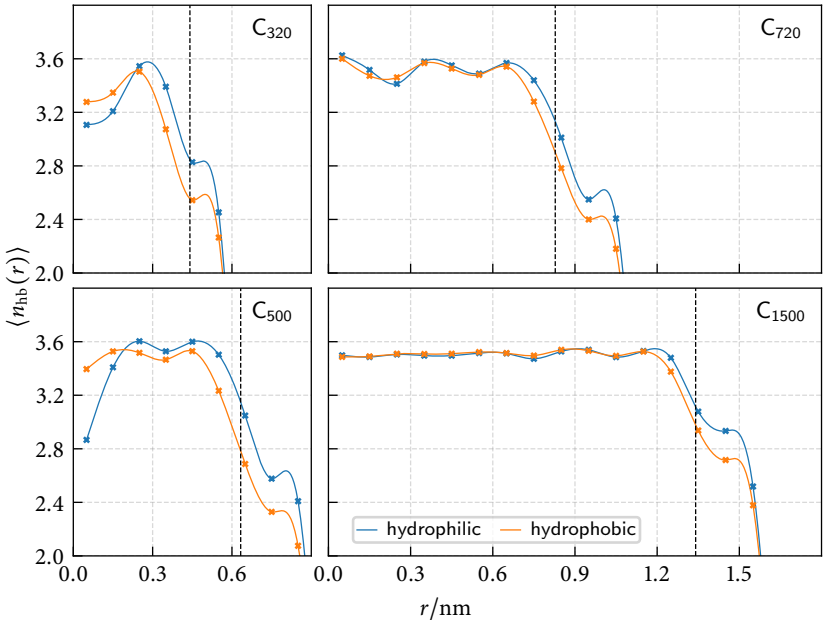


Figure 6.15: Average number of hydrogen bonds per water molecule $\langle n_{\mathrm{hb}}(r)\rangle$ as a function of the radial distance r from the fullerene center. Lines are meant to guide the eye. The dashed vertical line indicates the position of the interfacial density maximum.

water's hydrogen bond network in one way or another, and thus, share a number of common features:

1. The hydrogen bond network is being significantly disturbed by the interface. The average number of hydrogen bonds is almost halved within the outermost water shell, which is accompanied by significant distortion of tetrahedrality. These effects are more pronounced in hydrophobic than in hydrophilic environment.

2. Water molecules that reside beneath the outermost water shell are not hindered by the geometric restrictions anymore and all parameters normalize quickly. Exceptions are both C_{320} systems, which are presumably simply too small to build up a fully hydrogen

bonded, tetrahedral network, and C_{500} (hydrophilic) which shows a significant drop in hydrogen bond numbers and an associated distortion of the tetrahedral environment in the center of the cavity, even though a bulk-like hydrogen bond network was possible at intermediate distances. It is noteworthy that the same system exhibits a highly elevated central density peak (see Figure 6.12). The picture that emerges for this system is that the particular choice of system size (geometry of the cavity and occupation number) forces a single water molecule into the core region of the confinement, where it is unable, though, to fit into its preferred fully hydrogen bonded, tetrahedral environment.

From a structural point of view, all investigated properties (density profiles, order parameters, and hydrogen bond numbers) support the concept of a core/shell model,[255,256] wherein water molecules are divided into subensembles that are bulk-like and interfacial. However, the bulk-like phase occupies only a relatively minor fraction of the total system volume inside the small systems investigated here, so that most water molecules can be considered to be interfacial. A cage-like structure, such as that reported for $(H_2O)_{12}C_{180}$ could not be observed,[257] but such a small system was not subject of the present study.

We shall now turn to the analysis of dynamic properties. To get a first grasp of hydrogen bond kinetics, we shall start by looking at reorientation dynamics and diffusion, since both are tightly connected to the hydrogen bond network of water and its dynamics.

6.2.4 Reorientation Dynamics

Studying single-molecule reorientation is a popular approach to gain first insight into hydrogen bond dynamics; the rationale being that at some point, reorientation of a single molecule requires hydrogen bond breaking and formation. Reorientation dynamics is typically probed by time correlation functions of the type

$$C_n(t) = \langle P_n(\mathbf{u}(t)\mathbf{u}(0)) \rangle \,, \tag{6.13}$$

where $P_n(x)$ is the n-th order Legendre polynomial and $\mathbf{u}(t)$ is a unit vector describing molecular orientation. In this work, second-order orientational time correlation functions $C_2(t)$ were calculated, where the unit vector $\mathbf{u}(t)$ was set in direction of an OH bond of a water molecule.

Calculating these functions is particularly popular, since they are experimentally available from the anisotropy decay measured in infrared (IR) pump-probe spectroscopy.[266–268] This technique has the necessary time resolution to resolve even the fastest molecular motions contributing to reorientation (in the case of water, librational motions around the hydrogen bond axis) and can be made surface sensitive if coupled with appropriate methodology.[269,270] Further information about reorientation dynamics is available from nuclear magnetic resonance (NMR) spectroscopy,[271–273] quasi-elastic neutron scattering (QENS),[22,274] as well as dielectric relaxation, Terahertz and optical Kerr-effect (OKE) spectroscopy.[275–277]

Before turning back to the discussion of $C_2(t)$ in spherical confinement, reorientation dynamics in bulk water shall be briefly described. The most basic information to be extracted from $C_2(t)$ is the characteristic time of second-order reorientation

$$\tau_2^{\text{reor}} = \int_0^\infty C_2(t)\, dt\,, \tag{6.14}$$

which equals 1.7 ps for SPC/E water and varies between 1.7 ps and 2.6 ps in experiments.[278–282] While sufficient for many purposes, exact quantification of the processes contributing to reorientation requires additional mechanistic insight.

On sub-picosecond time scales, water molecules reorient rapidly, but are soon hindered by hydrogen bonds to nearest neighbors which exert restoring torques around the preferred hydrogen bond geometry.[268] These librational motions cause a rapid decay of $C_2(t < t_{\text{trans}})$ for times within the transient regime $t_{\text{trans}} \approx 200\,\text{fs}$. Afterwards, $C_2(t)$ decays mono-exponentially with a time constant varying between 1.9 ps and 3.0 ps, depending on experiment or computational model.[30,281–285]

If a time correlation function decays mono-exponentially, the process leading to that decay can be assumed to be Poissonian (i.e., events occur with a constant rate and independently of the time since the last event).

The rate of a Poisson process is equal to the exponential decay constant and can be easily extracted from the time correlation function by fitting $C_2(t > t_{\text{trans}})$ to an exponential function. While such rates give information about the time required for full reorientation, they do not provide information about the lifetime of a hydrogen bond, unless paired with appropriate phenomenology.

Since Debye's seminal work, rotational diffusion (i.e, slow and stepwise reorientation) has been assumed to be the mechanism describing reorientation in liquid water.[286] However, this picture has recently been challenged and a large-amplitude jump model has been shown to fit better to experimental and simulation data.[281] From the extended-jump model of Laage and Hynes, jump times can be extracted, which measure the time between jumps of a hydrogen atom from one stable hydrogen bond acceptor to another. For bulk SPC/E water, $\tau^{\text{jump}} = 3.3\,\text{ps}$ was reported.[281,282] While such jump times provide certain insight into the dynamics of hydrogen bonds, a jump requires the approach of a new partner and departure of another one. The complete hydrogen bond network dynamics is a collective phenomenon, that implies second-shell reorientation (and beyond) and is tightly coupled to diffusion.[116,258] In order to get true hydrogen bond lifetimes, the phenomenology presented in Section 6.2.6 is better suited.

Nevertheless, reorientation dynamics is ideal to get local insight and has been widely studied in literature, which permits validating results and putting them into a broader literature context. For our purpose, however, it will not be necessary to extract accurate jump times from the extended jump formalism. Characteristic time scales τ_2^{reor} are sufficient to make comparisons between differently sized and functionalized cavities and are reported here instead.

The second-order orientational time correlation functions $C_2(t)$ in spherical confinement are shown in Figure 6.16. In these cavities, $C_2(t)$ does not decay mono-exponentially beyond the transient regime (it is not a straight line on a semi-logarithmic scale). Recently, progress has been made in understanding this non-exponential behavior close to extended interfaces.[222] Next to such an interface, different OH bond populations must be considered; those that are hydrogen bonded and those that are dangling, because water is unable to maintain its full hydrogen bond

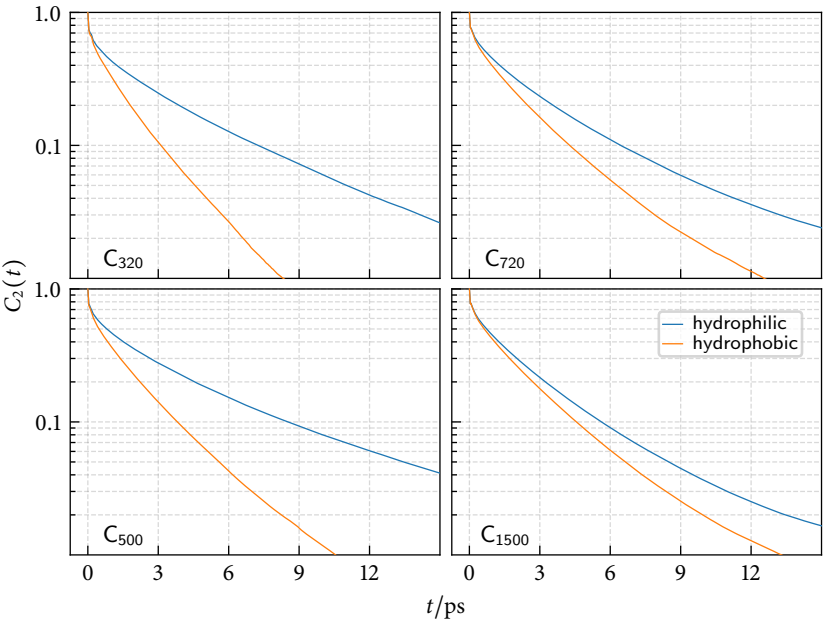


Figure 6.16: Second-order orientational time correlation functions $C_2(t)$.

network close to the interface (compare Figure 6.15). The OH bonds participating in a hydrogen bond must be further divided into those that point into the bulk and those that are tangential. These three different populations have been shown to reorient markedly different with jump times to a new stable hydrogen bond increasing in the order dangling < bulk < tangential. This interface-induced anisotropy causes the non-exponential decay of $C_2(t)$. Water molecules beneath the solvation layer reorient essentially bulk-like.

Inside the spherical cavities employed in this work, similar effects are expected to influence reorientation dynamics. A core/shell model could be invoked to describe the contribution of the small bulk-like phase to the global orientational time correlation function, and interfacial water molecules could be further divided into subensembles, such as those mentioned above. However, we shall refrain from mechanistic

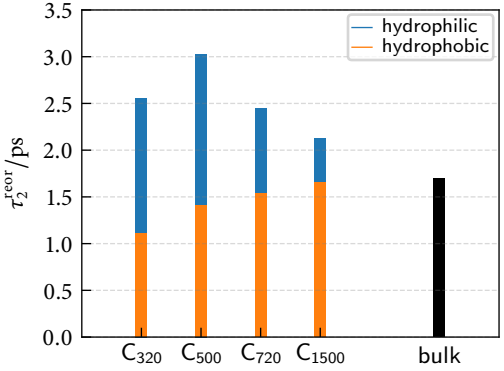


Figure 6.17:
The characteristic times of second-order reorientation τ_2^{reor} of nanoconfined water in different fullerenes and in the bulk.

interpretation here and focus on characteristic time scales τ_2^{reor}, instead. Like in Section 6.1, these were calculated by analytic continuation and subsequent integration of the time correlation functions (details are given in Appendix S6), and they are visualized in Figure 6.17.

Both in hydrophilic and hydrophobic confinement, τ_2^{reor} seems to converge to the bulk limit[282] of 1.7 ps with increasing cavity size. Since most molecules are still interfacial, even in C_{1500}, full convergence should not be expected. While there is a monotonic decrease in characteristic time scales with decreasing fullerene diameter in hydrophobic confinement, reorientation slows down non-monotonically inside hydrophilic cavities. From the data at hand, it is not clear whether this non-monotonic slow-down in hydrophilic confinement is real, or whether only C_{500} (hydrophilic) exhibits exceptionally slow reorientation, which could be well correlated to the elevated density in this system (Figure 6.11).

Nevertheless, this data demonstrates that spherical confinement has a marked influence on water dynamics that goes beyond what has been observed near extended interfaces. Results fit well into the literature context given in the introduction, except that in the present systems, decelerated reorientation dynamics cannot be attributed to hydrogen bonds between water and the interface, because, here, no such bonds exist. The effects are well correlated with density/occupation numbers on the other hand.

6.2.5 Diffusion

From an atomistic point of view, (self-)diffusion can be considered as the result of a random walk of a single particle in the mean field of all others. It is an attempt to describe the complex many-particle dynamics as an average, single-particle motion. A priori, there is no reason to believe that a diffusive picture should hold in small spherical cavities. The validity of such a picture will be investigated in this section.

Spherical cavities break the homogeneity and isotropy of the confined liquid, rendering conventional approaches to diffusion coefficients useless (i.e., fitting the mean-squared displacement as a function of time). Instead of a single diffusion coefficient D_0 that describes the motions of water molecules within the averaged potential exerted by their surroundings, an anisotropic and position dependent diffusion tensor $\mathbf{D}(\mathbf{r})$ must be employed. Spherical symmetry implies that diffusion in radial direction (orthogonal to the interface, denoted by $\perp$) differs from diffusion in angular direction (parallel to the interface, denoted by $\parallel$) and both depend on the radial position within the cavity.

The problem of calculating such diffusion tensors is conceptually challenging, since it requires the solution of the full Smoluchowsky equation,

$$\frac{\partial p(\mathbf{r},t)}{\partial t} = \nabla \cdot \left[\mathbf{D}(\mathbf{r}) e^{\beta F(\mathbf{r})} \nabla \left(e^{-\beta F(\mathbf{r})} p(\mathbf{r},t) \right) \right], \tag{6.15}$$

where $p(\mathbf{r},t)$ is the spatial probability density of a tracer particle as a function of time and $F(\mathbf{r})$ is the position dependent free energy. Fortunately, the problem is formally equivalent to pair diffusion, which has received significant theoretical attention.[287]

In this work, the approach of Hummer,[287,288] to self-consistently estimate local diffusion coefficients and free energies from simulation data, is being followed. The method relies on separation of the Smoluchowsky equation into a set of uncoupled, one-dimensional differential equations and subsequent solution by spatial discretization and inference of model parameters from simulation data through Bayesian inference,[288,289] as outlined in the following.

The diffusion tensor $\mathbf{D}(\mathbf{r})$ shows spherical symmetry in the investi-

gated cavities, that is,

$$\mathbf{D}(\mathbf{r}) = \begin{pmatrix} D_\perp(r) & 0 & 0 \\ 0 & D_\parallel(r) & 0 \\ 0 & 0 & D_\parallel(r) \end{pmatrix}, \tag{6.16}$$

in spherical coordinates r, θ, ϕ. The Smoluchowsky equation (eq. (6.15)) in such a coordinate system can be solved by means of Green's functions $P(r, \cos\theta, t|r', 0) = r^2 p(r, \theta, t|r', 0, 0)$, which measure the conditional probability of finding a tracer particle in the intervals $(r, r + dr)$ and $(\cos\theta, \cos\theta + d\cos\theta)$ starting from r' and $\theta = 0$ at time 0.

As demonstrated by Mittal and Hummer,[287] the diffusion equation for these functions becomes

$$\begin{aligned} \frac{\partial}{\partial t}P = \frac{\partial}{\partial r}\left\{D_\perp(r)\left[\beta V'(r) + \frac{\partial}{\partial r}\right]P\right\} \\ + \frac{D_\parallel(r)}{r^2}\frac{\partial}{\partial \cos\theta}\left[\left(1-\cos^2\theta\right)\frac{\partial}{\partial \cos\theta}P\right], \end{aligned} \tag{6.17}$$

where $V(r) = F(r) - 2kT\ln r$ and $V' = dV(r)/dr$. It can be further reduced by integration over the angular part, leading to

$$\frac{\partial}{\partial r}G = \frac{\partial}{\partial r}\left[D_\perp(r)\left(\beta V'(r) + \frac{\partial}{\partial r}G\right)\right], \tag{6.18}$$

wherein the radial propagator $G(r, t|r', 0) = \int_{-1}^{1} P(r, \cos\theta, t|r', 0)\, d\cos\theta$ measures the conditional probability of finding a tracer particle in the interval $(r, r + dr)$ at time t, starting from r' at time 0, irrespective of angular motion. The preceding equation permits the treatment of radial diffusion as a standard one-dimensional problem.

After realizing that the bottom term in eq. (6.17) corresponds to the angular momentum operator in quantum mechanics, the authors were also able to treat angular diffusion by a set of uncoupled one-dimensional equations, which are obtained after expanding the Green's function

$P(r, \cos\theta, t|r', 0)$ in Legendre polynomials $p_l(x)$:

$$P\big(r, \cos\theta, t\big|r', 0\big) = \sum_{l=0}^{\infty} \frac{2l+1}{2} p_l(\cos\theta) Q_l\big(r, t\big|r', 0\big)\,. \tag{6.19}$$

The associated evolution equations for each of the Q_l are given by:

$$\begin{aligned} \frac{\partial}{\partial t} Q_l = {} & \frac{\partial}{\partial r}\left\{ D_\perp(r)\left[\beta V'(r) + \frac{\partial}{\partial r} Q_l\right]\right\} \\ & - \frac{D_\parallel}{r^2} l(l+1) Q_l\,, \end{aligned} \tag{6.20}$$

which are equal to eq. (6.18) for $l = 0$ and differ therefrom by sink terms otherwise.

Note that for practical reasons, eq. (6.19) has to be truncated after a finite number of terms in computer codes. In this project, the sum was terminated at $l_{\max} = 30$. No significant changes in results could be observed by taking terms of higher order into account.

In the following, the algorithm to calculate radial diffusion coefficients $D_\perp(r)$ shall be outlined. To do so, simulation trajectories were discretized by assigning radial positions r into corresponding bins i along r, and then counting the number of events N_{ji} when a particular water molecule was in bin i at some time t during the simulation and in bin j at time $t + \Delta t$ later. In this work, bins of equal radial width $\Delta r = 500\,\text{pm}$ were chosen. Within the long (100 ns) equilibrium trajectories, transition counts are symmetric (i.e., $N_{ij} = N_{ji}$), as demanded by microscopic time reversibility.

Care must be taken in choosing the lag time Δt, so that the influence of the free energy surface underlying the particle dynamics has already been felt. In the present simulations, lag times of 1 ps, 2 ps, 5 ps and 10 ps were used.

After counting the transitions, radial diffusion coefficients $D_\perp(r_i)$ and free energies $V(r_i)$[a] that are consistent with the observed data were estimated by a Bayesian inference approach.[288,289] In the Bayesian formalism,

[a] Evaluated at the position r_i of the bin center for $D_\perp$ and at the position of the bin edges for V.

a posterior distribution $p(\text{parameters}|\text{data})$ of the model parameters is constructed from the simulation data through Bayes's theorem

$$p(\text{parameters}|\text{data}) \propto p(\text{data}|\text{parameters})p(\text{parameters}), \qquad (6.21)$$

wherein $p(\text{parameters})$ is the prior distribution of parameters, which is assumed to be uniform, and $p(\text{data}|\text{parameters})$ is given by a likelihood function, which is a product of Green's functions expressed in terms of matrix exponentials.[288] To construct posterior distributions, parameters were sampled using Metropolis Monte Carlo simulations in parameter space. For technical details, the reader is referred to refs. [73, 289].

Once $D_\perp(r)$ and $V(r)$ have been determined, angular diffusion coefficients $D_\parallel(r)$ can be estimated in a similar fashion. Here, the numbers $N_{j\alpha,i}$ for transitions of a particle from radial bin i to radial bin j and angular bin α have to be counted. Angular bins were indexed according to the cosine of the azimuthal angle

$$\cos\theta(t) = \frac{\mathbf{r}(t)\cdot\mathbf{r}(0)}{r(t)r(0)}, \qquad (6.22)$$

with $\theta(0) = 0$ by definition of the coordinate system and without loss of generality. Here, the azimuthal angle was divided into 50 bins of equal width. A second Bayesian inference approach can be devised as outlined above to find the set of angular diffusion coefficients $D_\parallel(r_i)$ that is most consistent with the observed simulation data.

To enforce a certain degree of smoothness on the resulting free energy and diffusion profiles, they were treated internally as a set of continuous cubic splines, with the number of spline nodes set to the number of distinct features (maxima + minima) seen in the radial density profiles. Finally, a procedure similar to shifting the origin of time in free diffusion (i.e., fitting the mean-square displacement to $6D_0(t+\tau)$, with τ being the origin of time) was employed to account for the initial non-diffusive spread of molecules due to fast molecular motions. Therefore, transition counts were collected at several different lag times and the origin of time was optimized during the MC parameter sampling, as well.

Python codes performing the Bayesian sampling procedure were pro-

vided by Gerhard Hummer and used with modifications to account for different spline-node resolutions for all quantities listed below.

A major advantage of the procedure described above is that free energy profiles and diffusion constants are obtained simultaneously from the same set of simulation data, without making a priori assumptions about their functional form. Naturally, free energy profiles are also directly accessible from counting statistics

$$V(r) = -kT \ln p(r)\,, \tag{6.23}$$

where $p(r)$ is the local probability density of finding a particle at position r. The preceding equation can be evaluated by discretization and results are shown in comparison to those obtained from the Bayesian sampling procedure in Appendix S6. Both are in excellent agreement, lending first credibility to the diffusion model (in radial direction).

The position dependent radial and angular diffusion constants are shown in Figures 6.18 and 6.19 and the following observations can be made:

1. Radial diffusion coefficients $D_{\perp}(r)$ are correlated with free energy profiles, that is, particles residing at energetically favorable positions are less likely to move away from there (in radial direction) and vice versa.

2. Angular diffusion coefficients $D_{\parallel}(r)$ are anti-correlated with free energy profiles, that is, angular motion is enhanced at positions of increased density.

