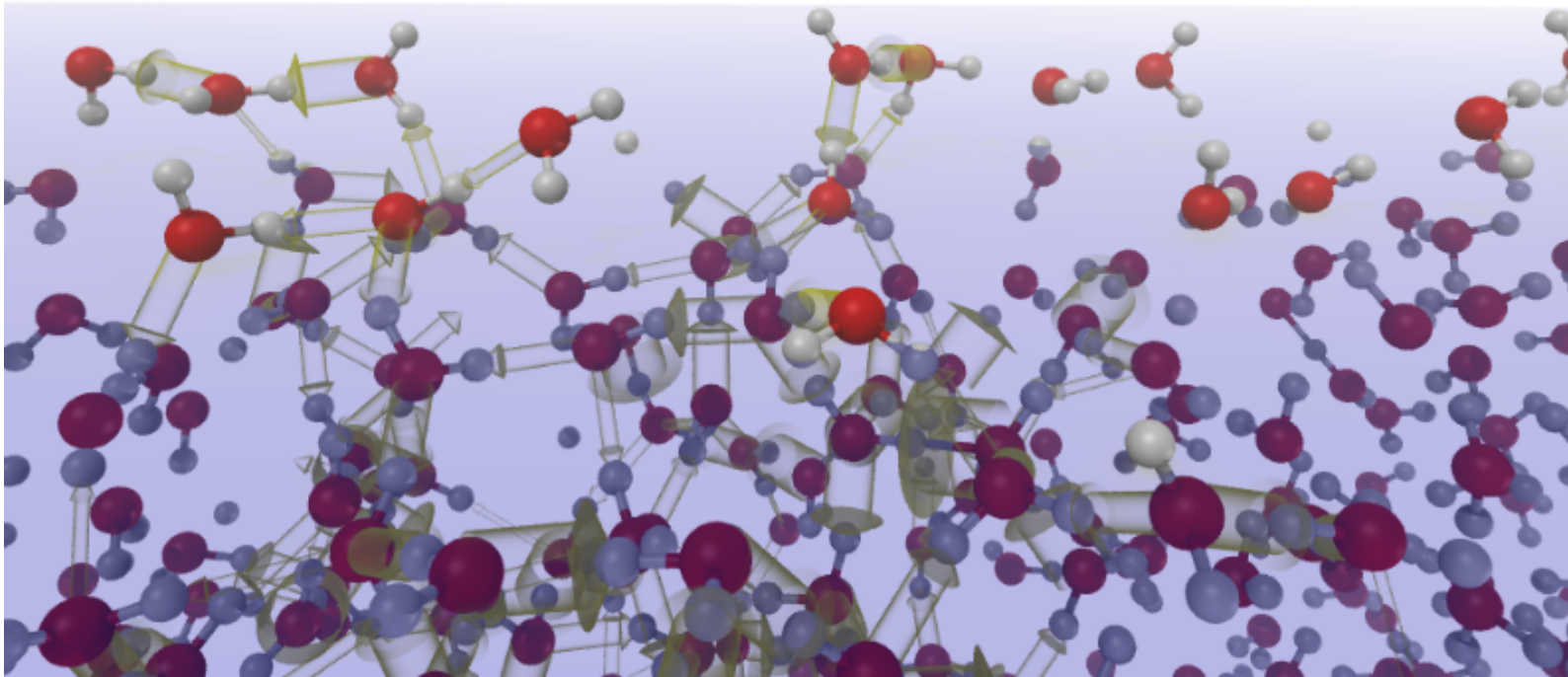
Energetics and Chemistry of Hydrogen Bonds in Water and at the Surface

Thesis to acquire the title Doktor der Naturwissenschaften

Submitted to
the University of Paderborn by:
Kristof Karhan



dedicated to Melanie

Contents

English Abstract

Water's ability to form hydrogen bonds is at the core of many of its astounding properties. Cooperation, or cooperativity, describes the notion that that a molecule forms additional hydrogen bonds more readily, when it is already hydrogen bonded. However, to this day the intricacies of cooperation in real water have not been elucidated. In this work, we set out to improve the current understanding of hydrogen bond cooperation by using a proper basis for its description, namely, its energy. To this end, we use data of the energy decomposition analysis of a molecular dynamics simulation of bulk water and the water surface, which has been been tuned to represent physical properties of real water at ambient temperature as accurately as possible. The energy decomposition analysis, ALMO EDA, offers a way to uniquely determine the energies of hydrogen bonds. We find and quantify a direct dependence of the energy of a hydrogen bond on the energy of other bonds. We further elucidate how cooperation and bond energy alters features of two surface specific spectroscopic methods, namely sum-frequency generation and x-ray absorption.

German Abstract

Die Fähigkeit des Wassers Wasserstoffbrückenbindungen auszubilden, hängt direkt mit vielen seiner erstaunlichen Eigenschaften zusammen. Kooperation, oder Kooperativität, bezeichnet nun die Beobachtung, dass ein Wassermolekül eher weiter Wasserstoffbrückenbindungen ausbildet, wenn es bereits über solcherart gebunden ist. Allerdings wurde bis jetzt das Wirken von Kooperation in realem Wasser nicht nachvollzogen. In dieser Arbeit gedenken wir das herrschende Verständnis von Kooperation zu verbessern, indem wir dessen Beschreibung auf ein solides Fundament stellen: Der Energie einer Wasserstoffbrückenbindung. Zu diesem Zwecke benutzen wir die Daten einer Energieverteilungsanalyse. Diese wurden aus einer Molekulardynamiksimulation gewonnen, welche derart gestaltet war, dass die die physikalischen Eigenschaften realen Wassers bei Raumtemperatur so gut wie möglich nachgebildet werden. Die verwendeted Energieverteilungsanalyse, ALMO EDA, bietet eine eindeutige Bestimmung der Energie einer Wasserstoffbrückenbindung. Wir beobachten und quantifizieren eine direkte Abhängigkeit zwischen der Energie einer Wasserstoffbrückenbindung und der Energie der sie umgebenden Bindungen. Weiter klären wir, wie Kooperation und Bindungsenergien sich auf die Signale zweier oberflächen-spezifischer Methoden auswirken.

1 Introduction

Despite the abundance of literature on the subject [1, 2, 3, 4, 5, 6, 7], the complex properties of water, which are rooted in its ever shifting hydrogen bond network, are yet to be fully understood [8]. Among the open questions concerning the hydrogen bond network are: How the phenomenon of cooperation alters hydrogen bonds, as well as the influence of the surface on the network and how these changes affect spectroscopic properties.

Cooperation, also known as cooperativity, between individual bonds is essential to the strength of the hydrogen bond network. Hence, the description of the liquid water state requires a proper understanding of cooperation. The phenomenon was first discribed by Frank and Wan [9], who stated the following: The existence of a hydrogen bond between two molecules increases the basicity on the acceptor as well as the acidity on the donor. Hence, both molecules are more susceptible to forming further hydrogen bonds, see Fig. 1.1 for a graphical depiction. Through the study of water trimers, Stillinger [10, 11] identified both cooperative and anti-cooperative behavior. Cooperation and anti-cooperation were deduced from the total energy of the trimers, thus relating the phenomenon to the bond energy. The former behavior is found between a donor and an acceptor bond, whereas the latter is found between two donor or two acceptor bonds. Since then, a plethora of articles, both experimental [12, 13, 14] and theoretical [15, 8, 16, 17] has been published indicating the influence of cooperation on a wide variety of properties. Water clusters in particular have attained special interest. Due their continually increasing size, and experimental accessibility, they can be used to monitor the transition of properties from the single molecule to the bulk value [18]. One well known example

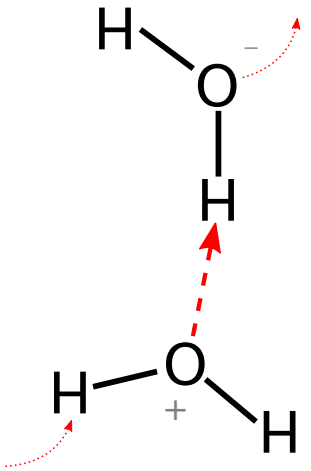


Figure 1.1: Schematic of charge separation through hydrogen bonds. Red curved lines show enhanced H-Bond formation capabilities

of such studies is the length of hydrogen bonds in water. Here, the study of clusters bridges the gap between the O-O distance in the dimer and bulk water. It was shown that, at 0 K, water molecules move closer together as the cluster size grows [18, 19]. As the dimer clearly lacks cooperation and the bulk clearly has it [13], the interpretation of these results saw cooperation as their source. Simultaneously, energy decomposition methods have been used together with static quantum mechanical calculations to investigate the energetic contributions to the bonding in the entirety of a cluster [20, 21].

Exploiting the computational efficiency of force fields, hydrogen bond networks have been investigated under ambient conditions. As its rich phase space demonstrates, temperature is crucial for the behavior of the hydrogen bond network [15]. Thus, a finite temperature simulation greatly enhances estimates of the properties of water. Using the additive AMOEBA [22] force field, Agmon and coworkers have observed that water molecules with a high amount of acceptor hydrogen bonds also tend to have a high amount of donor hydrogen bonds [23, 24].