3. With increasing cavity size, a convergence of both diffusion tensor components can be observed for $r \to 0$. This limiting value is around $2\,\mathrm{nm^2\,ns^{-1}}$, which is somewhat smaller than D_0 of bulk SPC/E water at ambient conditions ($D_0 = 2.97\,\mathrm{nm^2\,ns^{-1}}$).[290] The values reported here are expected to be smaller than D_0 because of the known system size dependence of diffusion coefficients.[291]

4. Overall, diffusion coefficients are larger in hydrophobic confinement. These effects are particularly pronounced for angular diffu-

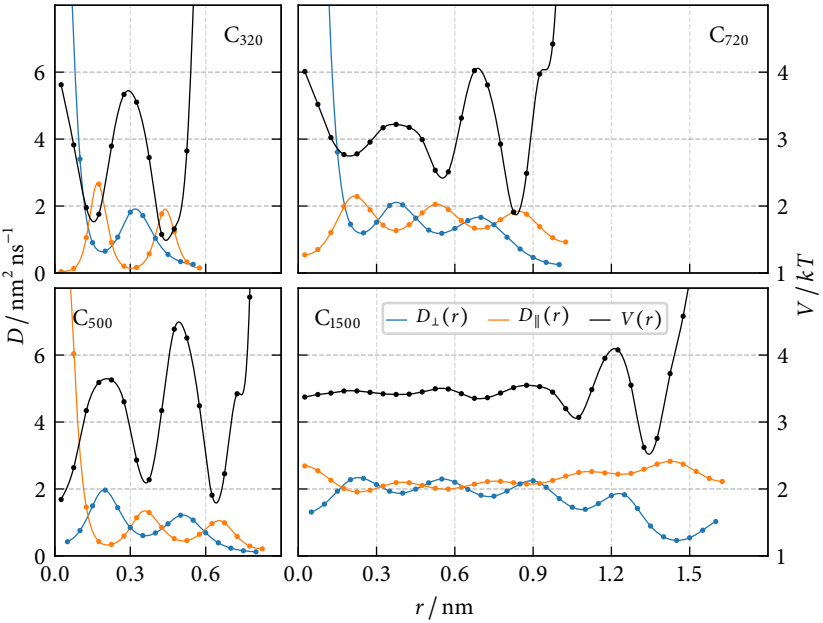


Figure 6.18: Position dependent diffusion constants $D_\perp(r)$ and $D_\parallel(r)$, as well as free energy profiles $V(r)$ as a function of the distance r from the fullerene center in hydrophilic confinement. Lines show the model-intrinsic spline representation of the discrete data.

sion coefficients, which are considerably elevated close to hydrophobic interfaces.

While the above-mentioned observations are expected, or can be easily rationalized, some other features are more surprising. We shall start with the discussion of radial diffusion coefficients. Some profiles (hydrophobic: C_{720}; hydrophilic: C_{320}, C_{720}) show extreme maxima in the center of the cavities. All of these systems exhibit a free energy barrier at $r \to 0$, which drives particles outwards, thus it is plausible that diffusion spikes at these positions. However, at such small radii, it is generally hard to accurately resolve radial motions, and transition counts may be subject to considerate noise. In order to judge the quality of the proposed model, the propagators of the radial diffusion model $G(r, t|r', 0)$ are plotted

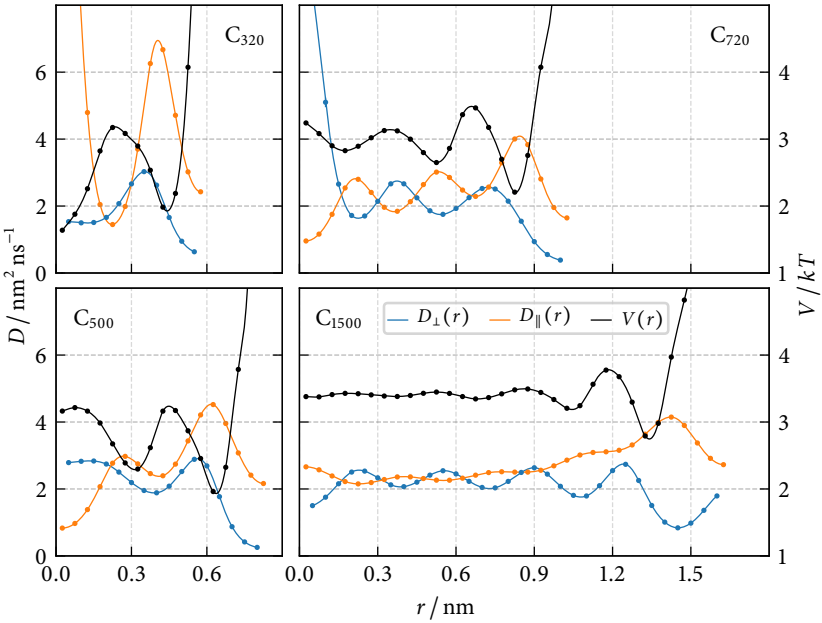


Figure 6.19: Position dependent diffusion constants $D_\perp(r)$ and $D_\parallel(r)$, as well as free energy profiles $V(r)$ as a function of the distance r from the fullerene center in hydrophobic confinement. Lines show the model-intrinsic spline representation of the discrete data.

in Figure 6.20 for different values of r' and compared to actual transition probabilities observed during the simulation. Results are shown for hydrophobic confinement; propagators in hydrophilic confinement are qualitatively similar and are shown in Appendix S6. These figures give the probability of finding a particle that originated from position r' (indicated by the different colors) at a position r, after the chosen lag time of 10 ps. In radial direction, perfect agreement between model propagators and simulation data can be observed, demonstrating that (a) a diffusive model provides a realistic description of the single-particle dynamics in the investigated systems and (b) the proposed model parameters are quantitatively accurate.

We shall now turn to angular diffusion. Just as for radial diffusion,

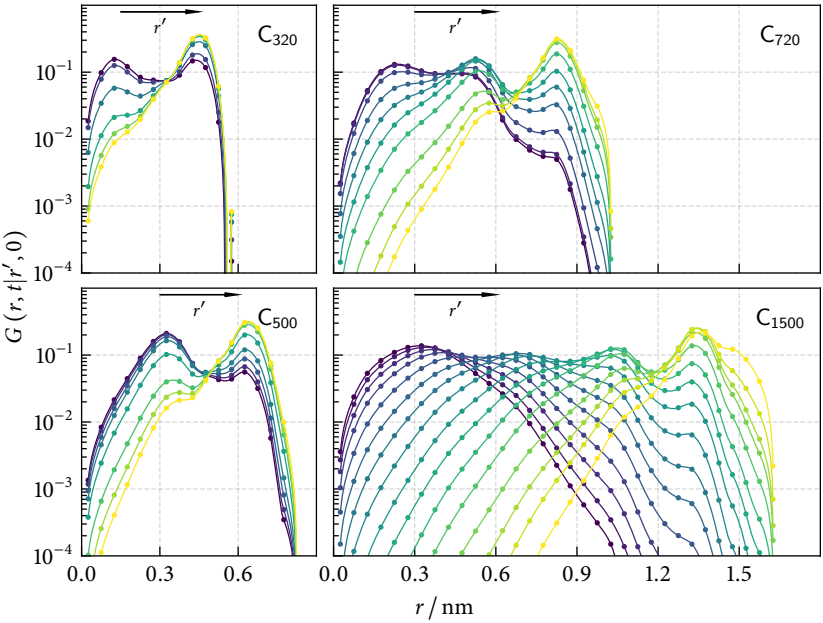


Figure 6.20: Model propagators $G(r, t|r', 0)$ (solid lines) compared to simulation data (symbols) in hydrophobic confinement for a lag time of 10 ps. The color gradient indicates rising r'. Lines show the model-intrinsic spline representation of the discrete data. Only every second bin is shown for clarity's sake (i.e., curves correspond to $r' = 0.025\,\text{nm}, 0.125\,\text{nm}, 0.225\,\text{nm}, \ldots$).

some systems (hydrophobic: C_{320}; hydrophilic: C_{500}) show highly elevated angular diffusion coefficients for $r \to 0$. These maxima exist, if the system exhibits a free energy minimum keeping particles in the center, thus a similar rationale as given above applies. In order to judge the quality of the angular diffusion model, propagators $P(r, \cos\theta, t|r', 0)$ were calculated and compared to transition probabilities observed during the simulation. As an example, the propagators $P(r, \cos\theta, t|r', 0)$ are shown for $r' = 0.175\,\text{nm}$ in Figure 6.21. This plot gives the probability of observing an angular motion of a particle, characterized by the angle of motion $\cos\theta$ and simultaneous transition from radial position r' to r during the

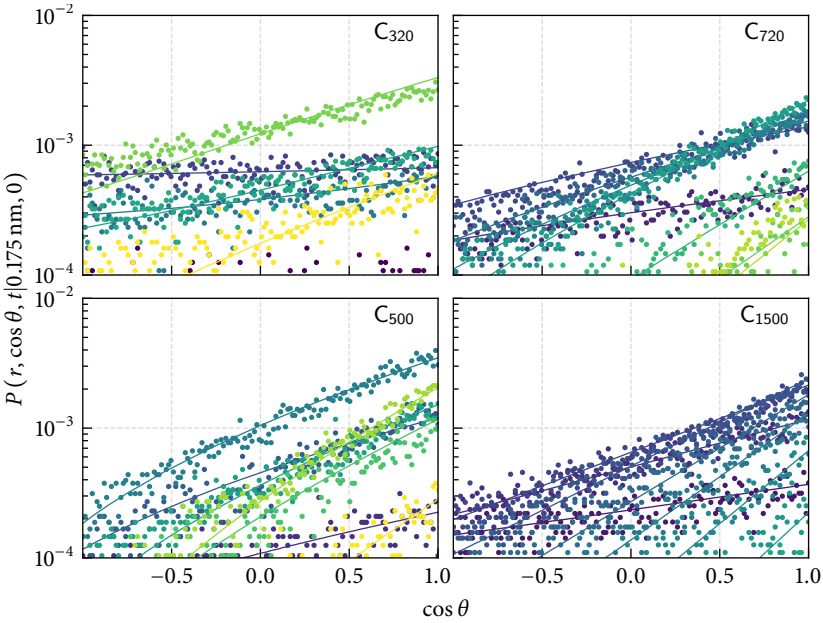


Figure 6.21: Model propagators $P(r, \cos\theta, t|r', 0)$ (solid lines) compared to simulation data (symbols) in hydrophobic confinement for a lag time of 10 ps and $r' = 0.175$ nm. The color gradient (blue → green → yellow) indicates rising r. Lines show the model-intrinsic spline representation of the discrete data. Only every second bin is shown for clarity's sake (i.e., curves correspond to $r = 0.025$ nm, 0.125 nm, 0.225 nm, . . .).

chosen lag time of 10 ps. Further examples are given in Appendix S6. The complete set of propagators was visually inspected but is not shown here, for brevity's sake. The following trends and observations can be made:

1. Observed transition probabilities are much more scattered around model propagators in angular direction, than in radial direction.

2. Apart from these imprecisions, systematic deviations (see Appendix S6) between model and simulation can be observed for

    a) transitions originating from the center of the cavities and

b) transitions originating from the outermost bins that involve large angular motions.

The latter are predicted to be less likely than observed.

These discrepancies could not be resolved by changing technical parameters, such as the number of radial or angular bins, the maximum order of the Legendre polynomials, or the spline-node-resolution. The systematic deviations are only relevant for very small and very large radii, where diffusion coefficients should not be interpreted quantitatively anymore. At intermediate distances (in particular at the important position of the interfacial density maximum layer), however, the model does provide an acceptable and quantitative description of the single-particle dynamics, that is valid over four orders of magnitude, as demonstrated by Figures 6.20 and 6.21. After having established the validity of a diffusion picture, hydrogen bond kinetics will be investigated in the next subsection, since it is known to couple to diffusion.[258]

6.2.6 Hydrogen Bond Kinetics

Here, we shall investigate whether hydrogen bond kinetics can be described by a similar formalism as in bulk water (i.e., by a single set of well-defined rate constants).[116] Since most molecules are interfacial in the relatively small fullerenes studied here and since there is only one type of hydrogen bonds present in these systems (water–water hydrogen bonds), assuming a single set of rate constants should be a reasonable approximation. For this purpose, hydrogen bond network dynamics was described by the kinetic model of Luzar and Chandler (in a slightly modified fashion, so that it can also be applied to the finite-sized systems investigated in this work).[116,258] The method relies on the definition of a hydrogen bond population operator $h(t)$, which equals unity if a particular tagged pair of water molecules is hydrogen bonded and zero otherwise. Its ensemble average $\langle h \rangle$ can be interpreted as the probability that a randomly chosen pair of water molecules is hydrogen bonded. In a macroscopic, infinite system, this probability approaches zero; in the finite systems investigated in this work, it won't.

With the aid of the hydrogen bond operator $h(t)$ and its fluctuations

from equilibrium $\delta h(t) = h(t) - \langle h \rangle$, a hydrogen bond time correlation function $c(t)$ can be defined:

$$c(t) = \frac{\langle \delta h(t) \delta h(0) \rangle}{\left\langle [\delta h(0)]^2 \right\rangle} . \tag{6.24}$$

This function measures the fluctuations of hydrogen bond populations in time, or, if interpreted probabilistically, gives the conditional probability that an initially hydrogen bonded pair is still bonded at a time t later. It corresponds to the family of intermittent time correlation functions introduced by Rapaport.[292] For easier illustration, $c(t)$ is shown in Figure 6.22 (top left) for bulk SPC/E water. The remaining plots and functions in Figure 6.22 are not relevant yet and will be introduced shortly. Note that for sufficiently large systems, $\langle h \rangle \to 0$, so that $c(t)$ reduces to its conventional definition

$$c(t) = \frac{\langle h(t) h(0) \rangle}{\langle h^2(0) \rangle} . \tag{6.25}$$

According to Onsager's regression hypothesis, the same laws that govern the time evolution of the hydrogen bond time correlation function, drive the decay of an initial non-equilibrium population $\overline{\delta h}(0)$ towards the equilibrium, that is,

$$c(t) = \frac{\langle h(t) h(0) \rangle}{\langle h^2(0) \rangle} = \frac{\overline{\delta h}(t)}{\overline{\delta h}(0)} . \tag{6.26}$$

The rate of relaxation is given by

$$k(t) = -\dot{c}(t) = \frac{\left\langle [1 - h(t)] \dot{h}(0) \right\rangle}{\left\langle [\delta h(0)]^2 \right\rangle} , \tag{6.27}$$

as explained in ref. [258]. This function is shown in Figure 6.22 (bottom left) for bulk SPC/E water. The quantity $-\dot{h}(0) = -(dh/dt)_{t=0}$ is the integrated flux departing the hydrogen bond configuration space at time zero. The function $k(t)$ measures the average of that flux for those trajectories, where the bond is broken at a time t later, hence its name: reactive

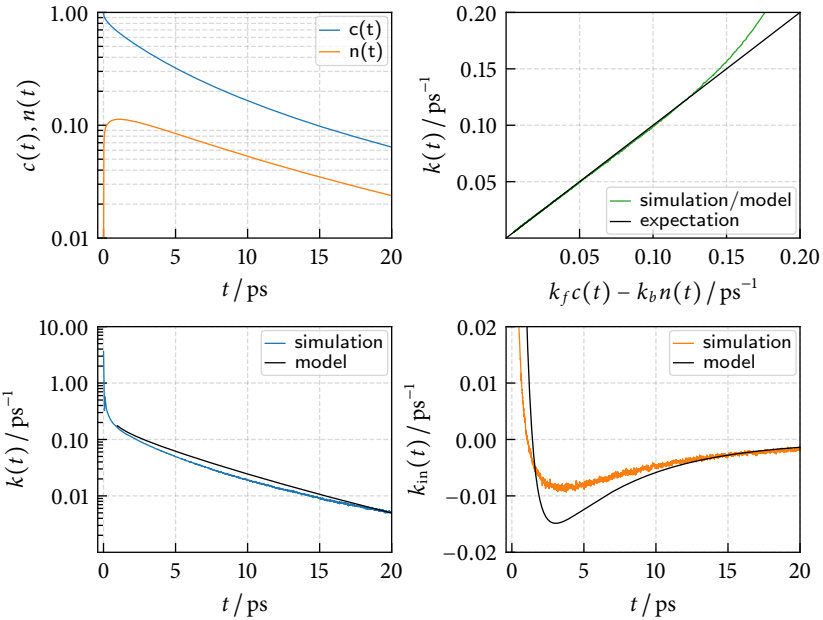


Figure 6.22: Hydrogen bond kinetics in bulk SPC/E water. Obtained from 5 ns equilibrium trajectories of 256 molecules at $\langle T \rangle = 298.15\,\mathrm{K}$ and $\rho = 0.998\,\mathrm{g\,cm^{-3}}$. Lennard-Jones interactions were cut off after 1.0 nm and Coulomb interactions were treated by a particle-particle particle-mesh solver. The integration time step was 1 fs. Other computational details were equal to those described in Section 6.2.2. The figure shows hydrogen bond time correlation functions $c(t)$ and $n(t)$ (top left), correlation between model and simulation data for $k_f = 0.35\,\mathrm{ps^{-1}}$ and $k_b = 0.72\,\mathrm{ps^{-1}}$ (top right), as well as the reactive flux functions $k(t)$ (bottom left) and $k_{\mathrm{in}}(t)$ (bottom right) and the respective model predictions for $\tau_{\mathrm{d}} = 0.46\,\mathrm{ps}$. See main text for an in-depth discussion.

flux hydrogen bond time correlation function. Its zero time value is the transition state theory estimate of the rate of relaxation.[293]

In order to extract hydrogen bond lifetimes from such correlation functions, a model is required that provides physical insight into the process. The easiest model is to assume that hydrogen bonds dissociate with a con-

stant rate and independent of the time since the last breaking event. Such a process is called a Poisson process and as explained previously, time correlation functions of such processes decay mono-exponentially. Thus, if an exponentially decaying hydrogen bond time correlation function is observed, the proposed model can be assumed to provide an adequate description of the underlying physics and the rate of the Poisson process (the inverse lifetime) can be identified as the exponential decay constant.

In bulk water, $k(t)$ does not decay mono-exponentially and reveals several motions leading to bond breaking.[116,258] A quick change within times up to 0.1 ps which is primarily due to librations is followed by interoxygen vibrations on time scales between 0.1 and 0.2 ps. Beyond that transient period, a continuum of time scales follows, resulting in a monotonic, non-exponential decay. Luzar and Chandler have shown[116,258] that this behavior can be understood with a simple diffusion picture: After bond breaking, a pair can drift apart, and similarly, diffusion of two molecules towards each other can lead to bond reforming. In order to investigate this, they separated the contributions to $k(t)$ according to whether a pair did or did not move apart after bond breaking, that is, by investigating the *restrictive* reactive flux hydrogen bond time correlation function

$$k_{\mathrm{in}}(t) = \frac{\left\langle H(t)[1-h(t)]\dot{h}(0)\right\rangle}{\left\langle [\delta h(0)]^2\right\rangle}, \tag{6.28}$$

which is also plotted in Figure 6.22 (bottom right). The vicinity operator $H(t)$ is responsible for that partitioning. It equals unity if the particular tagged pair has not yet drifted apart, and zero otherwise.[a] The probability that an initially bonded pair is broken, but has not drifted apart at a time t is given by

$$n(t) = \int_0^t k_{\mathrm{in}}(\tau)\, d\tau = \frac{\left\langle H(t)[1-\delta h(t)]\delta h(0)\right\rangle}{\left\langle [\delta h(0)]^2\right\rangle}. \tag{6.29}$$

The probabilities $c(t)$ and $n(t)$ (see Figure 6.22, top left) correspond to

[a] For this purpose, a simple geometric distance criterion is typically used. Here, $H(t)$ equals one if the oxygen–oxygen separation is below a cutoff of 350 pm, and zero otherwise.

local populations that can interconvert. A phenomenological description of their kinetics that is consistent with simulation data is

$$-\dot{c}(t) = k_f c(t) - k_b n(t)\,, \tag{6.30}$$

where the forward and backward rate constants k_f and k_b are the rate constants with which hydrogen bonds break and reform, respectively. The physical meaning of $\tau_{\mathrm{hb}} = 1/k_f$ is that of the average hydrogen bond lifetime. The required rate constants k_f and k_b can be obtained by comparing $k(t)$ to $k_f c(t) - k_b n(t)$, which should give a straight line on time scales where the model is expected to be valid (i.e., beyond the transient regime). The respective correlation plot for bulk SPC/E water is shown in Figure 6.22 (top right) for $k_f = 0.35\,\mathrm{ps}^{-1}$ and $k_b = 0.72\,\mathrm{ps}^{-1}$; values that agree very well with previously reported rate constants for the same water model.[197,294] As expected, deviations from unity slope occur only for transient times $t < 1\,\mathrm{ps}$.

By invoking a simple model, predictions of $k(t)$ and $k_{\mathrm{in}}(t)$ can be calculated and compared to simulation data. On time scales that are longer than the transient regime and on length scales that are larger than the spatial extent of a hydrogen bond, the following modified Fick's law for pair diffusion should hold:

$$\frac{\partial}{\partial t}\rho(\mathbf{r},t) = D\,\nabla\cdot\big[\nabla\cdot\rho(\mathbf{r},t)\big] + \delta(\mathbf{r})\big[k_f c(t) - k_b n(t)\big]\,, \tag{6.31}$$

Therein, $\rho(\mathbf{r},t)$ is the time and distance dependent density of the diffusing unbonded pair, $\mathbf{r}$ is the vector between the pair, and $D = 2D_0$ is the pair diffusion coefficient of bulk water.[116,258] In this equation, pair diffusion is subject to source and sink terms (expressed by the second term on the right hand side), since a pair may bond or break apart. Such effects are only relevant on length scales where a hydrogen bond may exist. The delta function $\delta(\mathbf{r})$ localizes the source and sink terms to such length scales. Eq. (6.31) can be solved in Laplace space for $k(t)$ and $k_{\mathrm{in}}(t)$.[116,258] The results depend on a free parameter τ_{d} which is expected to lie within the range 0.1–1.0 ps in bulk water.[116,258] In the language of stochastic methods, this diffusive time scale τ_{d} corresponds to the mean first passage time (MFPT) of an escape process.[295] Note that such times

typically depend on a length scale a and a diffusion coefficient D. For an isotropic bulk system,

$$\tau_\mathrm{d} = a^2/D\left(6\pi^2\right)^{2/3}. \tag{6.32}$$

Model predictions of $k(t)$ and $k_\mathrm{in}(t)$ obtained from solution of eq. (6.31) are shown in Figure 6.22 (bottom panels) for $\tau_\mathrm{d} = 0.46\,\mathrm{ps}$ and are in good agreement with the observed simulation data. Quantitative improvements could be expected by refining the resolution of eq. (6.31) in space and time, for example, by turning to the full Smoluchowsky equation for pair diffusion (eq. (6.15)). Technical details on the calculation of hydrogen bond time correlation functions and the solution of eq. (6.31) are given in Appendix S6.