Recently, scientfic interest has turned to the surface of water, not at least because of its suspected catalytic abilities [25, 26]. The premier experimental tools for investigation of the surface are Vibrational Sum Frequency Generation (VSFG, SFG) [27, 6, 7] and X-Ray Absorption (XA) [28, 29]. The surface sensitivity of SFG stems from the fact that the average orientation of the molecules with respect to the surface must not be zero [27]. In addition only probing the surface, the quantitiy measured in SFG, the susceptibility χ, features an imaginary part, which contains information regarding the average surface orientation of the vibrating groups [28]. The SFG-spectrum of the pure water surface features broad peaks at around 3200 cm^{-1} and 3450 cm^{-1}. Since these two frequencies resemble the O-H stretch frequency of ice and liquid water, respectively, the two SFG signals have been termed “ice-like” and “liquid-like”. In the wake of this finding the idea of an ice-like layer at the surface had been brought up [30], but such a layer has not been identified [27, 7]. Furthermore, even the existence of two distinct signals has been brought into question by groups around Bonn and Sovago [31, 32, 33, 34]. However, even to date the nature of the SFG signals remains under debate [7, 28]. Within this debate, Gao and coworkers presented a paper in which they identified a third signal at 3117 cm^{-1}. Through the imaginary part, they identified this signal as originating from upward facing O-H bonds in the surface, whereas the signals at 3222 cm^{-1} supposedly originated from downward facing molecules at the top of the surface, and the signal at 3450 cm^{-1} originated from deeper within the slab. The infrared frequency of a O-H vibration

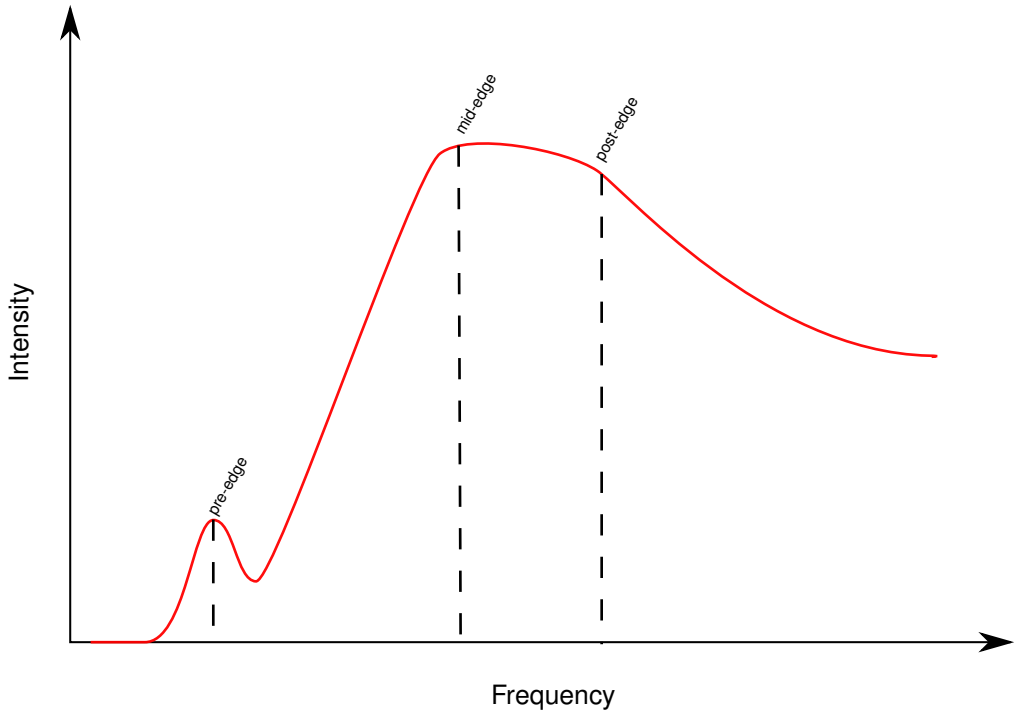


Figure 1.2: Schematic of an x-ray absorption spectrum of water

can be related to the length of an attached hydrogen bond, such that the longer the distance, the lower the frequency is [35]. Since the hydrogen bond length is strongly correlated to the hydrogen bond energy [36], the investigation of hydrogen bond energies can aid the understanding of the SFG results.

The second powerful tool available to probe the surface is x-ray absorption spectroscopy (XAS) [29]. The method is highly sensitive to changes of the virtual orbitals, which relate closely to the environment around the excited atom [37]. Due to the high frequency of the molecules, the lifetime of an excited molecule is of the order of a few femtoseconds, effectively resulting in a static picture of a molecule [38]. Therefore, this method is destined to monitor the constituency and structural surroundings of the molecule in a quasi-static setup [38]. Despite the huge prescision of the method, the interpretation of an XA spectrum is far from trivial [29]. Thus XA experiments are often accompanied by extensive theoretical work [39]. The water XA spectrum features three main peaks, titled pre-edge, mid-edge and post-edge. Previous work has led to increasing consensus that the pre-edge feature of the water XA spectrum can be related to broken hydrogen bonds. Several interpretations exist for the post edge feature, which is prominent in ice I_h. Many studies suggest that it is the result of strong hydrogen bonding, whereas the significance of the mid-edge feature is still under debate [29]. As previous studies suggest that the acceptor interactions of a

molecule have a more significant effect on its XA spectrum [40, 38, 41], this work will focus on those.