We shall now turn to the discussion of hydrogen bond kinetics within the fullerene cavities investigated in this work. All analyses described above were repeated for each system. As an example, the same plot shown in Figure 6.22 for bulk water is presented for C_{500} (hydrophobic) in Figure 6.23. Results for the other systems are qualitatively similar and are shown in Appendix S6 to avoid repetition. The following observations can be made for all systems:

1. No correlation function decays mono-exponentially beyond the transient regime, thus a non-Poissonian model must be employed in order to extract hydrogen bond lifetimes. For this purpose, the diffusion model described above was chosen.

2. The correlation plots show straight lines beyond times of 1.3–1.5 ps, demonstrating the validity of the Luzar–Chandler diffusion model in spherical confinement.

3. By optimizing the diffusive time scale τ_d, hydrogen bond kinetics can be described to some extent using the same model that is valid in bulk water (eq. (6.31)). The agreement is less convincing, though.

The full set of model parameters ($k_f, k_b, \tau_\mathrm{d}$) is listed in Table 6.5 and the characteristic time scales are visualized in Figure 6.24. Hydrogen

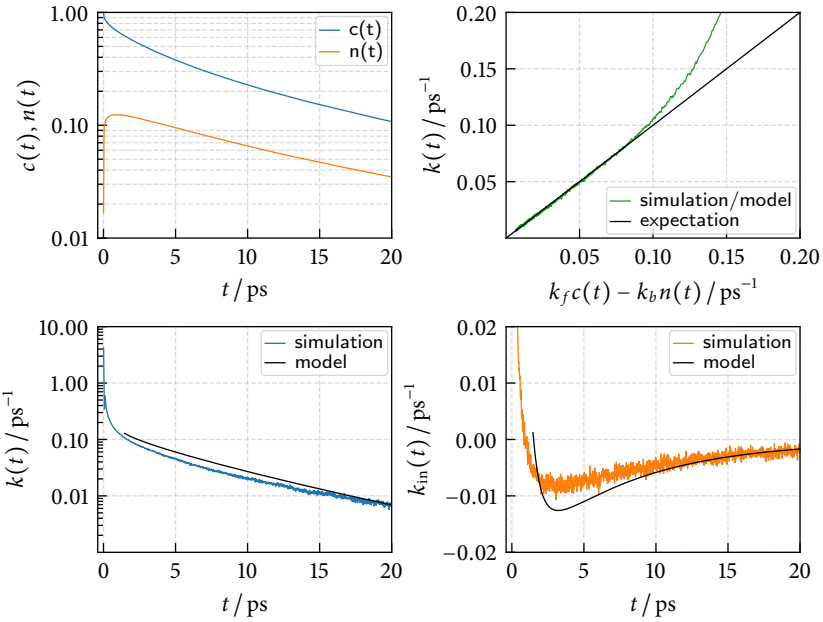


Figure 6.23: Hydrogen bond kinetics of SPC/E water confined inside C_{500} (hydrophobic). The figure shows hydrogen bond time correlation functions $c(t)$ and $n(t)$ (top left), correlation between model and simulation data for $k_f = 0.33\,\mathrm{ps}^{-1}$ and $k_b = 0.83\,\mathrm{ps}^{-1}$ (top right), as well as the reactive flux functions $k(t)$ (bottom left) and $k_{\mathrm{in}}(t)$ (bottom right) and the respective model predictions for $\tau_{\mathrm{d}} = 0.46\,\mathrm{ps}$.

bond lifetimes $\tau_{\mathrm{hb}} = k_f^{-1}$ are universally longer in confinement. The slow-down occurs monotonically with decreasing cavity size in hydrophobic confinement and non-monotonically in hydrophilic confinement. Hydrogen bond 'deadtimes' $\tau_{\mathrm{dead}} = k_b^{-1}$ on the other hand show the same trend as characteristic times of reorientation. There is a monotonic acceleration of reformation dynamics in hydrophobic systems and a non-monotonic slow-down in hydrophilic cavities with decreasing cavity size. This makes sense, since τ_{dead} is a measure for the time between stable HB configurations, and thus a major contribution to reorientation.

Characteristic time scales of diffusion τ_{d} are only slightly affected by

Table 6.5: Forward and backward rate constants k_f and k_b (in units of ps^{-1}), describing the hydrogen bond kinetics, as well as diffusive time scales τ_d and hydrogen bond lifetimes $\tau_{hb} = 1/k_f$ (in units of ps). Note that both k_f and k_b can be varied by ≈ 10 %, while still leading to acceptable model fits. Bulk values are $k_f = 0.35\,ps^{-1}$, $k_b = 0.72\,ps^{-1}$, $\tau_d = 0.46$ ps, and $\tau_{hb} = 2.90\,ps^{-1}$.

	hydrophilic				hydrophobic			
	k_f	k_b	τ_d	τ_{hb}	k_f	k_b	τ_d	τ_{hb}
C_{320}	0.21	0.54	0.77	4.76	0.30	0.89	0.56	3.33
C_{500}	0.19	0.40	0.66	5.26	0.33	0.82	0.52	3.08
C_{720}	0.26	0.57	0.59	3.92	0.33	0.76	0.52	3.08
C_{1500}	0.29	0.65	0.52	3.45	0.33	0.75	0.48	3.03

the confinement, rising with decreasing cavity size. This effect is more pronounced in hydrophilic confinement. While τ_d values are within a physically meaningful range of values according to eq. (6.32), deeper correlation with the diffusion profiles reported in Section 6.2.5 remains elusive. This is not surprising. Eq. (6.31) was designed for isotropic bulk systems and is only valid for time and length scales at which the influence of the potential energy surface describing pair diffusion has averaged out. However, the spherical cavities investigated here are anisotropic. The single-particle dynamics is position dependent and the same can be expected for pair diffusion. Future studies would involve extending eq. (6.31) to include the anisotropy inherent to spherical confinement.

6.2.7 Conclusion

Spherical confinement has a pronounced impact on the structure and dynamics of water. Geometric restrictions and specific interactions lead to an inhomogeneous potential of mean force, that organizes water molecules in distinct layers, or better, concentric shells. Despite of these inhomogeneities, water's tetrahedral hydrogen bond network is hardly disturbed beyond the interfacial water layer.

Much more drastic effects can be seen in the position dependent dif-

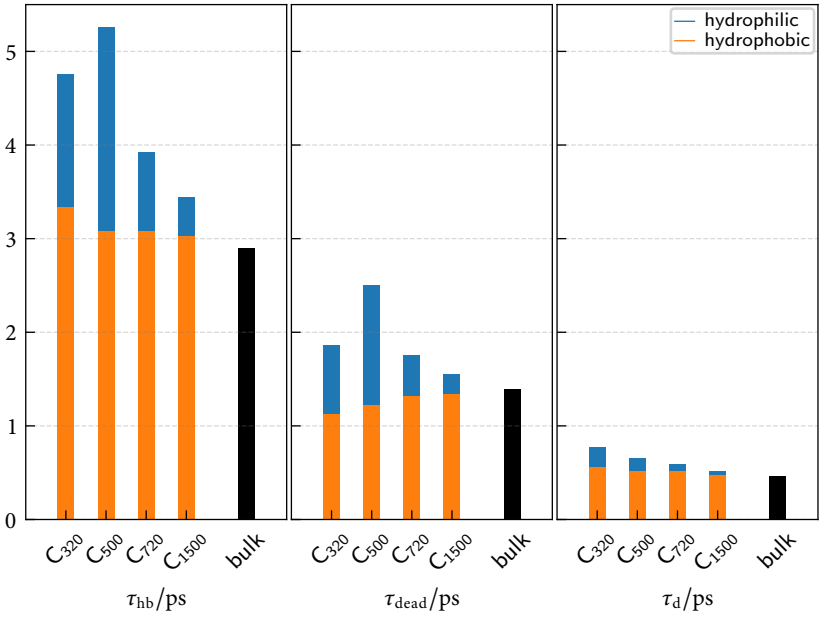


Figure 6.24: Hydrogen bond lifetimes $\tau_{hb} = k_f^{-1}$, 'deadtimes' $\tau_{dead} = k_b^{-1}$, and diffusion time scales τ_d in confinement and bulk.

fusion coefficients, which were calculated for the first time in spherical confinement. With the aid of Hummer's Bayesian inference scheme,[287,288] diffusion profiles that are correlated in a physically meaningful way with the position dependent free energies of the systems were obtained. In terms of overall magnitude, diffusion coefficients are larger in hydrophobic confinement, especially in angular direction. Similar effects have also been observed close to planar interfaces.[245] One system stands out, by showing unusually slow diffusion in any direction: hydrophilic C_{500}. The same system shows non-monotonic behavior in all other dynamic properties.

Hydrogen bond kinetics was analyzed by employing the diffusion model of Luzar and Chandler.[116,258] The validity of this model in spherical confinement has been established by means of the correlation plots (Figure S20, Appendix S6), which are straight lines on time scales where

the model can be expected to hold. Thus for the first time, well-defined hydrogen bond lifetimes were calculated in a systematic study of water in spherical confinement. With decreasing cavity size, hydrogen bond lifetimes increase non-monotonically in hydrophilic confinement and monotonically in hydrophobic systems. Just as for diffusion, the non-monotonic behavior is caused by a drastic deceleration of dynamics in C_{500} (hydrophilic). These results are qualitatively in agreement with OH bond reorientation dynamics and fit well into the literature context discussed in the beginning.

Finally, an attempt was made to predict the reactive flux time correlation functions $k(t)$ and $k_{\text{in}}(t)$ from a simple diffusion equation (eq. (6.31)). This equation employs a single constant pair diffusion coefficient for time and length scales on which effects causing variations of the pair diffusion coefficient have averaged out. Such an assumption should not be expected to hold in the inhomogeneous and anisotropic confinements investigated in this work, as demonstrated by the constructed diffusion model. Nevertheless, by optimizing the diffusive time scale τ_D, acceptable predictions of $k(t)$ and $k_{\text{in}}(t)$ could be obtained, albeit being less convincing than for bulk water. Deeper correlations between diffusion profiles and diffusive time scales could not be gained, however. Future work would address this issue, by studying a refined hydrogen bond kinetics/diffusion model that takes the intrinsic inhomogeneity and anisotropy of these systems into account.

The confinements investigated in this work do not form hydrogen bonds with water molecules, thus slowed down dynamics (in particular in hydrophilic confinement) cannot be attributed to the presence of such bonds. Instead, density appears to be a major determinant of the dynamics that would also explain the outstanding behavior of C_{500}, hydrophilic. It would also agree well with previously reported results by Kayal and Chandra in much less occupied fullerenes.[257] Future work should thus address density variations in fullerenes of a given size to confirm this assumption.

Conclusion and Outlook

The first part of this thesis was devoted to Quantum Cluster Equilibrium (QCE) theory,[17] the basic concepts of which were introduced in Chapter 2. The fundamental idea behind this theory is simple: QCE introduces quantum chemistry to the broad and well-established class of mixture theories.[7] Liquids are assumed to be mixtures of clusters, which may interconvert, and their equilibrium cluster distribution is sought in order to calculate the system's partition function and associated thermodynamic potentials. Quantum chemical calculations are applied to the internal cluster structure and an empirical mean field approximation is used to account for the residual cluster-cluster interactions.

In Chapter 2, the origin of the empirical correction parameters was analyzed in great detail and the tight connection between the van der Waals equation of state and QCE theory has been established. This connection was exploited to refine binary Quantum Cluster Equilibrium (bQCE) theory,[31] which has been achieved by sound modifications to the expressions for the binary mean field energy and the exclusion volume. A summary of all developments was given in Section 2.7. In combination with suitable mixing rules lent from cubic equations of state, these measures can eliminate the need for binary reference data, which was proven to be a viable approximation in Section 5.2. Even though such mixing rules are commonly associated with ideal mixing behavior, non-ideal mixing effects are inherent to QCE theory because of the internal quantum chemical description of the clusters. Predicting miscibilities without need for binary reference data or resorting to empiricism is an important and open task in modern theoretical research on binary mixtures that is well within the reach of QCE theory.

After realizing the connection between the van der Waals equation of state and QCE theory, a new route towards future improvements to QCE theory can be sketched. The van der Waals equation is a cubic equation of state—a simple and educationally highly appealing one, but certainly not one of the best.[28] Consequently, changing the underlying equation of state would be a straightforward measure to improve QCE theory. Many alternatives to van der Waals's equation exist, but most are empirical in their nature. In my opinion, future developments of QCE theory should be directed towards less empiricism, since there is already a large number of other empirical liquid and mixture models, whereas

few- or no-parameter theories present a niche that can be filled by QCE. It would thus be advisable to improve upon the van der Waals equation of state in a non-empirical fashion, for example, by using lessons learned from a perturbation theoretic derivation of the equation.[28]

Apart from the quality of the quantum chemical method that is employed, the success of a QCE calculation depends sensitively on the chosen cluster set. Such cluster set effects are most pronounced in binary mixtures, where system properties depend delicately on a properly balanced description of both components.[31] In the spirit of reducing empiricism, a first systematic and semi-automatic scheme to generate cluster sets has been proposed in Section 5.2. Only global minimum structures, obtained from a genetic-algorithm-based global geometry optimization, were employed therein. The convergence behavior with respect to the maximum cluster set size was studied and results for thermodynamic potentials of mixing were calculated for three different amide/water mixtures. Cluster sets based on global minima perform well for enthalpies of mixing, reaching almost chemical accuracy. However, correctly describing entropic contributions to the mixing Gibbs energy turned out to be out of their reach, since such cluster sets lack energetically less stable, yet entropically important cluster structures.[a] Fortunately, the proposed method can be easily extended. Many genetic-algorithm-based global structure optimization algorithms keep track of less stable energy minima to increase the genetic diversity of the population pool. These can (and should) be included in future studies. In combination with semi-empirical methods or low-cost quantum chemistry,[155] large, systematically constructed, and balanced cluster sets are within foreseeable reach.

QCE theory can go beyond the prediction of thermodynamic potentials and offer structural insight, as well. A novel application was presented in Section 5.1, where the ionic product of water was calculated by adding ion pair clusters to a conventional QCE cluster set. This study has

[a] Note that another source of errors to QCE theory, in particular to the entropy of the system, stems from the approximated treatment of internal degrees of freedom, such as rotations and vibrations. The neglect of anharmonicities or rotational-vibrational coupling is expected to contribute particularly and we have recently started to address such issues.[53]

opened the path to exciting new studies on acid-base related phenomena using QCE theory.

The proposed theoretical improvements and the performed QCE calculations were made possible by the Peacemaker QCE software,[117] which was rewritten from scratch as part of this thesis. Its implementation was covered in Chapter 4. Peacemaker 2 is a fast, robust, and well-documented code that permits running all described calculations on a laptop. The bottleneck of any QCE calculation is thus the generation of the cluster set.

In its very nature, QCE theory is static and applies to systems in equilibrium only. Some kinetic properties can be extracted from QCE calculations by making use of transition state theory (TST),[23,43,296] but for many problems, the approximations that enter into TST are inadequate. A prominent example is hydrogen bond kinetics, which is the driving force of liquid water dynamics.[21,22] Recrossings of the energy barrier, which are prohibited in TST, lead to reformation of hydrogen bonds and reduce the associated rate constant.[258] Accurate kinetics is accessible from molecular dynamics simulation, which will probably continue to be the method of choice for the investigation of dynamic or non-equilibrium processes. For this reason, MD was the second dominant method employed in this thesis and was applied to a selection of problems that occur in nanoscopically confined water systems.

One of the most fundamental questions that can be asked about the dynamics of water is: "How long does a hydrogen bond live?" In bulk water, Luzar and Chandler have demonstrated that the answer to this question is intimately connected to diffusion.[116] After uncoupling diffusion, they have found hydrogen bond kinetics to be characterized by simple, well-defined rate constants. In nature and technology, water is often confined inside spherical, nanometer-sized cavities. Whether relations similar to bulk water hold in such types of nanoconfinement had been an open question, and was investigated in Section 6.2. Therein, a diffusive model was constructed by Bayesian inference from simulation data, which describes the single-particle dynamics of water molecules inside model fullerenes. The propagators of the diffusion model are in good agreement with simulation data over four orders of magnitude, instilling great confidence in the model. There was no a priori reason

to believe in the existence of such a diffusion model, but after having established its validity, hydrogen bond kinetics could be meaningfully treated within the Luzar–Chandler diffusion model.[116] Overall, hydrogen bond lifetimes slow down with decreasing cavity diameter, though some non-monotonous effects arose, which should be investigated more carefully in future studies. An attempt was made to predict hydrogen bond time correlation functions from a simple, uniform diffusion model with sink and source terms, but could not be found to be reliable in spherical nanoconfinement. Various ways to improve upon this model have already been described in the conclusion of Section 6.2 and should be explored in follow-up investigations.

Hydrogen bonds are also the origin of some outstanding static properties of water.[21] Perhaps one of the most important features that becomes relevant on the nanoscale is the orientational polarization of water close to interfaces.[207,208,220–222] This polarization can couple with electric fields, giving rise to what has been described as a field-induced Janus interface.[168–172] In Section 6.1, interfacial dynamics in such systems was investigated, which is an important criterion in the design of electroswitchable, nanofluidic devices and may be of relevance in understanding the function of biological voltage-gated channels. A new and significant polarity dependence of interfacial polarization dynamics was uncovered, with asymmetries in response times that are even more pronounced than the asymmetries in static properties of a field-induced Janus interface.

Even though the topics and methods covered in this thesis offer only a brief glimpse at current water research, they are broad and illustrate the complexity. Computer simulations are omnipresent and powerful, and advanced applications to nanoscopic systems were presented herein. Yet there are problems that, right now, cannot be satisfactorily answered by simulation, necessitating the use of other techniques. Quantum Cluster Equilibrium theory is a promising candidate with great potential for the description of thermodynamic properties of neat and binary systems, and offers certain structural insight as well. Multiscale and multimethod modeling seems to be the answer to most challenging contemporary research problems, and both methods presented are likely to play a role in such approaches.

Appendices

Supplementary Material

S1 Modified QCE Polynomials

The modified binary mean field energy (2.42) does not only affect the electronic partition function, but also the volume polynomial, which is derived therefrom. The modified equations will be derived here.

The mean field energy is given by

$$\begin{aligned}\varepsilon_i^{\mathrm{mf}} &= -\frac{1}{V}\left(n_i^a N_1^{\mathrm{tot}} a_{\mathrm{mf}}^a + n_i^a N_2^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + n_i^b N_1^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + n_i^b N_2^{\mathrm{tot}} a_{\mathrm{mf}}^b\right) \\ &= -\frac{A_{\mathrm{mf}}}{V}\,,\end{aligned} \tag{S1}$$

where A_{mf} was introduced for convenience's sake. The derivative of $\ln q_i^{\mathrm{elec}}$ with respect to volume becomes

$$\frac{\partial \ln q_i^{\mathrm{elec}}}{\partial V} = -\frac{A_{\mathrm{mf}}}{kTV^2}\,. \tag{S2}$$

Repeating the steps that led to the volume polynomial (2.38) yields:

$$\begin{aligned}0 &= -P + kT\sum_{i=1}^{K} N_i\left[\frac{1}{V - V^{\mathrm{ex}}} - \frac{A_{\mathrm{mf}}}{kTV^2}\right] \\ &= -PV^2(V - V^{\mathrm{ex}}) + kT\sum_{i=1}^{K} N_i\left[V^2 - \frac{A_{\mathrm{mf}}(V - V^{\mathrm{ex}})}{kT}\right] \\ &= -PV^3 + \left(PV^{\mathrm{ex}} + kT\sum_{i=1}^{K} N_i\right)V^2 - \left(\sum_{i=1}^{K} N_i A_{\mathrm{mf}}\right)(V - V^{\mathrm{ex}})\,.\end{aligned} \tag{S3}$$

We shall now investigate the coefficient in front of $(V - V^{\mathrm{ex}})$ more closely.

Resubstitution of A_{mf}'s definition leads to:

$$\sum_{i=1}^{K} N_i A_{\mathrm{mf}} = \sum_{i=1}^{K} N_i \left(n_i^a N_1^{\mathrm{tot}} a_{\mathrm{mf}}^a + n_i^a N_2^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + n_i^b N_1^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + n_i^b N_2^{\mathrm{tot}} a_{\mathrm{mf}}^b \right). \tag{S4}$$

Therein, the following substitutions can be made:

$$\sum_{i=1}^{K} N_i n_i^a = N_1^{\mathrm{tot}}; \qquad \sum_{i=1}^{K} N_i n_i^b = N_2^{\mathrm{tot}}. \tag{S5}$$

Thus, eq. (S4) reduces to

$$\sum_{i=1}^{K} N_i A_{\mathrm{mf}} = \left(N_1^{\mathrm{tot}}\right)^2 a_{\mathrm{mf}}^a + 2 N_1^{\mathrm{tot}} N_2^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + \left(N_2^{\mathrm{tot}}\right)^2 a_{\mathrm{mf}}^b, \tag{S6}$$

which for $a_{\mathrm{mf}}^{ab} = \sqrt{a_{\mathrm{mf}}^a a_{\mathrm{mf}}^b}$ (i.e, for $k^{ab} = 0$) would conveniently reduce to

$$\sum_{i=1}^{K} N_i A_{\mathrm{mf}} = \left(N_1^{\mathrm{tot}} \sqrt{a_{\mathrm{mf}}^a} + N_2^{\mathrm{tot}} \sqrt{a_{\mathrm{mf}}^b} \right)^2. \tag{S7}$$

However, generally, this is not the case, thus the modified volume polynomial reads

$$\begin{aligned} 0 = & - PV^3 \\ & + \left(PV^{\mathrm{ex}} + kT \sum_{i=1}^{K} N_i \right) V^2 \\ & - \left[\left(N_1^{\mathrm{tot}}\right)^2 a_{\mathrm{mf}}^a + 2 N_1^{\mathrm{tot}} N_2^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + \left(N_2^{\mathrm{tot}}\right)^2 a_{\mathrm{mf}}^b \right] V \\ & + \left[\left(N_1^{\mathrm{tot}}\right)^2 a_{\mathrm{mf}}^a + 2 N_1^{\mathrm{tot}} N_2^{\mathrm{tot}} a_{\mathrm{mf}}^{ab} + \left(N_2^{\mathrm{tot}}\right)^2 a_{\mathrm{mf}}^b \right] V^{\mathrm{ex}}. \end{aligned} \tag{S8}$$

This equation was implemented in the latest version of Peacemaker.

S2 Peacemaker Input and Cluster Set Files

In this part of the appendix, the Peacemaker input and cluster set files will be explained in full detail. The following notational conventions apply: real numbers are referred to as A, B or C, integers are labeled as N or M, and file system paths are designated by the letter P.

Input File

[system]

`components N`
: The number of components in the system. Currently only 1 (pure) and 2 (binary) are supported. Optional. Default: 1

[qce]

`amf A`
`amf A B N`
: The mean field parameter a_{mf} in units of J m^3 mol^{-2}. Can be specified either as a single value A, or as a range A, B, N, where A is the start, B the end, and N the number of data points (including both boundaries). For binary mixtures, this keyword is equivalent to `amf_mix`. Optional. Default: 0.0

`amf_pure A B`
: The pure-component mean field parameters a^a_{mf} and a^b_{mf} in units of J m^3 mol^{-2}. May be specified for binary mixtures only. Optional. If unspecified, $a^a_{\mathrm{mf}} = a^b_{\mathrm{mf}} = a^{ab}_{\mathrm{mf}}$ is used.

`amf_mix A`
`amf_mix A B N`
: The mixed-species mean field parameter a^{ab}_{mf} in units of J m^3 mol^{-2}. Can be specified either as a single value A, or as a range A, B, N, where A is the start, B the end, and N the number of data points (including both boundaries). For binary mixtures, this keyword is equivalent to `amf`. Optional. Default: 0.0

`bxv A`

`bxv A B N`

The exclusion volume scaling parameter b_{xv}. Can be specified either as a single value A, or as a range A, B, N, where A is the start, B the end, and N the number of data points (including both boundaries). Optional. Default: 1.0

`bxv_pure A B`

The pure-component exclusion volume scaling parameters b^a_{xv} and b^b_{xv}. May be specified for binary mixtures only. Optional. Default: 1.0

`max_deviation A`

The maximum relative deviation of the Gibbs energy. Used to check convergence of the QCE iteration. A QCE cycle has converged, if

$$\left|\frac{G(\text{current step}) - G(\text{last step})}{G(\text{last step})}\right| < A.$$

Optional. Default: 1.0e-9

`volume_damping_factor A`

The volume damping factor used to damp the initial volume guess if one of the polynomials did not converge. Shall be between 0 and 1. Damping is performed by $\gamma_V = 1 \pm A$, depending on the mode of the temperature loop. Optional. Default: 0.01

`qce_iterations N`

The maximum number of iterations in a QCE cycle. Optional. Default: 100

`newton_iterations N`

The maximum number of iterations in a Newton–Raphson cycle. Optional. Default: 500

[reference]

This section is optional. It enables comparison to experimental reference data. It is disabled by default.

`density A B`

`density A B C`
: Reference density B in units of g cm^{-3} at reference temperature A in K and an optional error weight C. Optional.

`isobar P`
`isobar P A`
: Path to an isobar file P and an optional error weight A. Isobar files contain two columns representing the temperature in K and volume in L. Optional.

`phase_transition A`
`phase_transition A B`
: Reference temperature of phase transition A in units of K and an optional error weight B. Optional.

[output]

This section is optional. It enables output control. It is disabled by default.

`contributions`
`contributions <helmholtz|internal|entropy|cv>`
: Enables the output of contributions of each degree of freedom to the thermodynamic functions. If no arguments are given, contribution output is enabled for all possible thermodynamic quantities. If arguments are specified, contribution output is only enabled for the selected thermodynamic quantities. Optional.

Cluster Set File

The cluster set file contains one section ([label]) per cluster, where 'label' must be a unique identifier for each cluster. Unless otherwise specified, all cluster entries may include the following keywords.

`monomer`
: Sets the current cluster as monomer. Optional, but must be present once for each component.

`composition N M`
: Composition of the cluster in number of monomers. One number for each component.

`sigma N`
: The rotational symmetry number of the cluster.

`coordinates P`
: Path to a coordinate file in the xyz format. Units are Å.

`frequencies P`
: Path to a frequency file. It contains the number of frequencies in line 1, followed by a comment line, followed by one frequency per line. Units are cm^{-1}.

`energy A`
: The adiabatic interaction energy of the cluster in units of $kJ\ mol^{-1}$. Calculated according to eq. (2.17) (i.e., negative energies represent stable clusters).

`volume A`
: The volume of the cluster in units of $Å^3$. Must be specified for monomers, only.

`frequency_scale A`
: A frequency scaling factor. Optional.

`anharmonicity A`
: The anharmonicity constant. Optional.

S3 Supplementary Data: Ionic Product

Optimized values for a_{mf} and b_{xv} are given in Table S1 for all investigated methods. The adiabatic binding energies for all clusters and methods are listed in Table S2.

Table S1: Optimized QCE parameters a_{mf} and b_{xv} for the different methods applied in this study.

	a_{mf}/J m^3 mol^{-2}	b_{xv}
HF	0.3260	1.5122
B3LYP	0.2301	1.5147
B3LYP-D3	0.1549	1.4948
B3LYP-D3,gCP	0.2033	1.5081
PBE0	0.1906	1.5075
PBE0-D3	0.1470	1.4920
PBE0-D3,gCP	0.1914	1.5076
PBEh-3c	0.1746	1.5028
MP2	0.2725	1.5186
CCSD(T)	0.3152	1.5211

S4 Supplementary Data: Amide Mixtures

Adiabatic binding energies of the mixed clusters are given in Table S3 and the corresponding cluster pictures are shown in Figures S1 to S3.

Table S2: Adiabatic binding energies $\Delta\varepsilon^{\text{bind}}$ in kJ mol^{-1}. A dash (–) denotes an unstable geometry at the selected level of theory.

	HF	B3LYP	B3LYP -D3	B3LYP -D3,gCP	PBE0	PBE0 -D3	PBE0 -D3,gCP	PBEh-3c	MP2	CCSD(T)
$\mathbf{W_1}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\mathbf{W_2}$	−20.12	−23.26	−25.94	−24.25	−24.75	−26.27	−24.54	−25.15	−21.67	−20.26
$\mathbf{W_{3c}}$	−57.30	−68.54	−77.18	−71.39	−72.92	−77.73	−71.74	−71.83	−64.53	−60.82
$\mathbf{W_{3u}}$	−57.28	−73.00	−81.80	−75.82	−77.60	−82.48	−76.28	−78.30	−68.85	−64.07
$\mathbf{W_{5c}}$	−133.95	−168.93	−186.44	−175.10	−179.28	−189.70	−177.91	−190.30	−157.12	−146.82
$\mathbf{W_{5p}}$	−122.45	−157.96	−181.47	−167.46	−168.92	−182.63	−168.23	−177.41	−150.36	−140.68
$\mathbf{W_{5ip}}$	34.44	−85.06	−113.88	−93.82	−105.23	−121.24	−100.53	–	−75.66	−51.19
$\mathbf{W_6}$	−167.30	−209.09	−229.51	−216.26	−221.65	−233.83	−220.03	−234.16	−193.86	−181.31
$\mathbf{W_7}$	−194.04	−249.73	−279.52	−260.89	−265.34	−283.09	−263.74	−282.41	−236.05	−219.33
$\mathbf{W_{8c}}$	−251.05	−330.60	−380.09	−351.48	−352.49	−382.48	−352.95	−375.03	−317.94	−293.54
$\mathbf{W_{8cip}}$	−101.36	−265.91	−322.37	−287.78	−296.43	−329.79	−294.22	−317.67	−251.69	−215.18
$\mathbf{W_{8b}}$	−233.93	−301.99	−341.35	−317.67	−321.28	−345.05	−320.48	−342.37	−288.60	−267.49
$\mathbf{W_{8p}}$	−226.43	−287.78	−324.70	−303.01	−306.47	−329.65	−307.18	−325.05	−272.59	−253.05
$\mathbf{W_{8ip}}$	−77.32	−235.54	−276.60	−249.75	−264.57	−289.43	−261.72	−290.17	−212.34	−175.69
$\mathbf{W_9}$	−265.11	−337.76	−377.39	−352.97	−358.93	−383.31	−357.63	−383.79	−318.91	−296.08
$\mathbf{W_{10ip}}$	−165.57	−324.95	−387.90	−349.95	−360.31	−398.60	−359.45	−387.10	−306.37	−262.00
$\mathbf{W_{10ip2}}$	−177.55	−363.83	−432.05	−391.19	−400.30	−441.63	−399.58	−430.42	−345.72	−300.08
$\mathbf{W_{10}}$	−306.67	−400.25	−460.22	−426.43	−427.46	−464.52	−429.65	−455.90	−384.12	−354.36

Table S3: Adiabatic binding energies of the mixed clusters in kJ mol^{-1}.

FA		NMF		DMF	
$\mathbf{F_1W_1}$	−42.04	$\mathbf{C_1W_1}$	−39.80	$\mathbf{D_1W_1}$	−30.74
$\mathbf{F_1W_2}$	−100.00	$\mathbf{C_1W_2}$	−97.09	$\mathbf{D_1W_2}$	−70.47
$\mathbf{F_1W_3}$	−142.47	$\mathbf{C_1W_3}$	−139.92	$\mathbf{D_1W_3}$	−122.14
$\mathbf{F_1W_4}$	−192.15	$\mathbf{C_1W_4}$	−190.06	$\mathbf{D_1W_4}$	−176.36
$\mathbf{F_1W_5}$	−246.05	$\mathbf{C_1W_5}$	−244.19	$\mathbf{D_1W_5}$	−224.80
$\mathbf{F_1W_6}$	−310.19	$\mathbf{C_1W_6}$	−310.39	$\mathbf{D_1W_6}$	−302.36
$\mathbf{F_1W_7}$	−372.29	$\mathbf{C_1W_7}$	−370.81	$\mathbf{D_1W_7}$	−352.65
$\mathbf{F_1W_8}$	−428.83	$\mathbf{C_1W_8}$	−425.89	$\mathbf{D_1W_8}$	−409.67
$\mathbf{F_1W_9}$	−475.48	$\mathbf{T_1W_9}$	−477.83	$\mathbf{D_1W_9}$	−465.57
$\mathbf{F_2W_1}$	−83.27	$\mathbf{C_1T_1W_1}$	−82.31	$\mathbf{D_2W_1}$	−66.69
$\mathbf{F_2W_2}$	−144.20	$\mathbf{C_2W_2}$	−142.50	$\mathbf{D_2W_2}$	−104.55
$\mathbf{F_2W_3}$	−193.87	$\mathbf{T_2W_3}$	−191.41	$\mathbf{D_2W_3}$	−169.75
$\mathbf{F_2W_4}$	−241.35	$\mathbf{C_2W_4}$	−236.91	$\mathbf{D_2W_4}$	−222.46
$\mathbf{F_2W_5}$	−291.59	$\mathbf{T_2W_5}$	−295.98	$\mathbf{D_2W_5}$	−278.45
$\mathbf{F_2W_6}$	−355.41	$\mathbf{C_2W_6}$	−351.53	$\mathbf{D_2W_6}$	−343.13
$\mathbf{F_2W_7}$	−427.20	$\mathbf{C_2W_7}$	−415.68	$\mathbf{D_2W_7}$	−403.09
$\mathbf{F_2W_8}$	−471.85	$\mathbf{C_2W_8}$	−470.80	$\mathbf{D_2W_8}$	−455.70
$\mathbf{F_3W_1}$	−150.46	$\mathbf{C_1T_2W_1}$	−146.55	$\mathbf{D_3W_1}$	−107.38
$\mathbf{F_3W_2}$	−208.85	$\mathbf{T_3W_2}$	−195.75	$\mathbf{D_3W_2}$	−159.54
$\mathbf{F_3W_3}$	−236.34	$\mathbf{C_2T_1W_3}$	−256.08	$\mathbf{D_3W_3}$	−211.68
$\mathbf{F_3W_4}$	−301.57	$\mathbf{C_2T_1W_4}$	−295.58	$\mathbf{D_3W_4}$	−278.98
$\mathbf{F_3W_5}$	−352.57	$\mathbf{C_2T_1W_5}$	−356.94	$\mathbf{D_3W_5}$	−321.10
$\mathbf{F_3W_6}$	−414.49	$\mathbf{C_1T_2W_6}$	−410.60	$\mathbf{D_3W_6}$	−382.47
$\mathbf{F_3W_7}$	−457.90	$\mathbf{C_3W_7}$	−466.52	$\mathbf{D_3W_7}$	−433.24
$\mathbf{F_4W_1}$	−207.97	$\mathbf{C_4W_1}$	−203.85	$\mathbf{D_4W_1}$	−147.41
$\mathbf{F_4W_2}$	−247.56	$\mathbf{T_4W_2}$	−249.87	$\mathbf{D_4W_2}$	−198.99
$\mathbf{F_4W_3}$	−301.53	$\mathbf{C_2T_2W_3}$	−310.22	$\mathbf{D_4W_3}$	−259.93
$\mathbf{F_4W_4}$	−353.54	$\mathbf{T_4W_4}$	−363.82	$\mathbf{D_4W_4}$	−312.44
$\mathbf{F_4W_5}$	−419.69	$\mathbf{T_4W_5}$	−422.72	$\mathbf{D_4W_5}$	−352.89
$\mathbf{F_4W_6}$	−473.32	$\mathbf{C_2T_2W_6}$	−473.47	$\mathbf{D_4W_6}$	−436.03
$\mathbf{F_5W_1}$	−254.61	$\mathbf{C_2T_3W_1}$	−250.20	$\mathbf{D_5W_1}$	−192.37
$\mathbf{F_5W_2}$	−315.36	$\mathbf{C_1T_4W_2}$	−301.56	$\mathbf{D_5W_2}$	−244.47
$\mathbf{F_5W_3}$	−369.17	$\mathbf{C_3T_2W_3}$	−359.09	$\mathbf{D_5W_3}$	−297.65
$\mathbf{F_5W_4}$	−411.09	$\mathbf{T_5W_4}$	−415.77	$\mathbf{D_5W_4}$	−357.30
$\mathbf{F_5W_5}$	−475.81	$\mathbf{C_5W_5}$	−474.76	$\mathbf{D_5W_5}$	−403.40
$\mathbf{F_6W_1}$	−305.47	$\mathbf{C_4T_2W_1}$	−307.73	$\mathbf{D_6W_1}$	−225.37
$\mathbf{F_6W_2}$	−371.92	$\mathbf{C_4T_2W_2}$	−354.92	$\mathbf{D_6W_2}$	−282.93
$\mathbf{F_6W_3}$	−411.76	$\mathbf{T_6W_3}$	−427.82	$\mathbf{D_6W_3}$	−340.27
$\mathbf{F_6W_4}$	−486.81	$\mathbf{T_6W_4}$	−476.97	$\mathbf{D_6W_4}$	−376.18
$\mathbf{F_7W_1}$	−379.95	$\mathbf{C_4T_3W_1}$	−358.49	$\mathbf{D_7W_1}$	−265.57
$\mathbf{F_7W_2}$	−435.95	$\mathbf{T_7W_2}$	−410.32	$\mathbf{D_7W_2}$	−325.46
$\mathbf{F_7W_3}$	−469.86	$\mathbf{T_7W_3}$	−481.01	$\mathbf{D_7W_3}$	−374.54
$\mathbf{F_8W_1}$	−428.25	$\mathbf{C_2T_6W_1}$	−412.09	$\mathbf{D_8W_1}$	−314.33
$\mathbf{F_8W_2}$	−477.89	$\mathbf{C_3T_5W_2}$	−460.93	$\mathbf{D_8W_2}$	−379.06
$\mathbf{F_9W_1}$	−481.14	$\mathbf{C_1T_8W_1}$	−474.70	$\mathbf{D_9W_1}$	−351.48

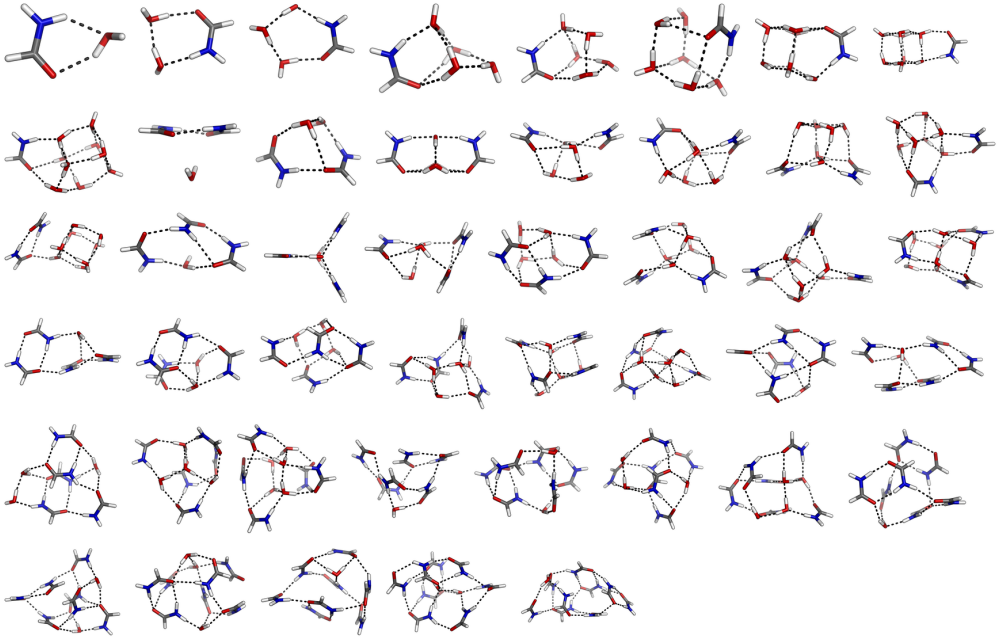


Figure S1: Mixed FA/water cluster set. Clusters are ordered as in Table S3 and arranged in rows first. Hydrogen bonds have been identified by a loose geometric criterion and are for visualization only.

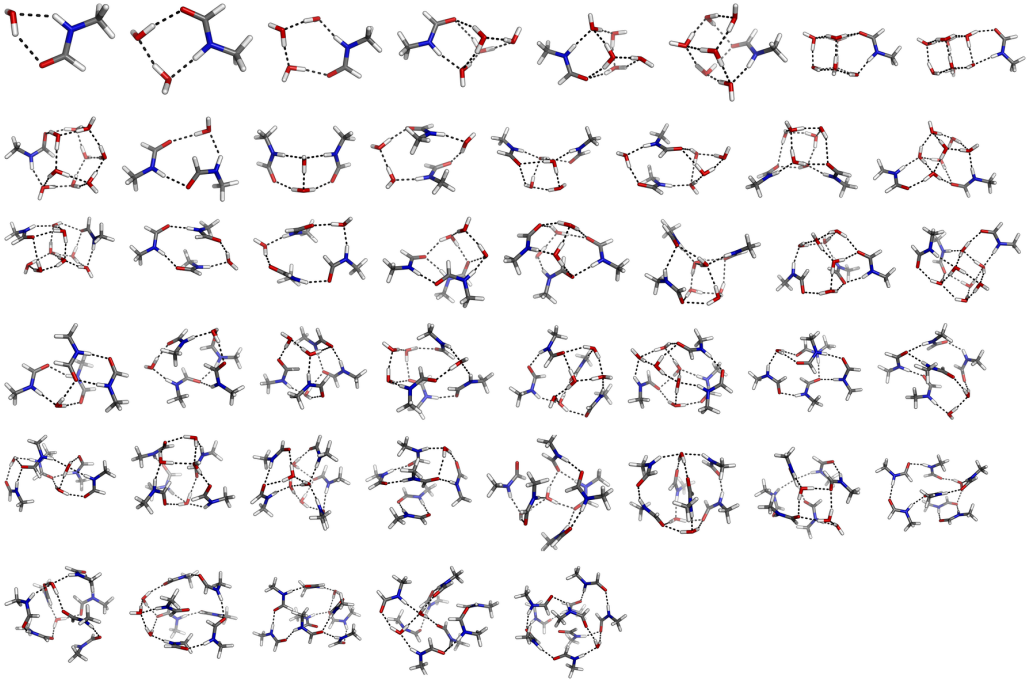


Figure S2: Mixed NMF/water cluster set. Clusters are ordered as in Table S3 and arranged in rows first. Hydrogen bonds have been identified by a loose geometric criterion and are for visualization only.

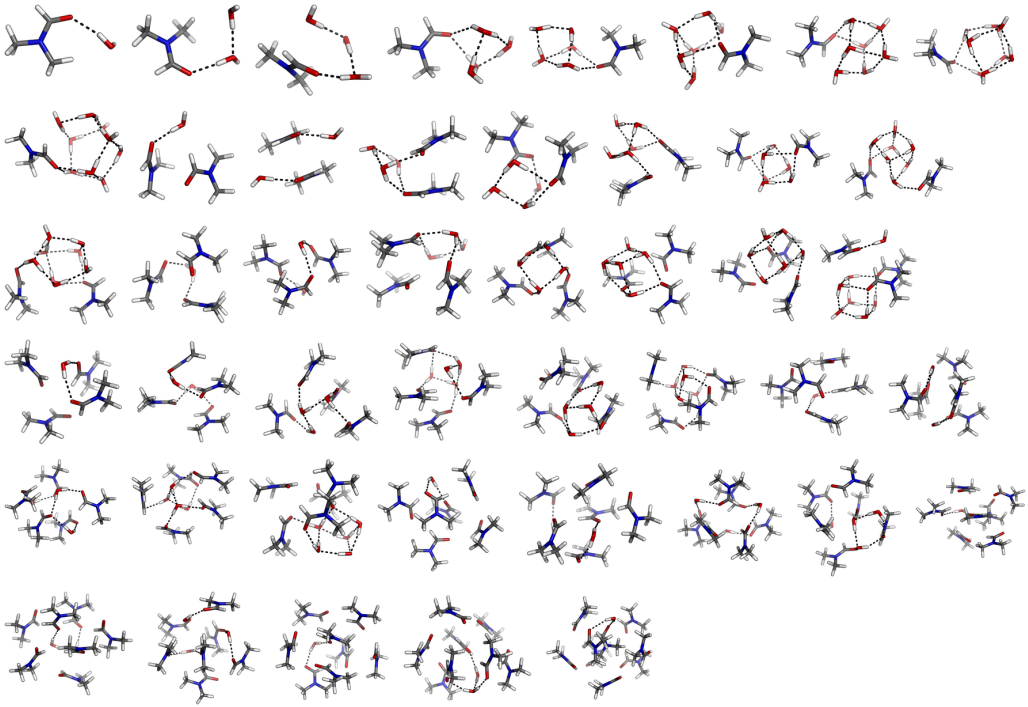


Figure S3: Mixed DMF/water cluster set. Clusters are ordered as in Table S3 and arranged in rows first. Hydrogen bonds have been identified by a loose geometric criterion and are for visualization only.