In Chapter 4, we use the energy decomposition method established by Khalliulin et al [42, 43, 44] in conjunction with an extensive ab-initio molecular dynamics simulation. The use of ab-initio methods allows for the accurate incorporation of the inherently non-additive cooperation effect, while ALMO-EDA enables the separation of bonding charge-transfer from non-bonding polarization interactions. Additionally MD ensures that the results hold for expansive water under ambient conditions. Thus, we here present the first account of the energetic impact hydrogen bonds have on each other under realistic circumstances. Chapter 5 transfers the newly established concept of cooperation to the surface and its validity for surface molecules is investigated. The link between energy and frequency, and the knowledge about the orientation of molecules relative to the surface, link the underlying energies to the SFG results. Finally, we make use of the electronic asymmetry parameter γ [45]. Developed to descibe hydrogen bonding in bulk, it was used for the interpretation of bulk water XA spectra. The usage of this parameter is motivated by the existing observations for the pre-edge and post-edge peak. The pre-edge peak is associated with broken molecules that feature broken hydrogen bonds, which necessitates a high asymmetry on that molecule. Furthermore, the post-edge peak is prominent in ice I_h and a previous study has indicated that the average asymmetry in ice is lower than in water [46]. Using the acceptor asymmetry γ_{Acc} for the reasons laid out above, we present a new avenue to the interpretation of surface XA spectra.

2 Theory

2.1 Theoretical descripton of atomic systems

The analytical solution of the Schrödinger equation can only be obtained for the simplest of systems, where the hydrogen atom belongs to the most complex problems. Hence, approximate methods have to be applied in order to calculate properties of real systems. Even then, the quantum mechanical methods become computationally unfeasible for tens of atoms in case of coupled cluster methods, or hundreds of atoms in case of density-functional theory [47]. The latter will be presented in more detail in section 2.1.3. However, many physical systems of interest, especially proteins easily contain several hundred thousand atoms, requiring classical parametric methods for computation. A brief introduction thereof will be given in section 2.1.4. These methods, by default, cannot capture chemical reactions. Methods coupling classical systems to quantum systems (QM / MM) have been developed to circumvent the deficiencies and combine the strengths of both former approaches; see section 2.1.5.

2.1.1 The Schrödinger Equation and the Born-Oppenheimer Approximation

The outset for every quantum mechanical approach is the non-relativistic, time-independent Schrödinger equation (SE),

$$\hat{H}\Psi_{\mathrm{eK}}(\mathbf{r}_1,\cdots,\mathbf{r}_N,\mathbf{R}_1,\cdots,\mathbf{R}_M) = E\Psi_{\mathrm{eK}}(\mathbf{r}_1,\cdots,\mathbf{r}_N,\mathbf{R}_1,\cdots,\mathbf{R}_M)\,, \tag{2.1}$$

obeyed by every stationary quantum mechanical system. Here, $\hat{H}$ is the Hamilton-, or energy operator, E is the observable energy of the system. $\mathbf{r},\mathbf{R}$ are the positions of the electrons and nuclei, respectively and N,M are the number of electrons and nuclei [48]. $\hat{H}$ is a linear operator that can be split up into the following contributions

$$\hat{H} = \sum_{I<J}^{M}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|} + \sum_{Ii}^{M} N\frac{Z_I}{\mathbf{R}_I-\mathbf{r}_i} + \sum_{i<j}^{N}\frac{1}{\mathbf{r}_i-\mathbf{r}_j} + \sum_{I}^{M} -\frac{1}{2M_I}\nabla_I + \sum_{i}^{N} -\frac{1}{2}\nabla_i\,, \tag{2.2}$$

where the capital letters I, J are indexes running over all nuclei, the letters i, j are indexes running over all electrons and M_I is the mass of nucleus I. The first two terms of Eq. 2.2 are the nucleus-nucleus $\hat{V}_{\mathrm{KK}}$ and nucleus electron potential interaction $\hat{V}_{\mathrm{Ke}}$, the third is the electron-electron interaction $\hat{U}$, the fourth the nucleic kinetic energy $\hat{T}_{\mathrm{K}}$, and the fifth the electron kinetic energy $\hat{T}$. The wavefunction Ψ, which depends on all spacial coordinates of the system describes said system completely, but is unknown [48]. Nuclei move on timescales much longer than electrons, because of the different masses. For instance, even the mass of the lightest nucleus m_p is about $1830 m_e$, where m_e is the electron mass. Thus, from the electrons' perspective, the nuclei appear to be frozen. In contrast, from the nuclei's perspective, the rapid movement of the electrons averages itself out [48, 49]. This approximate independece of the motion of the nuclei and electron leads to the following approximate ansatz for wavefunction

$$\Psi_{\mathrm{eK}}(\mathbf{r}, \mathbf{R}) = \Psi(\mathbf{r}; \mathbf{R}) X(\mathbf{R}) \,. \tag{2.3}$$