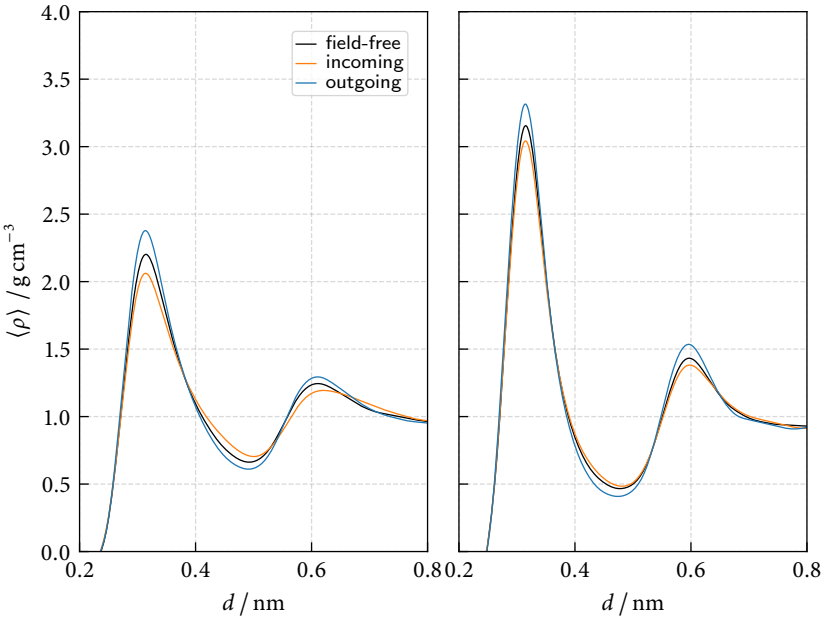
S5 Supplementary Data: Janus Interface

S5.1 Density Profiles

While macroscopic theory predicts an increase in water density (electrostriction) that is proportional to the compressibility of the liquid and the strength of the applied electric field squared,[172] the competition between orientational preferences and field alignment explained in the main text leads to an asymmetric wetting behavior at opposing surfaces[168,172] on the nanoscale. These effects are shown in Figures S4 and S5.

In the field-free systems, density profiles show the expected layering near flat surfaces,[207] which is more pronounced in hydrophilic confinement. Upon application of an electric field, asymmetries arise which are

Figure S4: Water density $\langle\rho\rangle$ as a function of d, the separation from the closest wall, shown in absence of an electric field (left) and for $E = 0.3\,\mathrm{V\,nm^{-1}}$. Left: hydrophobic confinement; right: hydrophilic confinement.



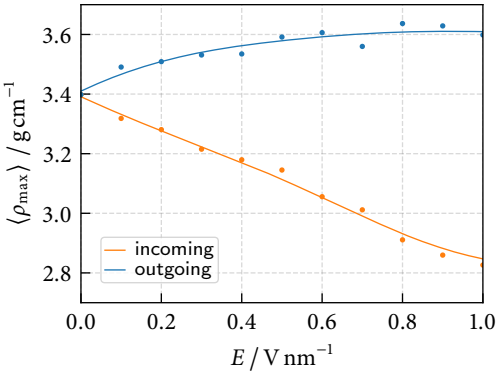


Figure S5: Maximum water density $\langle \rho_{max} \rangle$ within the first hydration layer at various incoming fields E_{in} (orange) and outgoing fields E_{out} (blue). Lines are meant to guide the eye. Hydrophilic confinement.

most pronounced in the hydration layers. Wetting is enhanced for incoming fields, but reduced for outgoing ones. By tuning the field strength, one of the walls can be rendered hydrophilic, while the other one remains hydrophobic, a situation that has been described as a field-induced[168] Janus[171] interface. The asymmetry in density becomes more pronounced with increasing field strengths (Figure S5).

For a detailed discussion of equilibrium properties, such as interfacial free energies and hydrogen bond network structure, the reader is referred to previous studies.[168,170,186] Note that at zero field, water density in the bulk-like portion of the hydrophilic system is almost that of bulk water, whereas the density in the bulk-like portion of the hydrophobic system containing the same amount of water is slightly elevated, because of comparatively weaker hydration (partial density depletion) at hydrophobic walls. The concomitant pressure change slightly affects density fluctuation dynamics, which will be described shortly.

S5.2 Dipole Angle Distributions

The orientational bias of water near interfaces does not depend on the interaction potential. In the absence of a field, $\langle \cos \phi \rangle$ is of the same magnitude at opposing walls, but differs in sign (Figure S6). Once more, orientational bias and field alignment compete, leading to an asymmetric shift of $\langle \cos \phi \rangle$ to higher values. Since the hydrophilic system derives its hydrophilicity only from enhanced Lennard-Jones interactions between

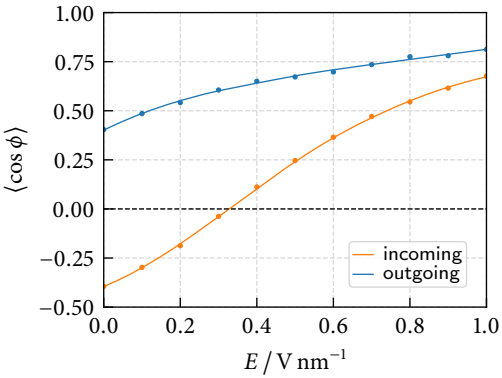


Figure S6: Field dependence of $\langle \cos \phi \rangle$, where ϕ is the angle between the collective interfacial dipole moment and the electric field **E**, applied perpendicular to the interface. Lines are meant to guide the eye. Hydrophilic confinement.

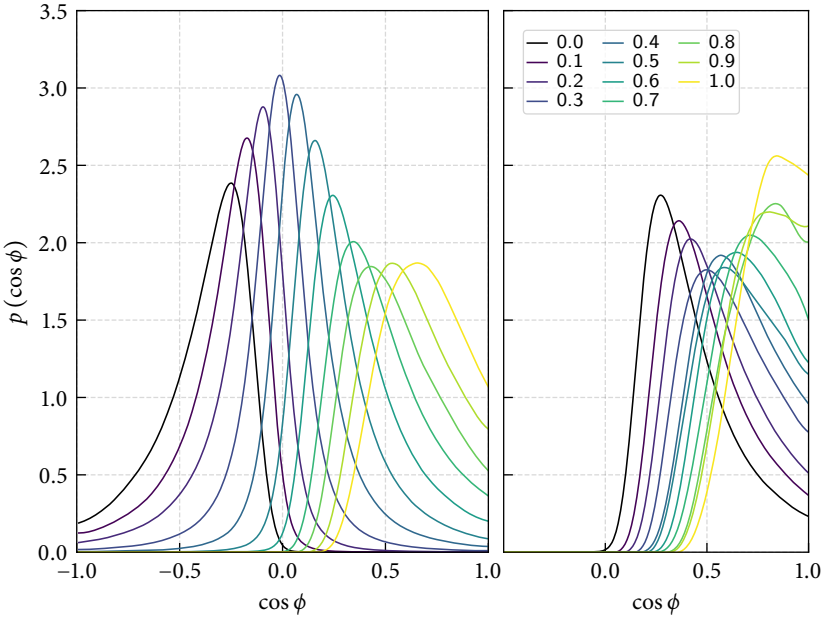


Figure S7: Probability density $p(\cos \phi)$ of the dipole angle ϕ for all investigated electric field strengths (given in V nm^{-1}). Left: incoming electric fields, and right: outgoing fields. The color gradient (blue → yellow) indicates growing field strengths. Note that the field-free distributions (black) are equal, apart from being mirrored. Hydrophilic confinement.

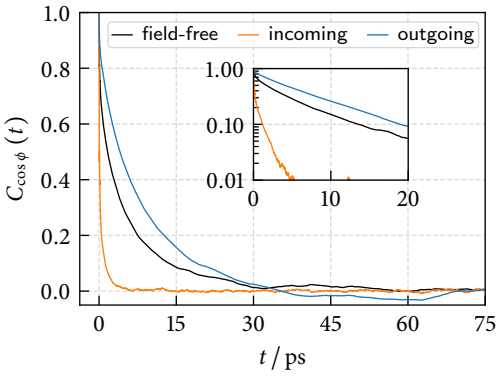


Figure S8:
Decay of dipole angle fluctuations without electric field (black), and in incoming (orange) and outgoing (blue) fields of strength 0.3 V nm^{-1}. The inset shows the correlation functions on a semi-logarithmic scale. Hydrophilic confinement.

water and the walls, Figure S6 is essentially the same as its hydrophobic counterpart (main text). The field strength at which the orientational bias is approximately canceled by the field is, however, slightly above 0.3 V nm^{-1}.

Dipole angle distributions $p(\cos\phi)$ in hydrophilic confinement are shown in Figure S7. Distributions resemble those in the hydrophobic system (main text). Their peaks are shifted systematically to higher values of $\cos\phi$ with increasing field strength. For incoming fields, their widths decrease at first, until they are narrowest around 0.3 V nm^{-1}, from where on they broaden again. For outgoing fields, they broaden monotonically with increasing field strength.

S5.3 Polarization Dynamics

In the absence of an electric field, first order collective orientational dynamics measured in terms of $C_{\cos\phi}(t)$ is slightly slower at hydrophilic walls ($t_{1/e}$ = 3.4 ps, Figures S8 and S9) than in the hydrophobic system (2.0 ps, main text). The influence of a field is analogous to the hydrophobic system. Fluctuations are monotonically slowed down in outgoing fields (≈ 8 ps), whereas in incoming fields, the trend is non-monotonic. Reorientation dynamics speeds up in relatively weak fields around 0.3 V nm^{-1}, but is slowed down again in higher fields. This trend correlates with the width in dipole angle distributions (Figure S7), as explained in the main text.

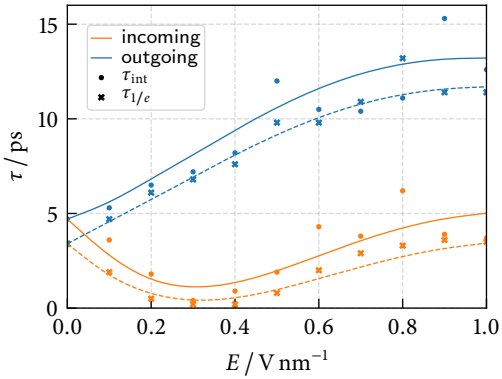


Figure S9:
Time scales $\tau_{1/e}$ (crosses, dashed lines) and τ_{int} (dots, solid lines) of the dipole angle fluctuation dynamics for incoming and outgoing fields. Lines are meant to guide the eye. Hydrophilic confinement.

S5.4 Density Fluctuation Dynamics

Density fluctuation dynamics in the first hydration layer was instigated by

$$C_N(t) = \frac{\langle \delta N(t)\, \delta N(0) \rangle}{\left\langle [\delta N(0)]^2 \right\rangle}, \tag{S9}$$

where $\delta N(t) = N(t) - \langle N \rangle$ denotes the deviation of the number of molecules within the first hydration layer N from its ensemble average $\langle N \rangle$.

No significant differences between incoming and outgoing fields can be observed (Figure S10). The decay is neither exponential, nor stretched-exponential, and cannot be described by a power law, neither at the hydrophobic nor at the hydrophilic wall. This is indicative of collective dynamics involving multiple processes. To quantify time scales, integrals $\int_0^\infty C_N(t)\, dt$ were calculated. These times are 1.3 ps in hydrophilic confinement and 1.1 ps in hydrophobic confinement.

Please note that density fluctuations between hydrophobic walls are subject to a plateau between 2 ps and 3 ps. This effect can be explained with the elevated pressure inside the hydrophobic system. In a closed system, density fluctuations in the hydration layers involve particle exchange with the bulk, which is known to be less compressible than the interfacial regions.[264] Thus, interfacial density fluctuations meet resistance which is expected to be stronger with elevated pressure. The plateau is reminiscent

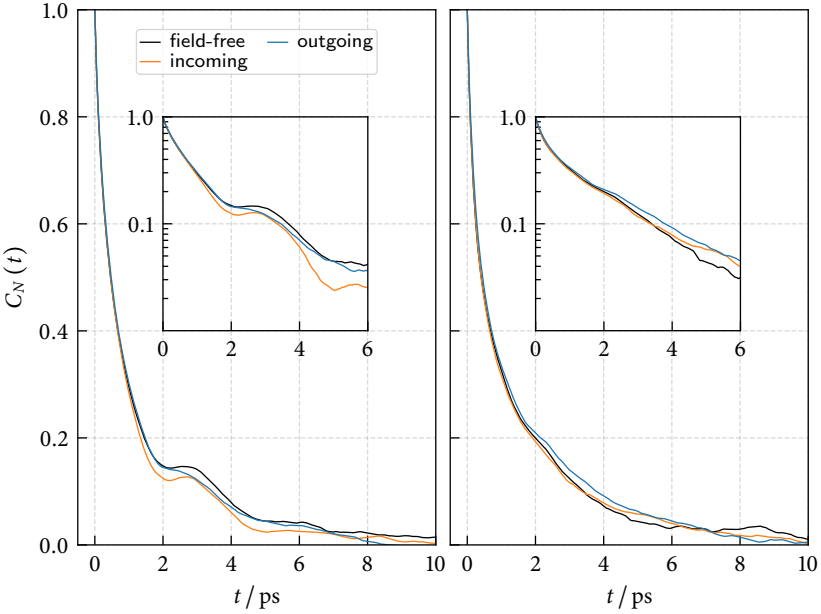


Figure S10: Decay of density fluctuations in the field-free system and for $E = 0.3\,\mathrm{V\,nm^{-1}}$ for hydrophobic confinement (left) and hydrophilic confinement (right). The inset shows the same function on a semi-logarithmic scale.

of that observed in velocity autocorrelation functions of water, where its existence is typically explained with intermolecular collisions.[297] Here, molecules push collectively against the bulk.

To verify this hypothesis, simulations were repeated in hydrophobic confinement with only 811 water molecules, leading to a bulk layer density of $0.9854\,\mathrm{g\,cm^{-3}}$. In these simulations, where the pressure is close to that of the hydrophilic system, the plateau can no longer be observed and the correlation function profiles (Figure S11) are nearly identical to those observed in the hydrophilic system (S10, right panel).

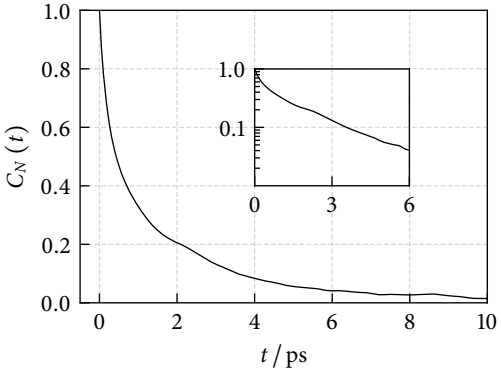


Figure S11:
Decay of density fluctuations in the field-free system with 811 water molecules (hydrophobic confinement). The inset shows the same function on a semi-logarithmic scale.

S6 Supplementary Data: Spherical Confinement

S6.1 Reorientation

In order to extract characteristic times of second order reorientation, simulation data was fit to a sum of four exponentials,

$$C_2^{\text{model}}(t) = \sum_{i=0}^{3} a_i e^{-t/\tau_i}\,, \tag{S10}$$

subject to

$$\sum_{i=0}^{3} a_i = 1\,. \tag{S11}$$

In this representation, the first two terms (i.e., parameters a_0, τ_0 and a_1, τ_1) are adequate to describe the rapid initial decay in the transient regime and the last two terms account for the remaining reorientation processes. Parameters are listed in Table S4.

Finally, characteristic times of reorientation τ_2^{reor} were calculated by

Table S4: Fit parameters a_i and τ_i/ps according to eq. (S10) in hydrophilic confinement (top) and hydrophobic confinement (bottom).

	a_0	a_1	a_2	a_3	τ_0	τ_1	τ_2	τ_3
C_{320}	0.23	0.20	0.38	0.19	0.02	0.32	2.60	8.08
C_{500}	0.21	0.18	0.37	0.24	0.02	0.33	2.48	8.67
C_{720}	0.21	0.19	0.48	0.12	0.02	0.35	2.55	9.81
C_{1500}	0.18	0.18	0.52	0.12	0.02	0.30	2.25	7.59
C_{320}	0.22	0.15	0.40	0.23	0.02	0.21	1.16	2.66
C_{500}	0.21	0.21	0.54	0.04	0.02	0.29	1.94	7.41
C_{720}	0.18	0.18	0.38	0.26	0.02	0.23	1.48	3.61
C_{1500}	0.17	0.17	0.41	0.24	0.01	0.24	1.64	3.85

integration:

$$\tau_2^{\text{reor}} = \int_0^\infty C_2^{\text{model}}(t)\, dt = \sum_{i=0}^{3} a_i \tau_i \,. \tag{S12}$$

S6.2 Diffusion

Free energy profiles are also directly accessible from counting statistics

$$V(r) = -kT \ln p(r) \,, \tag{S13}$$

where $p(r)$ is the local probability density of finding a particle at position r. The preceding equation can be evaluated by discretization and results are shown in comparison to those obtained from the Bayesian sampling procedure in Figure S12. Agreement is perfect.

Propagators in radial direction $G(r, t, |r', 0)$ are shown in Figure S13 for hydrophilic confinement. Figures S14 and S15 show selected angular propagators $P(r, \cos\theta, t|r', 0)$ in hydrophobic confinement. Other propagators are qualitatively similar. Trends have already been discussed in the main text.

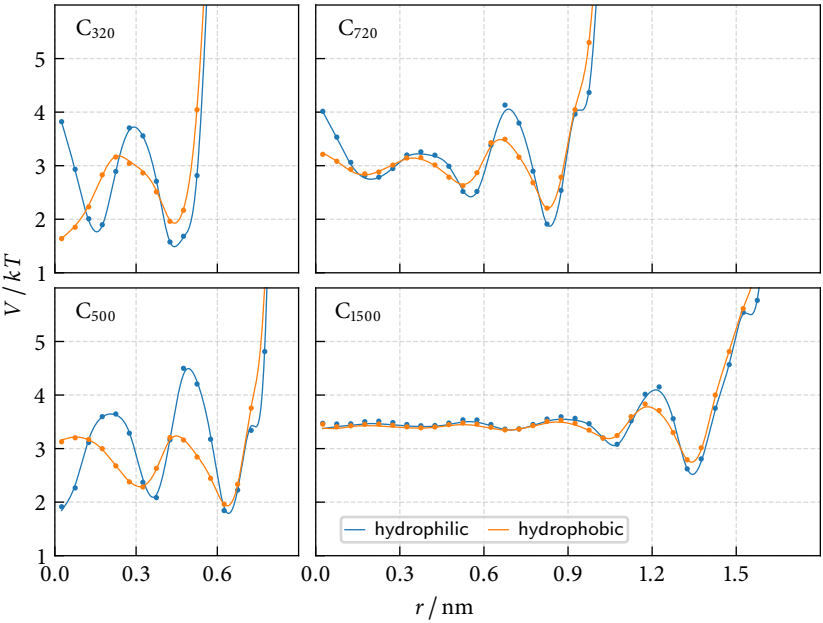


Figure S12: Free energy profiles $V(r)$ as a function of the radial distance r from the fullerene center. Results were obtained from counting statistics (solid lines) and from Bayesian sampling (dots).

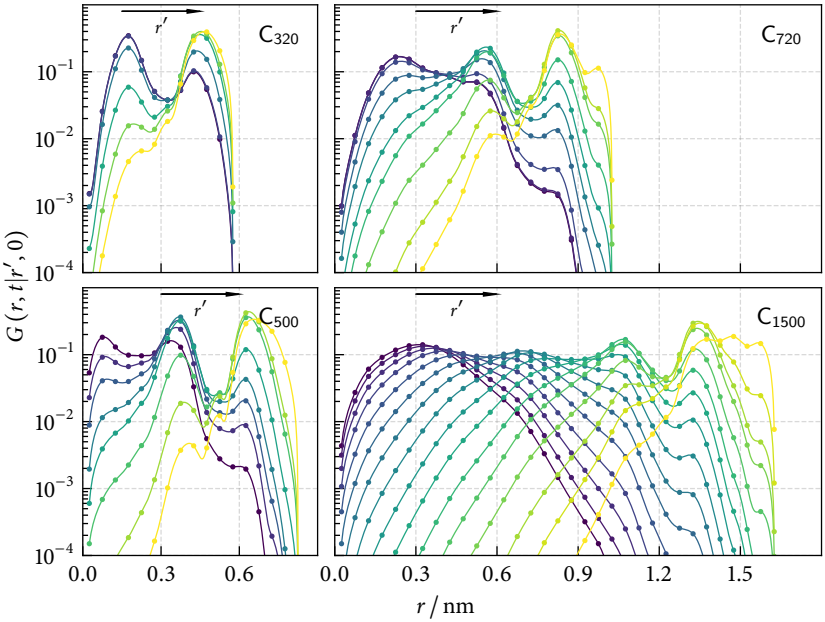


Figure S13: Model propagators $G(r, t|r', 0)$ (solid lines) compared to simulation data (symbols) in hydrophilic confinement for a lag time of 10 ps. The color gradient indicates rising r'. Lines show the model-intrinsic spline representation. Only every second bin is shown for clarity's sake (i.e., curves correspond to $r' = 0.025\,\text{nm}, 0.125\,\text{nm}, 0.225\,\text{nm}, \ldots$).

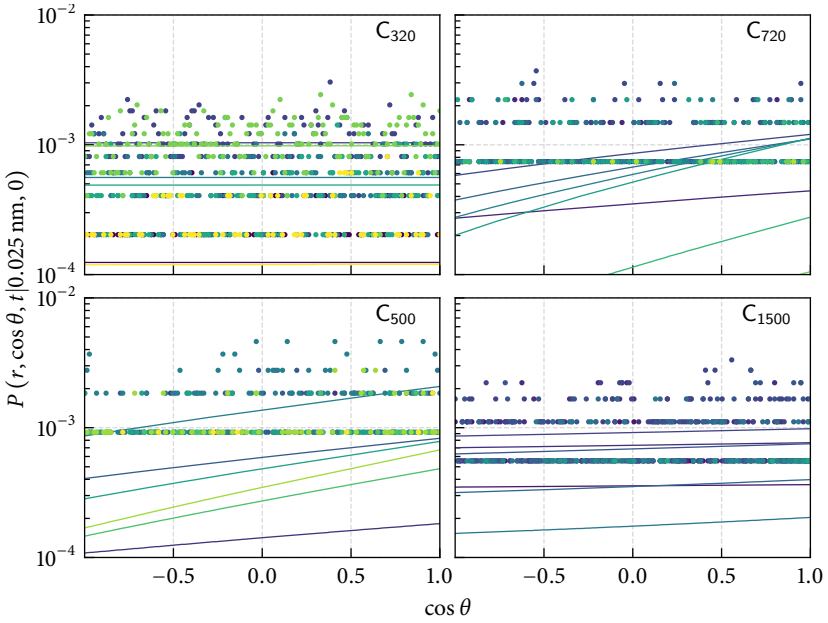


Figure S14: Model propagators $P(r, \cos\theta, t|r', 0)$ (solid lines) compared to simulation data (symbols) in hydrophobic confinement for a lag time of 10 ps and $r' = 0.025$ nm. The color gradient indicates rising r. Lines show the model-intrinsic spline representation. Only every second bin is shown for clarity's sake (i.e., curves correspond to $r = 0.025\,\text{nm}, 0.125\,\text{nm}, 0.225\,\text{nm}, \ldots$).

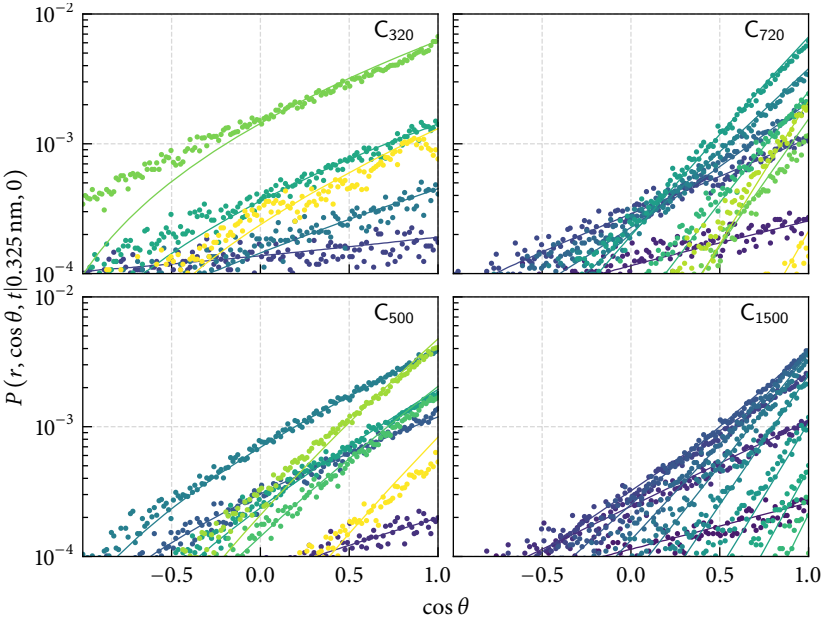


Figure S15: Model propagators $P(r, \cos\theta, t|r', 0)$ (solid lines) compared to simulation data (symbols) in hydrophobic confinement for a lag time of 10 ps and $r' = 0.225\,\text{nm}$. The color gradient indicates rising r. Lines show the model-intrinsic spline representation. Only every second bin is shown for clarity's sake (i.e., curves correspond to $r = 0.025\,\text{nm}, 0.125\,\text{nm}, 0.225\,\text{nm}, \ldots$).

S6.3 Calculation of Time Correlation Functions

Here, a few technical notes shall be given on the calculation of time correlation functions and the solution of the diffusion model (eq. (6.31)).

The time correlation functions (6.26) and (6.29) can be expanded as follows:

$$c(t) = \frac{\langle \delta h(t)\, \delta h(0) \rangle}{\left\langle [\delta h(0)]^2 \right\rangle} = \frac{\langle h(t)h(0) \rangle - \langle h \rangle^2}{\langle h \rangle (1 - \langle h \rangle)} \quad \text{and} \tag{S14}$$

$$n(t) = \frac{\langle H(t)[1 - \delta h(0)]\, \delta h(0) \rangle}{\left\langle [\delta h(0)]^2 \right\rangle} = \frac{\langle H(t)h(0) \rangle - \langle H \rangle \langle h \rangle}{\langle h \rangle (1 - \langle H \rangle)} - c(t)\,. \tag{S15}$$

Therein, the only computationally demanding terms are the averages $\langle h(t)h(0) \rangle$ and $\langle H(t)h(0) \rangle$, which represent discrete time correlation functions of the type

$$\langle A(t)B(0) \rangle = \frac{1}{T - t} \sum_{t_0=0}^{T-t} A(t + t_0)B(t_0)\,. \tag{S16}$$

Apart from the normalizing factor $1/(T - t)$ (which can be reintroduced later), such sums are efficiently accessible by exploitation of the cross-correlation theorem and use of fast discrete Fourier transforms.

Let $A \star B$ denote the cross-correlation of the functions (or time discrete time series) A and B, and $\mathcal{F}$ the (discrete) Fourier transform. Overlines shall denote the complex conjugate. Then,

$$\mathcal{F}\{A \star B\} = \overline{\mathcal{F}\{A\}} \mathcal{F}\{B\}\,, \tag{S17}$$

according to the cross-correlation theorem. Applying the reverse Fourier transform to both sides yields the desired cross-correlation function. This relationship was used in the post-processing code. The fast and parallel FFTW library was used for all Fourier transforms.[298] Zero-padding was applied to the input vectors to avoid mixing up of results due to circular convolution.

Since storing the values of $h(t)$ for each time step and pair of water

molecules is impractical, as it can easily exceed the amount of memory available to common workstations, only a list of changes in $h(t)$ is stored for each pair of molecules. The full time series is then reconstructed individually for each pair.

As shown by Luzar and Chandler, eq. (6.31) can be solved in Laplace space for

$$\tilde{k}(s) = \frac{k_f}{s + k_f + k_b s f(s)} \tag{S18}$$

and

$$\tilde{k}_{\mathrm{in}}(s) = s f(s) \tilde{k}(s)\,, \tag{S19}$$

where

$$f(s) = 3\tau_{\mathrm{d}}\left[1 - \sqrt{s\tau_{\mathrm{d}}}\arctan\left(1/\sqrt{s\tau_{\mathrm{d}}}\right)\right]. \tag{S20}$$

Eqns. (S18) and (S19) were transferred to their time domain representations by numerical inverse Laplace transformation, using an improved Talbot method.[299]

S6.4 Hydrogen Bond Kinetics

The reactive flux time correlation functions $k(t)$ are shown in Figures S16 and S17 for hydrophilic and hydrophobic confinement, respectively. Restrictive reactive flux time correlation functions $k_{\mathrm{in}}(t)$ for the same values of τ_{d} are shown in Figures S18 and S19. Correlation plots are shown in Figure S20. Note that deviations from the expected slope of unity occur only for times $t < 1.5$ ps. Finally the transient regimes of $k(t)$ are shown in Figure S21, showing no appreciable changes.

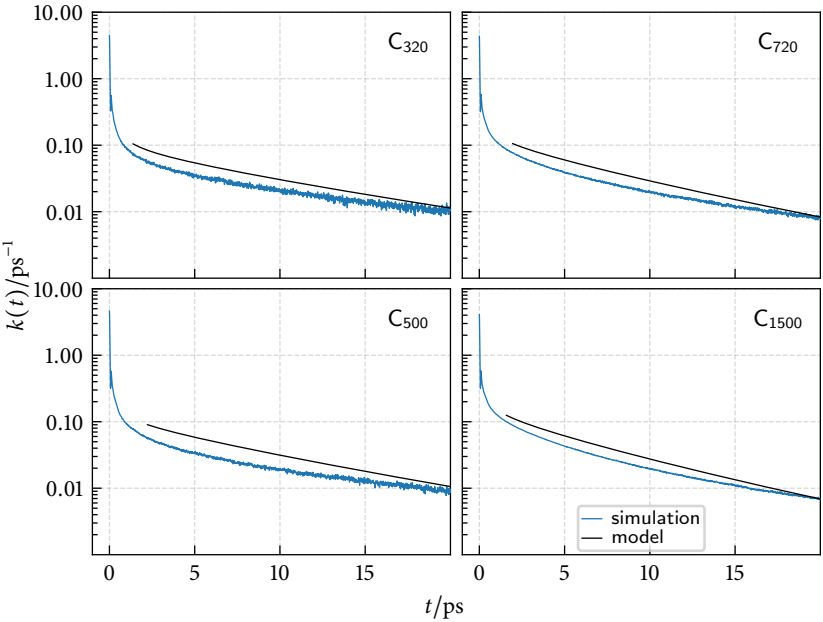


Figure S16: Reactive flux hydrogen bond time correlation functions $k(t)$ obtained from simulation and predicted by the diffusion model (eq. (6.31)). Hydrophilic confinement.

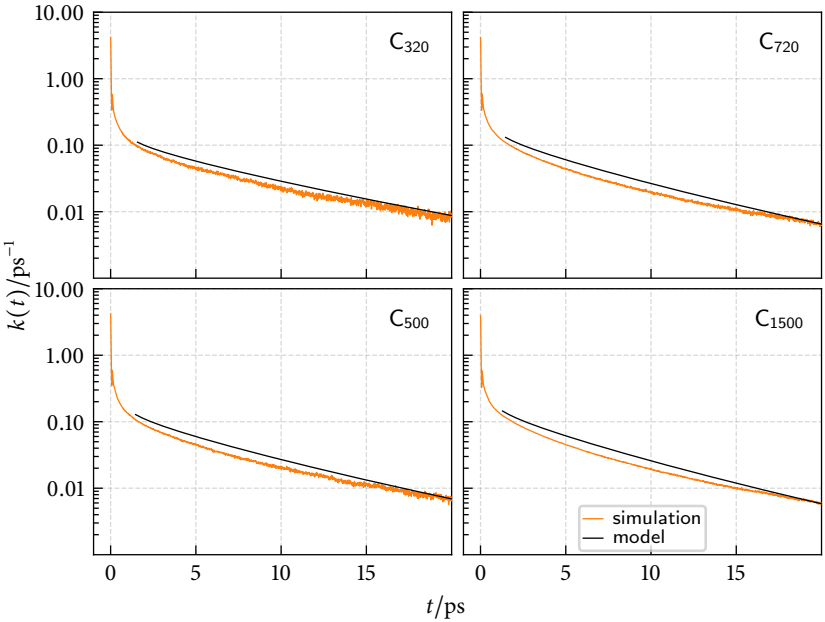


Figure S17: Reactive flux hydrogen bond time correlation functions $k(t)$ obtained from simulation and predicted by the diffusion model (eq. (6.31)). Hydrophobic confinement.

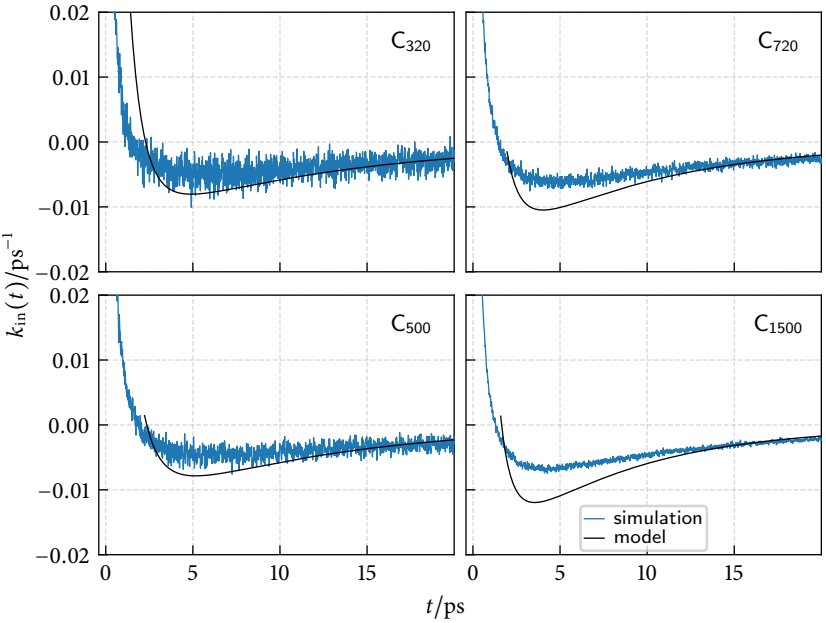


Figure S18: Restricted reactive flux hydrogen bond time correlation functions $k_{in}(t)$ obtained from simulation and predicted by the diffusion model (eq. (6.31)). Hydrophilic confinement.

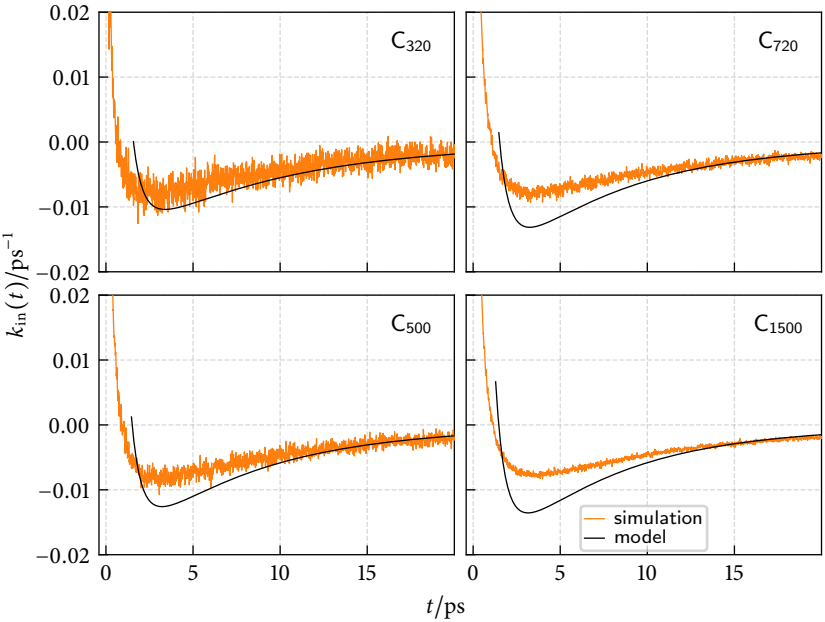


Figure S19: Restricted reactive flux hydrogen bond time correlation functions $k_{in}(t)$ obtained from simulation and predicted by the diffusion model (eq. (6.31)). Hydrophobic confinement.

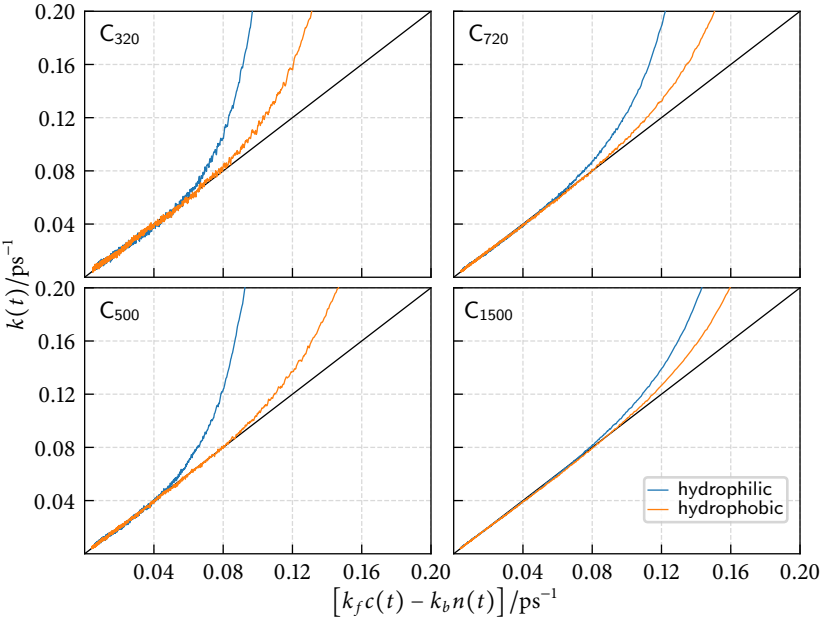


Figure S20: Hydrogen bond dynamics correlation plots in spherical nanoconfinement.

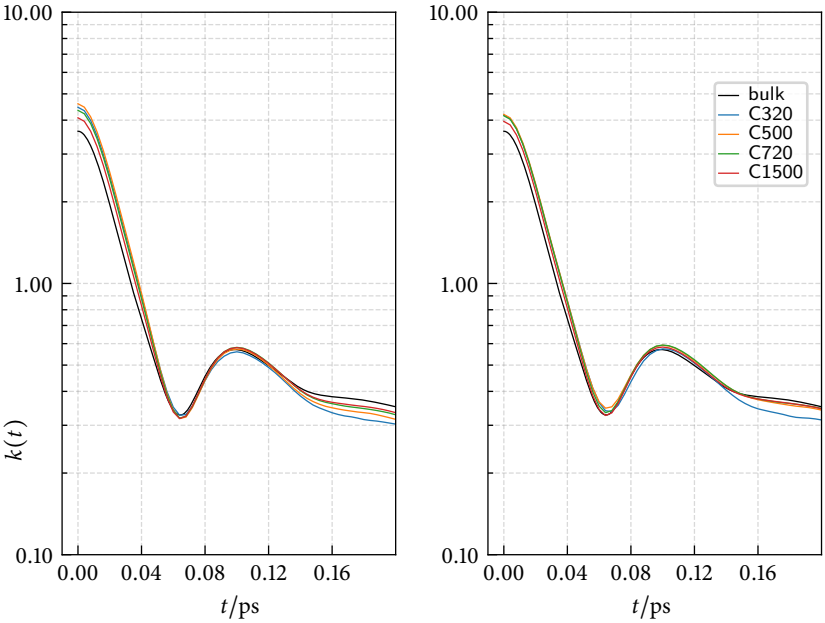


Figure S21: Reactive flux hydrogen bond time correlation functions $k(t)$. Focus on the transient regime in hydrophilic confinement (left) and hydrophobic confinement (right).

Bibliography

(1) Eisenberg, D. S.; Kauzmann, W., *The structure and properties of water*; Clarendon Press: 1969 (cit. on p. 3).

(2) Ben-Naim, A., *Water and aqueous solutions: Introduction to a molecular theory*; Plenum Press: 1974 (cit. on p. 3).

(3) Internet Encyclopedia of Philosophy: A Peer-reviewed Academic Resource. *Entry on Thalos of Miletus.* `http://www.iep.utm.edu/thales` (cit. on p. 3).

(4) Clarivate Analytics. *The Web of Science.* `https://apps.webofknowledge.com` (cit. on p. 3).

(5) Angell, C. A.; Bressel, R. D.; Hemmati, M.; Sare, E. J.; Tucker, J. C. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1559–1566 (cit. on p. 3).

(6) Dorsey, N. E., *Properties of ordinary water-substance*; Reinhold Publishing Corporation: 1940 (cit. on p. 3).

(7) Ben-Naim, A., *Molecular theory of water and aqueous solutions: Understanding water*; Part 1; World Scientific: 2009 (cit. on pp. 3, 4, 33, 159).

(8) *The physics and physical chemistry of water*; Franks, F., Ed.; Water a Comprehensive Treatise, Vol. 1; Springer US: 1972 (cit. on p. 3).

(9) *Water in crystalline hydrates: Aqueous solutions of simple nonelectrolytes*; Franks, F., Ed.; Water a Comprehensive Treatise, Vol. 2; Springer US: 1973 (cit. on p. 3).

(10) *Aqueous solutions of simple electrolytes*; Franks, F., Ed.; Water a Comprehensive Treatise, Vol. 3; Springer US: 1973 (cit. on p. 3).

(11) *Aqueous solutions of amphiphiles and macromolecules*; Franks, F., Ed.; Water a Comprehensive Treatise, Vol. 4; Springer US: 1975 (cit. on p. 3).

(12) *Water in disperse systems*; Franks, F., Ed.; Water a Comprehensive Treatise, Vol. 5; Springer US: 1975 (cit. on p. 3).

(13) *Recent advances*; Franks, F., Ed.; Water a Comprehensive Treatise, Vol. 6; Springer US: 1979 (cit. on p. 3).

(14) *Water and aqueous solutions at subzero temperatures*; Franks, F., Ed.; Water a Comprehensive Treatise, Vol. 7; Springer US: 1982 (cit. on p. 3).

(15) Robinson, G. W.; Singh, S.; Zhu, S.-B.; Evans, M. W., *Water in biology, chemistry and physics*, 1996 (cit. on p. 3).

(16) Ben-Naim, A. *J. Chem. Phys.* **1973**, *59*, 6535–6555 (cit. on p. 4).

(17) Weinhold, F. *J. Chem. Phys.* **1998**, *109*, 367–372 (cit. on pp. 4, 33, 43, 159).

(18) Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*; OUP Oxford: 2017 (cit. on pp. 4, 48, 51, 54).

(19) Barker, J. A.; Watts, R. O. *Chem. Phys. Lett.* **1969**, *3*, 144–145 (cit. on p. 4).

(20) Rahman, A.; Stillinger, F. H. *J. Chem. Phys.* **1971**, *55*, 3336–3359 (cit. on p. 4).

(21) Eisenberg, D.; Kauzmann, W., *The structure and properties of water*; Oxford University Press: 2005 (cit. on pp. 5, 77, 122, 161, 162).

(22) Teixeira, J.; Bellissent-Funel, M.-C.; Chen, S. H.; Dianoux, A. J. *Phys. Rev. A* **1985**, *31*, 1913–1917 (cit. on pp. 5, 133, 161).