Here, X is the nucleic wave function and Ψ is the electronic wave function that depends parametrically on the position of the nuclei [49]. This separation leads to the following set of coupled equations:

$$\left(\hat{T} + \hat{U} + \hat{V}_{\mathrm{Ke}}\right) \Psi = E \Psi \quad \text{(electronic SE)} \tag{2.4}$$

$$(\hat{T}_{\mathrm{K}} + \hat{V}_{\mathrm{KK}} + E) X = E_{\mathrm{tot}} X \quad \text{(nucleic-SE)} \,. \tag{2.5}$$

Here, E is the electronic energy and E_{tot} is the total energy of the system. For fixed nuclei, the term $\hat{V}_{\mathrm{KK}} X$ reduces to a constant and can be combined with $\hat{V}_{\mathrm{Ke}}$ into an external potential operator $\hat{v}_{\mathrm{ext}}$ to obtain the total energy of the system.

2.1.2 Constructing Imaginable Wavefunctions

Even Eq. 2.4 faces the issue of being unsolveable for more than two independent but inseperable variables [49]. Moreover, the complexity of the wavefunction makes it hard to imagine it and thus makes it difficult to work with. In the context of a molecular system, the wavefunction is therefore modeled to be constructed from one-electron orbitals φ, also called molecular orbitals, which can then be considered seperately. If these φ are then used to construct a so-called *Slater-Determinant*, see Ref [48], a viable approximation for Ψ is obtained.

Unfortunately, the molecular orbitals φ are as obscure as the wavefunction itself [50, 47]. Their form, however, can be rationalized in the following way [51]. Considering the case of a one electron system, H_2^+, if the atom and the proton are far apart, the electron will only be in the field of one nucleus. Solving Eq. 2.4 for such a system

will yield two solutions, each with the electron located at one nucleus, where both solutions φ_1 and φ_2 take the shape of 1s orbital of the hydrogen atom. The energy E_H will be equal to that of the lone hydrogen atom. If the distance of the two nuclei is reduced slightly , the new solutions will largely retain the characteristics of φ_1 and φ_2. Therefore the new wavefunction may be constructed from the old solutions [49].

$$\Psi = \varphi_1(\mathbf{r}_1) + \varphi_2(\mathbf{r}_2) \; . \tag{2.6}$$

This is known as the Linear Combination of Atomic Orbitals (LCAO) approach [51, 52]. In mathematical phrasing, Ψ is expanded in the basis of φ_1 and φ_2. In analogy to vector algebra, functions like Ψ exist in a special vector space called *Hilbert-Space* and can, as any vector, be represented by functions that span the space [**?**] . A mathematically rigorous version of Eq. 2.6 thus reads,

$$\Psi = \sum_i c_i \varphi_i \; , \tag{2.7}$$

where the c_i account for the fact that not every φ_i may contribute equally to Ψ [**?**]. The basis set expansion of functions will be encountered again in Sec. 2.1.3.

Solving Eq. 2.4 with Eq. 2.6 gives again two solutions

$$E_{s/a} = E_H + \frac{1}{\mathbf{R}_{12}} \frac{J \pm K}{1 \pm \langle \varphi_1 | \varphi_2 \rangle} \tag{2.8}$$

where subscript s, denotes the positve ($+$-sign) combination of φ_1 and φ_2 and subscript a denotes the negative combination. E_H is the energy of the lone hydrogen atom, the fraction $1/\mathbf{R}_{12}$ denotes the Coulomb repulsion between the two nuclei at distance $\mathbf{R}_{12}$, J the Coulomb attraction between the elctron and the second nucleus, $\langle \varphi_1 | \varphi_2 \rangle$ is the overlap of the atomic orbitals, and K is the exchange term [49]. K is a nonclassical term that can be thought of as the energy gain – K is always negative when negative means more stable – from the electron having more space to move in [49]. The state corresponding to E_s is referred to as bonding state, the state corresponding to E_a as anti-bonding state. The difference between E_a and E_s is

$$E_a - E_s = \frac{K}{1 - \langle \varphi_1 | \varphi_2 \rangle} \; . \tag{2.9}$$

The functions φ are assumed to be normalized, therefore the overlap is smaller than 1. The value of 1 is only assumed if the two functions are equal. Due to the negative denominator in Eq. 2.9, the split between bonding and anti-bonding state scales with the overlap, i.e. the closer the atoms are, the stronger the bonding state is. The

observant reader will have noticed that despite claiming an intermediate distance for the nuclei, no restrictions for their relative positions has been put into the equations. Therefore, the LCAO approach, Eq. 2.6, is also applicable to the H_2^+ molecule, i.e. when both nuclei are in close range.