(23) Perlt, E.; von Domaros, M.; Kirchner, B.; Ludwig, R.; Weinhold, F. *Sci. Rep.* **2017**, *7*, 10244 (cit. on pp. 9, 33, 60, 77, 82, 89, 91, 161).

(24) Von Domaros, M. *Water and ionic liquids in spherical nanoconfinement: A molecular dynamics simulation study*., MA thesis, Leipzig University, 2013 (cit. on pp. 10, 106, 124).

(25) Feynman, R. P., *Statistical mechanics: A set of lectures*; Frontiers in Physics; W. A. Benjamin: 1972 (cit. on p. 13).

(26) Reif, F., *Statistische Physik und Theorie der Wärme*, 3rd ed.; Muschik, W., Ed.; De Gruyter: 1987 (cit. on pp. 13, 14).

(27) McQuarrie, D. A.; Simon, J. D., *Physical chemistry: A molecular approach*, 1st ed.; University Science Books: 1997 (cit. on p. 13).

(28) McQuarrie, D. A., *Statistical mechanics*, 1st ed.; University Science Books: 2000 (cit. on pp. 13, 159, 160).

(29) Carey, V. P., *Statistical thermodynamics and microscale thermophysics*; Cambridge University Press: 1999 (cit. on p. 30).

(30) Ludwig, R.; Weinhold, F.; Farrar, T. C. *J. Chem. Phys.* **1995**, *103*, 6941–6950 (cit. on pp. 33, 133).

(31) Brüssel, M.; Perlt, E.; Lehmann, S. B. C.; von Domaros, M.; Kirchner, B. *J. Chem. Phys.* **2011**, *135*, 194113 (cit. on pp. 33, 34, 44, 46, 59, 88, 96, 159, 160).

(32) Andrews, T. *Philos. Trans. Roy. Soc. London* **1869**, *159*, 575–590 (cit. on p. 33).

(33) Weinhold, F. *J. Chem. Phys.* **1998**, *109*, 373–384 (cit. on p. 33).

(34) Ludwig, R.; Weinhold, F. *J. Chem. Phys.* **1999**, *110*, 508–515 (cit. on p. 33).

(35) Ludwig, R.; Weinhold, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1613–1619 (cit. on pp. 33, 81, 82, 86).

(36) Borowski, P.; Jaroniec, J.; Janowski, T.; Woliński, K. *Mol. Phys.* **2003**, *101*, 1413–1421 (cit. on p. 33).

(37) Kirchner, B. *J. Chem. Phys.* **2005**, *123*, 204116 (cit. on pp. 33, 59, 81, 82, 86, 88).

(38) Ludwig, R. *ChemPhysChem* **2007**, *8*, 938–943 (cit. on pp. 33, 81, 82, 86, 88).

(39) Spickermann, C.; Lehmann, S. B. C.; Kirchner, B. *J. Chem. Phys.* **2008**, *128*, 244506 (cit. on p. 33).

(40) Lenz, A.; Ojamäe, L. *J. Chem. Phys.* **2009**, *131*, 134302 (cit. on p. 33).

(41) Lehmann, S. B. C.; Spickermann, C.; Kirchner, B. *J. Chem. Theory Comput.* **2009**, *5*, 1640–1649 (cit. on p. 33).

(42) Lehmann, S. B. C.; Spickermann, C.; Kirchner, B. *J. Chem. Theory Comput.* **2009**, *5*, 1650–1656 (cit. on p. 33).

(43) Weinhold, F. *J. Phys. Chem. B* **2014**, *118*, 7792–7798 (cit. on pp. 33, 161).

(44) Ludwig, R.; Weinhold, F.; Farrar, T. C. *Mol. Phys.* **1999**, *97*, 465–477 (cit. on p. 33).

(45) Ludwig, R.; Weinhold, F.; Farrar, T. C. *Mol. Phys.* **1999**, *97*, 479–486 (cit. on p. 33).

(46) Huelsekopf, M.; Ludwig, R. *J. Mol. Liq.* **2000**, *85*, 105–125 (cit. on p. 33).

(47) Ludwig, R.; Weinhold, F. *Z. Phys. Chem.* **2002**, *216*, 659 (cit. on p. 33).

(48) Huelsekopf, M.; Ludwig, R. *J. Mol. Liq.* **2002**, *98–99*, 163–171 (cit. on p. 33).

(49) Ludwig, R. *ChemPhysChem* **2005**, *6*, 1369–1375 (cit. on p. 33).

(50) Ludwig, R. *ChemPhysChem* **2005**, *6*, 1376–1380 (cit. on p. 33).

(51) Perlt, E.; Friedrich, J.; von Domaros, M.; Kirchner, B. *ChemPhysChem* **2011**, *12*, 3474–3482 (cit. on pp. 33, 81, 88).

(52) Spickermann, C.; Perlt, E.; von Domaros, M.; Roatsch, M.; Friedrich, J.; Kirchner, B. *J. Chem. Theory Comput.* **2011**, *7*, 868–875 (cit. on p. 33).

(53) Von Domaros, M.; Perlt, E. *J. Chem. Phys.* **2017**, *146*, 124114 (cit. on pp. 33, 160).

(54) Ludwig, R.; Weinhold, F.; Farrar, T. C. *J. Phys. Chem. A* **1997**, *101*, 8861–8870 (cit. on p. 33).

(55) Ludwig, R.; Reis, O.; Winter, R.; Weinhold, F.; Farrar, T. C. *J. Phys. Chem. B* **1998**, *102*, 9312–9318 (cit. on p. 33).

(56) Huelsekopf, M.; Ludwig, R. *Magn. Reson. Chem.* **2001**, *39*, S127–s134 (cit. on p. 33).

(57) Von Domaros, M.; Jähnigen, S.; Friedrich, J.; Kirchner, B. *J. Chem. Phys.* **2016**, *144*, 064305 (cit. on p. 33).

(58) Wendt, M. A.; Weinhold, F.; Farrar, T. C. *J. Chem. Phys.* **1998**, *109*, 5945–5947 (cit. on p. 33).

(59) Hansen, M. J.; Wendt, M. A.; Weinhold, F. *Mol. Phys.* **2003**, *101*, 1147–1153 (cit. on p. 33).

(60) Ludwig, R.; Weinhold, F.; Farrar, T. C. *Berich. Bunsen Gesell.* **1998**, *102*, 197–204 (cit. on p. 33).

(61) Ludwig, R.; Weinhold, F.; Farrar, T. C. *Berich. Bunsen Gesell.* **1998**, *102*, 205–212 (cit. on p. 33).

(62) Ludwig, R.; Behler, J.; Klink, B.; Weinhold, F. *Angew. Chem. Int. Ed.* **2002**, *41*, 3199–3202 (cit. on p. 33).

(63) Matisz, G.; Fabian, W. M. F.; Kelterer, A.-M.; Kunsági-Máté, S. *J. Mol. Struct. Theochem* **2010**, *956*, 103–109 (cit. on p. 34).

(64) Matisz, G.; Kelterer, A.-M.; Fabian, W. M. F.; Kunsági-Máté, S. *J. Phys. Chem. B* **2011**, *115*, 3936–3941 (cit. on p. 34).

(65) Matisz, G.; Kelterer, A.-M.; Fabian, W. M. F.; Kunsági-Máté, S. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8467–8479 (cit. on p. 34).

(66) Ingenmey, J.; von Domaros, M.; Kirchner, B. *J. Chem. Phys.* **2017**, *146*, 154502 (cit. on pp. 34, 44, 45, 89, 91).

(67) Kirchner, B. *Phys. Reports* **2007**, *440*, 1–111 (cit. on p. 34).

(68) Kirchner, B.; Weinhold, F.; Friedrich, J.; Perlt, E.; Lehmann, S. B. C. In *Many-Electron Approaches in Physics, Chemistry and Mathematics*, Bach, V., Delle Site, L., Eds.; Springer International Publishing: 2014, pp 77–96 (cit. on pp. 34, 81, 82, 86, 88, 90).

(69) Atilhan, M.; Aparicio, S.; Hall, K. R. *Ind. Eng. Chem. Res.* **2012**, *51*, 9687–9699 (cit. on p. 45).

(70) Hansen, J.-P.; McDonald, I. R., *Theory of simple liquids: With applications to soft matter*; Elsevier Science: 2013 (cit. on p. 47).

(71) Allen, M. P. In *Computational Soft Matter: From Synthetic Polymers to Proteins*, Attig, N., Binder, K., Grubmüller, K. K., Eds., 23 vols.; Nic Series; John Von Neumann Institute for Computing: Jülich, 2004, pp 1–28 (cit. on p. 47).

(72) Berendsen, H. J. C., *Simulating the physical world*; Cambridge University Press: 2007 (cit. on p. 47).

(73) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092 (cit. on pp. 48, 140).

(74) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208–1209 (cit. on p. 48).

(75) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459–466 (cit. on p. 48).

(76) Frenkel, D.; Smit, B., *Understanding molecular simulation: From algorithms to applications*; Elsevier Science: 2001 (cit. on pp. 48, 51, 53, 54).

(77) Maitland, G. C., *Intermolecular forces: Their origin and determination*; Clarendon Press: 1981 (cit. on p. 51).

(78) Gray, C. G.; Gubbins, K. E., *Theory of molecular fluids: I. Fundamentals*; Oxford University Press: 1984 (cit. on p. 51).

(79) Sprik, M. In *Computer Simulation in Chemical Physics*, Allen, M. P., Tildesley, D. J., Eds.; Springer Netherlands: 1993, pp 211–259 (cit. on p. 51).

(80) Stone, A. J., *The theory of intermolecular forces*; Clarendon Press: 1997 (cit. on p. 51).

(81) Rahman, A. *Phys. Rev.* **1964**, *136*, A405–a411 (cit. on p. 51).

(82) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566 (cit. on p. 52).

(83) Allinger, N. L.; Chen, K.; Lii, J. H. *J. Comput. Chem.* **1996**, *17*, 642–668 (cit. on p. 52).

(84) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174 (cit. on pp. 52, 90).

(85) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666 (cit. on p. 52).

(86) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236 (cit. on p. 52).

(87) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103 (cit. on p. 53).

(88) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384–2393 (cit. on p. 55).

(89) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519 (cit. on p. 55).

(90) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697 (cit. on p. 55).

(91) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643 (cit. on p. 55).

(92) Von Domaros, M.; Perlt, E.; Kirchner, B. *Peacemaker.* `http://www.thch.uni-bonn.de/tc/peacemaker` (cit. on pp. 60, 84, 94).

(93) Stallman, R. *GNU General Public License.* `https://www.gnu.org/licenses/gpl.html` (cit. on p. 60).

(94) Pascual-ahuir, J. L.; Silla, E.; Tuñon, I. *J. Comput. Chem.* **1994**, *15*, 1127–1138 (cit. on pp. 65, 84).

(95) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169 (cit. on p. 65).

(96) Ahlrichs, R. et al. *Turbomole.* `http://www.turbomole.com` (cit. on p. 65).

(97) Neese, F. *Wires Comput. Mol. Sci.* **2012**, *2*, 73–78 (cit. on pp. 65, 82, 83, 91).

(98) Neese, F. et al. *Orca.* `http://orcaforum.cec.mpg.de` (cit. on p. 65).

(99) Press, W. H., *Numerical recipes 3rd edition: The art of scientific computing*; Cambridge University Press: 2007 (cit. on pp. 66, 68).

(100) Ingenmey, J.; von Domaros, M.; Perlt, E.; Verevkin, S. P.; Kirchner, B. *J. Chem. Phys.* **2018**, *148*, 193822 (cit. on p. 77).

(101) Sato, H.; Hirata, F. *J. Phys. Chem. A* **1998**, *102*, 2603–2608 (cit. on p. 78).

(102) Sato, H.; Hirata, F. *J. Phys. Chem. B* **1999**, *103*, 6596–6604 (cit. on p. 78).

(103) Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. *J. Phys. Chem. A* **2003**, *107*, 9380–9386 (cit. on p. 78).

(104) Yagasaki, T.; Iwahashi, K.; Saito, S.; Ohmine, I. *J. Chem. Phys.* **2005**, *122*, 144504 (cit. on p. 78).

(105) Yoshida, N.; Ishizuka, R.; Sato, H.; Hirata, F. *J. Phys. Chem. B* **2006**, *110*, 8451–8458 (cit. on p. 78).

(106) Halstead, S. J.; Masters, A. J. *Mol. Phys.* **2010**, *108*, 193–203 (cit. on p. 78).

(107) Himmel, D.; Goll, S. K.; Leito, I.; Krossing, I. *Chem. Eur. J.* **2012**, *18*, 9333–9340 (cit. on p. 78).

(108) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Phys. Chem.* **1995**, *99*, 5749–5752 (cit. on p. 78).

(109) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. *Nature* **1999**, *397*, 601–604 (cit. on p. 78).

(110) Sprik, M. *Chem. Phys.* **2000**, *258*, 139–150 (cit. on p. 78).

(111) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. *Science* **2001**, *291*, 2121–2124 (cit. on p. 78).

(112) Tuckerman, M. E.; Marx, D.; Parrinello, M. *Nature* **2002**, *417*, 925–929 (cit. on p. 78).

(113) Hassanali, A.; Prakash, M. K.; Eshet, H.; Parrinello, M. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20410–20415 (cit. on pp. 78–80, 86).

(114) Chandler, D.; Dellago, C.; Geissler, P. L. *Nat. Chem.* **2012**, *4*, 245 (cit. on pp. 78–80, 86).

(115) Agmon, N. *Chem. Phys. Lett.* **1995**, *244*, 456–462 (cit. on pp. 79, 86).

(116) Luzar, A.; Chandler, D. *Nature* **1996**, *379*, 55–57 (cit. on pp. 79, 121, 122, 134, 146, 149, 150, 154, 161, 162).

(117) Kirchner, B.; Spickermann, C.; Lehmann, S. B. C.; Perlt, E.; Langner, J.; von Domaros, M.; Reuther, P.; Uhlig, F.; Kohagen, M.; Brüssel, M. *Comput. Phys. Commun.* **2011**, *182*, 1428–1446 (cit. on pp. 81, 82, 88, 161).

(118) Weinhold, F. *J. Chem. Phys.* **1998**, *109*, 373–384 (cit. on pp. 81, 82, 86).

(119) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654 (cit. on p. 82).

(120) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. *J. Comput. Chem.* **1983**, *4*, 294–301 (cit. on p. 82).

(121) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627 (cit. on p. 82).

(122) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170 (cit. on p. 82).

(123) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305 (cit. on p. 82).

(124) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104 (cit. on p. 82).

(125) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456–1465 (cit. on p. 82).

(126) Kruse, H.; Grimme, S. *J. Chem. Phys.* **2012**, *136*, 154101 (cit. on pp. 82, 84).

(127) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. *J. Chem. Phys.* **2015**, *143*, 054107 (cit. on pp. 82, 91).

(128) Dunning Jr., T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023 (cit. on pp. 82, 83).

(129) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2016**, *144*, 024109 (cit. on p. 83).

(130) Liakos, D. G.; Neese, F. *J. Phys. Chem. A* **2012**, *116*, 4801–4816 (cit. on p. 83).

(131) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566 (cit. on p. 83).

(132) National Institute of Standards and Technology. *Chemistry webbook*. `http://webbook.nist.gov/cgi/cbook.cgi?ID=C7732185` (cit. on p. 84).

(133) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451 (cit. on p. 84).

(134) Marshall, W. L.; Franck, E. U. *J. Phys. Chem. Ref. Data* **1981**, *10*, 295–304 (cit. on pp. 84, 85).

(135) Spickermann, C., *Entropies of condensed phases and complex systems: A first principles approach*; Springer Berlin Heidelberg: 2011 (cit. on pp. 89, 99).

(136) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 205–211 (cit. on p. 89).

(137) Kuntz Jr., I. D.; Kauzmann, W. *Adv. Protein. Chem.* **1974**, *28*, 239–345 (cit. on p. 89).

(138) Russell, R. A.; Thompson, H. W. *Spectrochim. Acta* **1956**, *8*, 138–141 (cit. on p. 89).

(139) LaPlanche, L. A.; Rogers, M. T. *J. Am. Chem. Soc.* **1964**, *86*, 337–341 (cit. on p. 89).

(140) Aroney, M. J.; Le Fevre, R. J. W.; Singh, A. N. *J. Chem. Soc.* **1965**, 3179–3184 (cit. on p. 89).

(141) Nakanishi, H.; Roberts, J. D. *Org. Magn. Resonance* **1981**, *15*, 7–12 (cit. on p. 89).

(142) Radzicka, A.; Pedersen, L.; Wolfenden, R. *Biochemistry (Mosc.)* **1988**, *27*, 4538–4541 (cit. on p. 89).

(143) Shin, S.; Kurawaki, A.; Hamada, Y.; Shinya, K.; Ohno, K.; Tohara, A.; Sato, M. *J. Mol. Struct.* **2006**, *791*, 30–40 (cit. on p. 89).

(144) Case, D. A. et al. *Amber.* `http://www.ambermd.org` (cit. on p. 90).

(145) Dieterich, J. M.; Hartke, B. *Mol. Phys.* **2010**, *108*, 279–291 (cit. on p. 90).

(146) Buck, U.; Pradzynski, C. C.; Zeuch, T.; Dieterich, J. M.; Hartke, B. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6859–6871 (cit. on p. 90).

(147) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. *J. Phys. Chem. A* **2001**, *105*, 10525–10537 (cit. on p. 91).

(148) Qian, P.; Song, W.; Lu, L.; Yang, Z. *Int. J. Quantum Chem.* **2010**, *110*, 1923–1937 (cit. on p. 91).

(149) Egorov, G. I.; Makarov, D. M. *J. Chem. Eng. Data* **2017**, *62*, 1247–1256 (cit. on p. 94).

(150) Rumble, J., *CRC handbook of chemistry and physics*, 98th ed.; CRC Press LLC: 2017 (cit. on p. 94).

(151) Islam, M. N.; Ali, M. A.; Monirul Islam, M.; Nahar, M. K. *Phys. Chem. Liq.* **2003**, *41*, 271–282 (cit. on p. 94).

(152) Bernal-García, J. M.; Guzmán-López, A.; Cabrales-Torres, A.; Estrada-Baltazar, A.; Iglesias-Silva, G. A. *J. Chem. Eng. Data* **2008**, *53*, 1024–1027 (cit. on p. 94).

(153) Wagner, W.; Pruß, A. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387–535 (cit. on p. 94).

(154) Marcus, Y., *Solvent mixtures: Properties and selective solvation*; Taylor & Francis: 2002 (cit. on pp. 97–100).

(155) Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. *J. Phys. Chem. Lett.* **2014**, *5*, 4275–4284 (cit. on pp. 102, 160).

(156) Benjamin, I. In *Reviews in Computational Chemistry*; John Wiley & Sons, Inc: 2015; Vol. 28, pp 205–313 (cit. on pp. 103, 126).

(157) Björneholm, O.; Hansen, M. H.; Hodgson, A.; Liu, L.-M.; Limmer, D. T.; Michaelides, A.; Pedevilla, P.; Rossmeisl, J.; Shen, H.; Tocci, G.; Tyrode, E.; Walz, M.-M.; Werner, J.; Bluhm, H. *Chem. Rev.* **2016**, *116*, 7698–7726 (cit. on p. 103).

(158) Shapiro, B.; Moon, H.; Garrell, R. L.; Kim, C.-j. *J. Appl. Phys.* **2003**, *93*, 5794–5811 (cit. on pp. 103, 104).

(159) Mugele, F.; Baret, J.-C. *J. Phys.: Condens. Matter* **2005**, *17*, R705–r774 (cit. on pp. 103, 104).

(160) Baret, J.-C.; Mugele, F. *Phys. Rev. Lett.* **2006**, *96*, 016106 (cit. on p. 103).

(161) Krupenkin, T.; Taylor, J. A.; Kolodner, P.; Hodes, M. *Bell Labs Tech. J.* **2005**, *10*, 161–170 (cit. on p. 103).

(162) Chen, J. Y.; Kutana, A.; Collier, C. P.; Giapis, K. P. *Science* **2005**, *310*, 1480–1483 (cit. on p. 103).

(163) Berge, B.; Peseux, J. *Eur. Phys. J. E* **2000**, *3*, 159–163 (cit. on p. 103).

(164) Krupenkin, T.; Yang, S.; Mach, P. *Appl. Phys. Lett.* **2003**, *82*, 316–318 (cit. on p. 103).

(165) Murade, C. U.; Oh, J. M.; an den Ende, D.; Mugele, F. *Opt. Express* **2011**, *19*, 15525–15531 (cit. on p. 103).

(166) Frank, H. S. *J. Chem. Phys.* **1955**, *23*, 2023 (cit. on p. 104).

(167) Striolo, A. *Adsorpt. Sci. Technol.* **2011**, *29*, 211–258 (cit. on p. 104).

(168) Bratko, D.; Daub, C. D.; Leung, K.; Luzar, A. *J. Am. Chem. Soc.* **2007**, *129*, 2504–2510 (cit. on pp. 104, 105, 107, 124, 162, 177, 178).

(169) Daub, C. D.; Bratko, D.; Leung, K.; Luzar, A. *J. Phys. Chem. C* **2007**, *111*, 505–509 (cit. on pp. 104, 107, 162).

(170) Bratko, D.; Daub, C. D.; Luzar, A. *Faraday Discuss.* **2009**, *141*, 55–66 (cit. on pp. 104, 105, 107, 162, 178).

(171) Zhang, X.; Zhu, Y.; Granick, S. *Science* **2002**, *295*, 663–666 (cit. on pp. 104, 162, 178).

(172) Daub, C. D.; Bratko, D.; Luzar, A. *Top. Curr. Chem.* **2012**, *307*, 155–180 (cit. on pp. 104, 105, 107, 112, 162, 177).

(173) Uchida, K.; Izumi, N.; Sukata, S.; Kojima, Y.; Nakamura, S.; Irie, M. *Angew. Chem. Int. Ed.* **2006**, *45*, 6470–6473 (cit. on p. 104).

(174) Lahann, J.; Mitragotri, S.; Tran, T.-N.; Kaido, H.; Sundaram, J.; Choi, I. S.; Hoffer, S.; Somorjai, G. A.; Langer, R. *Science* **2003**, *299*, 371–374 (cit. on p. 104).

(175) Lim, H. S.; Han, J. T.; Kwak, D.; Jin, M.; Cho, K. *J. Am. Chem. Soc.* **2006**, *128*, 14458–14459 (cit. on p. 104).