From this simple example the essentials of LCAO Theory can already be extracted. First, atomic orbitals combine to form molecular orbitals. The number of molecular orbitals always corresponds to the number of atomic orbitals, i.e. from two atomic orbitals two molecular orbitals are constructed [51]. Second, the energy difference between the molecular orbitals scales with the overlap of the atomic orbitals [**?**].

2.1.3 Density-Functional Theory

From its inception in the 1920s [53] Density-Functional Theory (DFT) has developed into an extremely successful and widely applied method for the quantum mechanical calculation of micro- and macroscopic properties of real systems [54]. At its core, DFT is based on the idea that a system in its ground state may be completely described by the electron density $n(\mathbf{r})$. As a consequence, a system would turn from an interacting N-particle problem into an effective one-particle problem. In 1964 Kohn and Hohenberg proved that such an immense reduction in complexity is indeed possible [53, 55]. Their findings are summed up in the following two statements:

> **1. Hohenberg-Kohn-Theorem:** The external potential v_{ext} is determined, within a trivial additive constant, by the electron density.
> **2. Hohenberg-Kohn-Theorem**: The energy of the exact electron densityis always below the energy of any approximate density.

The validity of the first Hohenberg-Kohn (HK) theorem becomes clear, when one realizes that the external potential is the only part of the Schrödinger equation (2.4) where external information is fed into it. The external potential defines the nuclear frame of the system, i.e. position, type and number of atoms, which in turn also fixes the electron number. These factors alltogether determine all electronic properties [55]. The proof of the second theorem is straightforward. Any approximate trial density $\tilde{n}$ determines its own wavefunction $\tilde{\Psi}$, approximate to the exact wavefunction Ψ of the exact density n. Since the variational priciple holds for wavefunctions, and wavefunctions and densities are bijective, it must hold for densities as well [55].

Using the HK-Theorems, the Schrödinger equation becomes, for any v - repre-

sentable (vide-infra) density n:

$$E[n(\mathbf{r})] = \int v_{\text{ext}} n(\mathbf{r})\,\mathrm{d}\mathbf{r} + F[n(\mathbf{r})]\ , \tag{2.10}$$

where $F[n(\mathbf{r})]$ is the *universal functional*, which encapsulates all electron-electron interaction. Its name is derived from the fact that it is valid for any electron density in any system with any external potential, even systems whose external potentials are not Coulomb potentials [55] . Unfortunately its explicit form is unknown. Hence, approximations to the functional have to be made, which will be discussed in section 2.1.3.

N-representability and v-representability

Strictly speaking, the above formulation of the first and second HK theorem is only valid for v representable densities. A density $\tilde{n}$ is v-representable if there exists a potential $\tilde{v}$ that determines an anti-symmetric ground-state wavefunction $\tilde{\Psi}$ that produces $\tilde{n}$. [55, 56]. However, densities, even 'reasonable' ones that are non-v-representable are easily found [56, 57]. This poses a problem, because a non-v representable trial density $\tilde{n}$, used in a minimization process for a potential v of a system, is *not* guaranteed to yield the ground state density of the system [55]. Even worse, the mathematical treatment of this quesion is cumbersome and complicated [56]. Luckily, a reformulation of the HK theorems has been achieved that instead relied only on N-representability [55]. A density is N-representable if the following conditions are satisfied [58]:

$$n(\mathbf{r}) \geq 0\,, \quad \int n(\mathbf{r})\,\mathrm{d}\mathbf{r} = N\,, \quad \int \sqrt{|\nabla n(\mathbf{r})|^2}\,\mathrm{d}\mathbf{r} < \infty\,, \tag{2.11}$$

which holds true for virtually any density one might apply in practice [55]. We can therefore safely assume the validity of the HK theorem from here on.

Making DFT Work: The Kohn-Sham equations

The universal functional can be divided into two terms. Both of them are as unknown as the entire functional:

$$F[n(\mathbf{r})] = T[n(\mathbf{r})] + U[n(\mathbf{r})] \quad \equiv \quad \min_{\Psi} \langle \Psi | \hat{U} + \hat{T} | \Psi \rangle\ , \tag{2.12}$$

where T is the exact interacting electron kinetic energy expression and U the exact correlated electron-electron exchange. All practical approaches to DFT, from the Thomas-Fermi model [56] onwards, have tried to find accurate approximations for

those functionals. The earliest approaches unfortunately were too simplistic to accurately model concepts such as molecular bonds [47]. In 1965, Kohn and Sham achieved a breakthrough, laying the foundation for modern DFT [59].

$$E\left[n\left(\mathbf{r}\right)\right] = T_{\mathrm{s}}\left[n\left(\mathbf{r}\right)\right] + U_{\mathrm{H}} + E_{\mathrm{xc}}\left[n\left(\mathbf{r}\right)\right] + \int v_{\mathrm{ext}} n\left(\mathbf{r}\right) \mathrm{d}\mathbf{r}\,, \tag{2.13}$$

where T_{s} is the non-interacting electron kinetic energy, U_{H} is the continuous form of the classical Coulomb interaction and $E_{\mathrm{xc}} = (T - T_{\mathrm{s}}) + (U - U_{\mathrm{H}})$ is the so-called exchange-correlation term, a correction for the approximation made for the other two terms [56, 60]. Kohn and Sham then showed that for slowly varying densites, the functional derivative of Eq. 2.13 is precisely equal to Kohn's original many-body formulation [53] when applied to a system of non-interacting electrons moving in an external potential $v_{\mathrm{ext}} + v_{\mathrm{xc}}$ [60]. It is therefore possible to use the one-particle Schrödinger equation

$$\varepsilon_i \phi_i = (v_{\mathrm{ext}} + v_H - \frac{1}{2}\nabla^2 + v_{\mathrm{xc}})\phi_i\,, \tag{2.14}$$

to determine the density. Here, $v_H = \delta U_{\mathrm{H}}/\delta n$, $v_{\mathrm{xc}} = \delta E_{\mathrm{xc}}/\delta n$. In combination with v_{ext} they form the effective single-particle Kohn-Sham potential v_{s} [56]. The orbitals ϕ_i and the orbital energy ε_i have no physical meaning beyond reproducing the density via $n\left(\mathbf{r}\right) = \sum_i \phi_i^\star \phi_i$ [47]. The complete set of orbitals forms a Kohn-Sham (slater) determinant that is unknown. Hence, Eq. 2.14 has to be solved iteratively for all i [60]. Once such as self-consistent calculation is converged, the energy functional can be written as

$$E\left[n\left(\mathbf{r}\right)\right] = \underbrace{\sum_i \langle\phi_i| - \frac{1}{2}\nabla^2 |\phi_i\rangle}_{T_{\mathrm{s}}} + \underbrace{\int \frac{n\left(\mathbf{r}\right) n\left(\mathbf{r}'\right)}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'}_{U_{\mathrm{H}}} + \underbrace{\int n\left(\mathbf{r}\right) v_{\mathrm{xc}}\, \mathrm{d}\mathbf{r}}_{E_{\mathrm{xc}}}\,. \tag{2.15}$$

Thy entire complexity of the system is now hidden in v_{xc}. However, the contribution of E_{xc} to the total energy are comparatively small. Thus, errors made by approximating E_{xc} are not as grave as errors made by approximating the much larger kinetic energy T for instance [61].

The meaning of the exchange-correlation functional

The exchange correlation functional can be thought of to encompass two quantum effects, namely exchange E_{x} and correlation E_{c}, hence its name. Exchange, or Fermi-correlation hereby arises from the Pauli principle. Even near the location of an electron, the probability to find another electron of the same spin is reduced. Coulomb-correlation on the other hand, arises from the fact that electrons have the

same charge and therefore tend to avoid the actual current position of the respective other electrons. This as well leads to a reduced probability of finding one electron close to another. However, Coulomb correlation does not prohibit the electrons to be at the exact same position [48, 61, 55]. The reduced probability to find an electron in the presence of another can be described with the exchange-correlation hole

$$n_{\mathrm{xc}}(\mathbf{r},\mathbf{r}') = \left(g[n](\mathbf{r},\mathbf{r}') - 1\right) n(\mathbf{r}) \ . \tag{2.16}$$

The term $g(\mathbf{r},\mathbf{r}')$ is the electron-electron pair correlation function [61]. By introducing a fictitious hamiltonian H_λ this result can be generalized to obtain an average density $\bar{n}_{\mathrm{xc}}$. Here, $0 \leq \lambda \leq 1$ marks the fractional presence of an electron. In the fictitious system H_λ the potential is modified as such that the fictitious system's density $n_\lambda = n_{\lambda=1} = n(\mathbf{r})$ for all values of λ [47]. Its significance lies in the fact that $\bar{n}_{\mathrm{xc}}$ can describe E_{xc} exactly. [47]. The exact form of v_{xc} is:

$$E_{\mathrm{xc}} = \frac{1}{2}\int v_{\mathrm{xc}}\, n(\mathbf{r})\, \mathrm{d}\mathbf{r} = \frac{1}{2}\int\int \frac{\bar{n}_{\mathrm{xc}}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, \mathrm{d}\mathbf{r}'\, \mathrm{d}\mathbf{r} \ . \tag{2.17}$$

The objects $\bar{n}_{\mathrm{xc}}, E_{\mathrm{x}}, E_{\mathrm{c}}$ have several known properties that can be used to benchmark actual approximate functionals [61]. For instance, $\int n_{\mathrm{xc}}\, \mathrm{d}\mathbf{r} = -1$ and $E_{\mathrm{xc}} = -U_{\mathrm{H}}$ for a one-electron system [61]. The last property might be surprising, but it is a direct consequence of the self-interaction inherently present in U_{H}.