(176) Wan, Z.; Zeng, H.; Feinerman, A. *Appl. Phys. Lett.* **2006**, *89*, 201107 (cit. on p. 104).

(177) Hsieh, J.; Mach, P.; Cattaneo, F.; Yang, S.; Krupenkine, T.; Baldwin, K.; Rogers J., A. *IEEE Photonic. Tech. Lett.* **2003**, *15*, 81–83 (cit. on p. 104).

(178) Daub, C. D.; Bratko, D.; Ali, T.; Luzar, A. *Phys. Rev. Lett.* **2009**, *103*, 207801 (cit. on pp. 105, 107).

(179) Bonthuis, D. J.; Gekle, S.; Netz, R. R. *Langmuir* **2012**, *28*, 7679–7694 (cit. on pp. 105, 108).

(180) Werder, T.; Walther, J. H.; Jaffe, R. L.; Halicioglu, T.; Koumoutsakos, P. *J. Phys. Chem. B* **2003**, *107*, 1345–1352 (cit. on pp. 106, 124).

(181) Wang, J.; Bratko, D.; Luzar, A. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6374–6379 (cit. on p. 106).

(182) Hautman, J.; Klein, M. L. *Phys. Rev. Lett.* **1991**, *67*, 1763–1766 (cit. on p. 106).

(183) Daub, C. D.; Bratko, D.; Luzar, A. *J. Phys. Chem. C* **2011**, *115*, 22393–22399 (cit. on pp. 106, 107).

(184) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271 (cit. on pp. 107, 122, 125).

(185) Vaitheeswaran, S.; Yin, H.; Rasaiah, J. C. *J. Phys. Chem. B* **2005**, *109*, 6629–6635 (cit. on p. 107).

(186) Bratko, D.; Daub, C. D.; Luzar, A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6807–6813 (cit. on pp. 107, 110, 117, 178).

(187) Guillot, B. *J. Mol. Liq.* **2002**, *101*, 219–260 (cit. on p. 107).

(188) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. *Faraday Discuss.* **2009**, *141*, 251–276 (cit. on p. 107).

(189) Van der Spoel, D.; Van Maaren, P. J.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *108*, 10220 (cit. on pp. 107, 108).

(190) Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954–9960 (cit. on p. 107).

(191) Vega, C.; de Miguel, E. *J. Chem. Phys.* **2007**, *126*, 154707 (cit. on p. 107).

(192) Chandler, D. *Nature* **2005**, *437*, 640–646 (cit. on p. 107).

(193) Patel, A. J.; Varilly, P.; Chandler, D. *J. Phys. Chem. B* **2010**, *114*, 1632–1637 (cit. on p. 107).

(194) Bhide, S. Y.; Berkowitz, M. L. *J. Chem. Phys.* **2005**, *123*, 224702 (cit. on p. 107).

(195) Chowdhary, J.; Ladanyi, B. M. *J. Phys. Chem. B* **2006**, *110*, 15442–15453 (cit. on p. 107).

(196) Chowdhary, J.; Ladanyi, B. M. *J. Phys. Chem. B* **2008**, *112*, 6259–6273 (cit. on pp. 107, 114, 116, 120).

(197) Chowdhary, J.; Ladanyi, B. M. *J. Phys. Chem. B* **2009**, *113*, 4045–4053 (cit. on pp. 107, 114, 116, 150).

(198) Johnson, M. E.; Malardier-Jugroot, C.; Murarka, R. K.; Head-Gordon, T. *J. Phys. Chem. B* **2009**, *113*, 4082–4092 (cit. on p. 107).

(199) Argyris, D.; Cole, D. R.; Striolo, A. *J. Phys. Chem. C* **2009**, *113*, 45 (cit. on p. 107).

(200) Lambeth Jr., B. P.; Junghans, C.; Kremer, K.; Cecilia, C.; Delle Site, L. *J. Chem. Phys.* **2010**, *133*, 221101 (cit. on p. 107).

(201) Rosenfeld, D. E.; Schmuttenmaer, C. A. *J. Phys. Chem. B* **2011**, *115*, 1021–1031 (cit. on pp. 107, 120).

(202) Malani, A.; Ayappa, K. G. *J. Chem. Phys.* **2012**, *136*, 194701 (cit. on p. 107).

(203) Ho, T. A.; Striolo, A. *J. Chem. Phys.* **2013**, *138*, 054117 (cit. on p. 107).

(204) Brodsky, A. *Chem. Phys. Lett.* **1996**, *261*, 563–568 (cit. on p. 107).

(205) Baranyai, A.; Kiss, P. T. *J. Chem. Phys.* **2010**, *133*, 144109 (cit. on p. 107).

(206) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. B* **2013**, *117*, 9956–9972 (cit. on p. 107).

(207) Lee, C.-Y.; Mccammon, J. A.; Rossky, P. J. *J. Chem. Phys.* **1984**, *80*, 4448–4455 (cit. on pp. 107, 109, 162, 177).

(208) Shelley, J. C.; Patey, G. N. *Mol. Phys.* **1996**, *88*, 385–398 (cit. on pp. 107, 109, 162).

(209) Stillinger, F. H. In *The liquid state of matter: Fluids, simple and complex*, Montroll, E. W., Lebowitz, J. L., Eds.; North-Holland Publishing Company: 1982, pp 341–431 (cit. on p. 107).

(210) Saitta, A. M.; Saija, F.; Giaquinta, P. V. *Phys. Rev. Lett.* **2012**, *108*, 207801 (cit. on p. 107).

(211) Stuve, E. M. *Chem. Phys. Lett.* **2012**, *519-520*, 1–17 (cit. on p. 107).

(212) Tainter, C. J.; Pieniazek, P. A.; Lin, Y.-S.; Skinner, J. L. *J. Chem. Phys.* **2011**, *134*, 184501 (cit. on p. 108).

(213) Boresch, S.; Steinhauser, O. *Berich. Bunsen Gesell.* **1997**, *101*, 1019–1029 (cit. on p. 108).

(214) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 3169–3174 (cit. on p. 108).

(215) Svishchev, I. M.; Kusalik, P. G.; Wang, J.; Boyd, R. J. *J. Chem. Phys.* **1996**, *105*, 4742–50 (cit. on p. 108).

(216) Zasetsky, A. Y.; Petelina, S. V.; Lyashchenko, A. K.; Lileev, A. S. *J. Chem. Phys.* **2010**, *133*, 134502 (cit. on p. 108).

(217) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1–19 (cit. on pp. 108, 126).

(218) Plimpton, S. et al. *Lammps.* `http://lammps.sandia.gov` (cit. on p. 108).

(219) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341 (cit. on pp. 108, 125).

(220) Luzar, A.; Svetina, S.; Žekš, B. *Studia Biophysica* **1982**, *91*, 89–90 (cit. on pp. 109, 162).

(221) Luzar, A.; Svetina, S.; Žekš, B. *Chem. Phys. Lett.* **1983**, *96*, 485–490 (cit. on pp. 109, 162).

(222) Stirnemann, G.; Rossky, P. J.; Hynes, J. T.; Laage, D. *Faraday Discuss.* **2010**, *146*, 263–281 (cit. on pp. 109, 115, 119, 134, 162).

(223) Sutmann, G. *J. Electroanal. Chem.* **1998**, *450*, 289–302 (cit. on pp. 110, 117).

(224) Schweighofer, K. J.; Benjamin, I. *J. Electroanal. Chem.* **1995**, *391*, 1–10 (cit. on pp. 110, 117).

(225) Suresh, S. J.; Satish, A. V.; Choudhary, A. *J. Chem. Phys.* **2006**, *124*, 074506 (cit. on pp. 110, 117).

(226) Suresh, S. J. *J. Chem. Phys.* **2007**, *126*, 2722745 (cit. on pp. 110, 117).

(227) Kiselev, M.; Heinzinger, K. *J. Chem. Phys.* **1996**, *105*, 650–657 (cit. on pp. 110, 117).

(228) Maerzke, K. A.; Siepmann, J. I. *J. Phys. Chem. B* **2010**, *114*, 4261–4270 (cit. on pp. 110, 117).

(229) Dzubiella, J.; Allen, R. J.; Hansen, J.-P. *J. Chem. Phys.* **2004**, *120*, 5001–5004 (cit. on p. 110).

(230) Wu, J.; Bratko, D.; Prausnitz, J. M. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15169–15172 (cit. on p. 110).

(231) Smith, J. D.; Saykally, R. J.; Geissler, P. L. *J. Am. Chem. Soc.* **2007**, *129*, 13847–13856 (cit. on p. 110).

(232) Benjamin, I. *J. Phys. Chem. B* **2005**, *109*, 13711–13715 (cit. on pp. 114, 116).

(233) Milischuk, A. A.; Ladanyi, B. M. *J. Chem. Phys.* **2011**, *135*, 174709 (cit. on pp. 114, 116, 120).

(234) Luzar, A. *Faraday Discuss.* **1996**, *103*, 29–40 (cit. on p. 114).

(235) Lee, S. H.; Rossky, P. J. *J. Chem. Phys.* **1994**, *100*, 3334–3345 (cit. on p. 115).

(236) Farrer, R. A.; Fourkas, J. T. *Acc. Chem. Res.* **2003**, *36*, 605–612 (cit. on p. 115).

(237) Saito, S.; Ohmine, I. *J. Chem. Phys.* **1994**, *101*, 6063 (cit. on p. 115).

(238) Castrillón, S. R.-V.; Giovambattista, N.; Aksay, I. A.; Debenedetti, P. G. *J. Phys. Chem. B* **2009**, *113*, 1438–1446 (cit. on pp. 116, 119).

(239) Pal, S. K.; Peon, J.; Bagchi, B.; Zewail, A. H. *J. Phys. Chem. B* **2002**, *106*, 12376–12395 (cit. on p. 120).

(240) Rasaiah, J. C.; Garde, S.; Hummer, G. *Annu. Rev. Phys. Chem.* **2008**, *59*, 713–740 (cit. on p. 120).

(241) Paranthaman, S.; Coudert, F.-X.; Fuchs, A. H. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8124–8130 (cit. on p. 120).

(242) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. *Nature* **2001**, *414*, 188 (cit. on p. 120).

(243) Hanasaki, I.; Nakatani, A. *J. Chem. Phys.* **2006**, *124*, 144708 (cit. on p. 120).

(244) Hummer, G. *Mol. Phys.* **2007**, *105*, 201–207 (cit. on p. 120).

(245) Liu, P.; Harder, E.; Berne, B. J. *J. Phys. Chem. B* **2004**, *108*, 6595–6602 (cit. on pp. 120, 154).

(246) Liu, P.; Harder, E.; Berne, B. J. *J. Phys. Chem. B* **2005**, *109*, 2949–2955 (cit. on p. 120).

(247) Fábián, B.; Senćanski, M. V.; Cvijetić, I. N.; Jedlovszky, P.; Horvai, G. *J. Phys. Chem. C* **2016**, *120*, 8578–8588 (cit. on p. 120).

(248) Mitra, T.; Miró, P.; Tomsa, A.-R.; Merca, A.; Bögge, H.; Ávalos, J.; Poblet, J. M.; Bo, C.; Müller, A. *Chem. Eur. J.* **2009**, *115*, 1844–1852 (cit. on p. 120).

(249) Garcia-Ratés, M.; Miró, P.; Poblet, J. M.; Bo, C.; Avalos, J. B. *J. Phys. Chem. B* **2011**, *19*, 5980–5992 (cit. on p. 120).

(250) Moilanen, D. E.; Levinger, N. E.; Spry, D. B.; Fayer, M. D. *J. Am. Chem. Soc.* **2007**, *129*, 14311–14318 (cit. on p. 120).

(251) Park, S.; Moilanen, D. E.; Fayer, M. D. *J. Phys. Chem. B* **2008**, *112*, 5279–5290 (cit. on p. 120).

(252) Faeder, J.; Ladanyi, B. M. *J. Phys. Chem. B* **2000**, *104*, 1033–1046 (cit. on p. 120).

(253) Fayer, M. D. *Acc. Chem. Res.* **2012**, *45*, 3–14 (cit. on pp. 120, 121).

(254) Van der Loop, T. H.; Ottosson, N.; Lotze, S.; Kentzinger, E.; Vad, T.; Sager, W. F. C.; Bakker, H. J.; Woutersen, S. *J. Chem. Phys.* **2014**, *141*, 535 (cit. on p. 120).

(255) Cringus, D.; Bakulin, A.; Lindner, J.; Vöhringer, P.; Pshenichnikov, M. S.; Wiersma, D. A. *J. Phys. Chem. B* **2007**, *111*, 14193–14207 (cit. on pp. 120, 121, 132).

(256) Cringus, D.; Lindner, J.; Milder, M. T. W.; Pshenichnikov, M. S.; Vöhringer, P.; Wiersma, D. A. *Chem. Phys. Lett.* **2005**, *408*, 162–168 (cit. on pp. 121, 132).

(257) Kayal, A.; Chandra, A. *Mol. Simul.* **2015**, *41*, 463–470 (cit. on pp. 121, 128, 132, 155).

(258) Luzar, A. *J. Chem. Phys.* **2000**, *113*, 10663 (cit. on pp. 121, 122, 129, 130, 134, 146, 147, 149, 150, 154, 161).

(259) Schwerdtfeger, P.; Wirz, L.; Avery, J. *J. Comput. Chem.* **2013**, *34*, 1508–1526 (cit. on p. 122).

(260) Fowler, P. W.; Manolopoulos, D. E., *An atlas of fullerenes*; Clarendon Press: 1995 (cit. on pp. 122, 123).

(261) Wu, Z. C.; Jelski, D. A.; George, T. F. *Chem. Phys. Lett.* **1987**, *137*, 291–294 (cit. on p. 122).

(262) Babic, D.; Balaban, A. T.; Klein, D. J. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 515–526 (cit. on p. 123).

(263) Norman, G. E.; Filinov, V. S. *High Temp.* **1969**, *7*, 216–222 (cit. on p. 124).

(264) Bratko, D.; Curtis, R. A.; Blanch, H. W.; Prausnitz, J. M. *J. Chem. Phys.* **2001**, *115*, 3873–3877 (cit. on pp. 124, 181).

(265) Chau, P.-L.; Hardwick, A. J. *Mol. Phys.* **1998**, *93*, 511–518 (cit. on pp. 128–130).

(266) Woutersen, S.; Emmerichs, U.; Bakker, H. J. *Science* **1997**, *278*, 658 (cit. on p. 133).

(267) Bakker, H. J.; Skinner, J. L. *Chem. Rev.* **2010**, *110*, 1498–1517 (cit. on p. 133).

(268) Laage, D.; Stirnemann, G.; Sterpone, F.; Rey, R.; Hynes, J. T. *Annu. Rev. Phys. Chem.* **2011**, *62*, 395–416 (cit. on p. 133).

(269) Shen, Y. R. *Annu. Rev. Phys. Chem.* **1989**, *40*, 327–350 (cit. on p. 133).

(270) Shen, Y. R. *Nature* **1989**, *337*, 519–525 (cit. on p. 133).

(271) Jonas, J.; Defries, T.; Wilbur, D. J. *J. Chem. Phys.* **1976**, *65*, 582–588 (cit. on p. 133).

(272) Ropp, J.; Lawrence, C.; Farrar, T. C.; Skinner, J. L. *J. Am. Chem. Soc.* **2001**, *123*, 8047–8052 (cit. on p. 133).

(273) Qvist, J.; Halle, B. *J. Am. Chem. Soc.* **2008**, *130*, 10345–10353 (cit. on p. 133).

(274) Di Cola, D.; Deriu, A.; Sampoli, M.; Torcini, A. *J. Chem. Phys.* **1996**, *104*, 4223–4232 (cit. on p. 133).

(275) Barthel, J.; Bachhuber, K.; Buchner; Hetzenauer, H. *Chem. Phys. Lett.* **1990**, *165*, 369–373 (cit. on p. 133).

(276) Rønne, C.; Åstrand, P.-O.; Keiding, S. R. *Phys. Rev. Lett.* **1999**, *82*, 2888–2891 (cit. on p. 133).

(277) Winkler, K.; Lindner, J.; Bürsing, H.; Vöhringer, P. *J. Chem. Phys.* **2000**, *113*, 4674–4682 (cit. on p. 133).

(278) Tan, H.-S.; Piletic, I. R.; Fayer, M. D. *J. Chem. Phys.* **2005**, *122*, 174501 (cit. on p. 133).

(279) Fecko, C. J.; Loparo, J. J.; Roberts, S. T.; Tokmakoff, A. *J. Chem. Phys.* **2005**, *122*, 054506 (cit. on p. 133).

(280) Rezus, Y. L. A.; Bakker, H. J. *J. Chem. Phys.* **2005**, *123*, 114502 (cit. on p. 133).

(281) Laage, D.; Hynes, J. T. *Science* **2006**, *311*, 832–835 (cit. on pp. 133, 134).

(282) Laage, D.; Hynes, J. T. *J. Phys. Chem. B* **2008**, *112*, 14230–14242 (cit. on pp. 133, 134, 136).

(283) Smith, D. W. G.; Powles, J. G. *Mol. Phys.* **1966**, *10*, 451–463 (cit. on p. 133).

(284) Lankhorst, D.; Schriever, J.; Leyte, J. C. *Berich. Bunsen Gesell.* **1982**, *86*, 215–221 (cit. on p. 133).

(285) Van der Maarel, J. R. C.; Lankhorst, D.; de Bleijser, J.; Leyte, J. C. *Chem. Phys. Lett.* **1985**, *122*, 541–544 (cit. on p. 133).

(286) Debye, P. J. W., *Polar molecules*; Dover Publications: 1929 (cit. on p. 134).

(287) Mittal, J.; Hummer, G. *J. Chem. Phys.* **2012**, *137*, 034110 (cit. on pp. 137, 138, 154).

(288) Hummer, G. *New J. Phys.* **2005**, *7*, 34 (cit. on pp. 137, 139, 140, 154).

(289) Sriraman, S.; Kevrekidis, I. G.; Hummer, G. *J. Phys. Chem. B* **2005**, *109*, 6479–6484 (cit. on pp. 137, 139, 140).

(290) Tazi, S.; Boţan, A.; Salanne, M.; Marry, V.; Turq, P.; Rotenberg, B. *J. Phys.: Condens. Matter* **2012**, *24*, 284117 (cit. on p. 141).

(291) Yeh, I.-C.; Hummer, G. *J. Phys. Chem. B* **2004**, *108*, 15873–15879 (cit. on p. 141).

(292) Rapaport, D. C. *Mol. Phys.* **1983**, *50*, 1151–1162 (cit. on p. 147).

(293) Chandler, D. *J. Chem. Phys.* **1978**, *68*, 2959 (cit. on p. 148).

(294) Xu, H.; Stern, H. A.; Berne, B. J. *J. Phys. Chem. B* **2002**, *106*, 2054–2060 (cit. on p. 150).

(295) Gardiner, C., *Stochastic methods: A handbook for the natural and social sciences*; Springer Berlin Heidelberg: 2009 (cit. on p. 150).

(296) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107–115 (cit. on p. 161).

(297) Bauer, B. A.; Ou, S.; Patel, S.; Siva, K. *Phys. Rev. E* **2012**, *85*, 051506 (cit. on p. 182).

(298) Frigo, M.; Johnson, S. G. *Proc. IEEE* **2005**, *93*, 216–231 (cit. on p. 190).

(299) Dingfelder, B.; Weideman, J. A. C. *arXiv e-prints* **2013**, 1304.2505 (cit. on p. 191).

Curriculum Vitae

Publications

(13) J. Ingenmey, **M. von Domaros**, E. Perlt, S. Verevkin, and B. Kirchner: "Thermodynamics and proton activities of protic ionic liquids with quantum cluster equilibrium theory."
J. Chem. Phys. (2018), **148**, 193822.

(12) S. Gehrke, **M. von Domaros**, R. Clark, O. Hollóczki, M. Brehm, T. Welton, A. Luzar, and B. Kirchner: "Structure and lifetimes in ionic liquids and their mixtures."
Faraday Discuss. (2018), **206**, 219–245.

(11) E. Perlt, **M. von Domaros**, B. Kirchner, R. Ludwig, and F. Weinhold: "Predicting the Ionic Product of Water."
Sci. Rep. (2017), **7**, 10244.

(10) **M. von Domaros**, E. Perlt: "Anharmonic effects in the quantum cluster equilibrium method."
J. Chem. Phys. (2017), **146**, 124114.

(9) J. Ingenmey, **M. von Domaros**, and B. Kirchner: "Predicting miscibility of binary liquids from small cluster QCE calculations."
J. Chem. Phys. (2017), **146**, 154502.

(8) **M. von Domaros**, S. Jähnigen, J. Friedrich, and B. Kirchner: "Quantum cluster equilibrium model of N-methylformamide–water binary mixtures."
J. Chem. Phys. (2016), **144**, 064305.

(7) D. Czurlok, **M. von Domaros**, M. Thomas, J. Gleim, J. Lindner, B. Kirchner, and P. Vöhringer: "Femtosecond 2DIR spectroscopy of the nitrile stretching vibration of thiocyanate anions in liquid-to-supercritical heavy water. Spectral diffusion and libration-induced hydrogen-bond dynamics."
Phys. Chem. Chem. Phys. (2015), **17**, 29776–29785.

(6) **M. von Domaros**, D. Bratko, B. Kirchner, and A. Luzar: "Dynamics at a Janus interface."
J. Phys. Chem. C (2013), **117**, 4561–4567.

(5) M. Brüssel, E. Perlt, **M. von Domaros**, M. Brehm, and B. Kirchner: "A one-parameter quantum cluster equilibrium approach."
J. Chem. Phys. (2012), **137**, 164107.

(4) M. Brüssel, E. Perlt, S. B. C. Lehmann, **M. von Domaros**, and B. Kirchner: "Binary systems from quantum cluster equilibrium theory."
J. Chem. Phys. (2011), **135**, 194113.

(3) B. Kirchner, C. Spickermann, S. B. C. Lehmann, E. Perlt, J. Langner, **M. von Domaros**, P. Reuther, F. Uhlig, M. Kohagen, and M. Brüssel: "What can clusters tell us about the bulk? Peacemaker: Extended quantum cluster equilibrium calculations."
Comput. Phys. Commun. (2011), **182**, 1428–1446.

(2) E. Perlt, J. Friedrich, **M. von Domaros**, and B. Kirchner: "Importance of structural motifs in liquid hydrogen fluoride."
ChemPhysChem (2011), **12**, 3474–3482.