Approximations to the exchange-correlation functional

Since $v_{\mathrm{xc}}[n(\mathbf{r})]$ is a functional of $n(\mathbf{r})$ that is expected to be predominantly short-sighted, one can expand $n(\mathbf{r})$ into a power series.

$$n(\mathbf{r}) = n + n_i\mathbf{r} + \frac{1}{2}\sum n_{ij}\mathbf{r}_i\mathbf{r}_j \ , \tag{2.18}$$

where $n = n(0)$ and $n_i = \nabla_i n(\mathbf{r})$. This series can be resummed to obtain

$$E^0_{\mathrm{xc}} = \int \tilde{v}_{\mathrm{xc}}(n(\mathbf{r}))n(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{2.19}$$

$$E^1_{\mathrm{xc}} = \int \tilde{v}_{\mathrm{xc}}(n(\mathbf{r}), |\nabla n(\mathbf{r})|)n(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{2.20}$$

$$E^2_{\mathrm{xc}} = \int \tilde{v}_{\mathrm{xc}}(n(\mathbf{r}), |\nabla n(\mathbf{r})|)\nabla^2 n(\mathbf{r})\, \mathrm{d}\mathbf{r} \ , \tag{2.21}$$

where $\tilde{v}_{\mathrm{xc}}$ is a nondescript function that mimics the exact potential. Eq. 2.19 describes the simplest of all functionals, the local-density approximation (LDA). It assumes that the density is (virtually) constant on microscopic distances such as the Fermi-wavelength $\lambda_{\mathrm{F}} = [3\pi^2 n(\mathbf{r})]^{-1/3}$ [47]. Such condition is often violated in molec-

ular setups [47]. Consequently a wide variety of functionals in the style of Eq. 2.20 have been developed. These so called generalized-gradient-approximation (GGA) functionals make intricate use of properties of the exact functional [47]. Among the most successful functionals are the works of Perdew, Burke and Enzerhof [62] and Becke, Lee Young and Parr [63, 64]. Beyond the GGA approach people have made use of the kinetic energy density (meta-GGA) [61], or used the exact exchange in a hybrid approach

$$E_{\mathrm{xc}}^{hyb} = \alpha E_{\mathrm{x}}^{K} S + (1-\alpha) E_{\mathrm{xc}}^{GGA} \,, \tag{2.22}$$

where E_{xc}^{GGA} is an appropriate GGA functional and α is a fitting parameter [47].

Implementation of Density-Functional Theory

Molecular orbitals ϕ_i are complicated and unknown [50, 47]. An attempt to sample it numerically leads to the same problem as the attempt to represent the entire wavefunction numerically. As has been discussed in Sec. 2.1.2, a possible way to handle unknown functions, is to expand them in a basis of known functions χ.

$$\phi_i = \sum_{\mu} c_{\mu i} \chi_{\mu} \,. \tag{2.23}$$

This reduces the problem of finding an unknown function to the problem of finding a set of numbers, namely $c_{\mu i}$ [48]. Two representations for the description of molecular orbitals are popular:

Gaussian basis-functions Predominantly popular in computational chemistry, basis set made of Gaussians are of the form

$$\chi_{\mathbf{m}} = N_{\mathbf{m}} r_x^{m_x} r_y^{m_y} r_z^{m_z} \exp\left[-\xi \mathbf{r}^2\right] \,, \tag{2.24}$$

where where $N_{\mathbf{m}}$, m_x, m_y, m_z and ξ represent parameters and m represents an angular momentum quantum number [65]. These functions are modeled after the quantum states of atomic wavefunctions [61] and intend to mimic the atomic orbitals of chemical elements [65].

Plane-wave basis In solid state physics, plane wave basis sets have been popular from early on, because of their inherent periodicity and their delocalization [65]. The properties are ideal to describe electrons in a periodic crystal.

$$\phi_i = \sum_{\mathbf{G}} n_i(\mathbf{G}) \mathrm{e}^{i\mathbf{G}\mathbf{r}} \,, \tag{2.25}$$

where $\mathbf{G}$ is a wave vector that fits the periodicity of v_{ext} and $n_i(\mathbf{G})$ is the expansion coefficient.

In principle, any function can serve as a basis, however, the Hilbert space (see Sec. 2.1.2) is infinite-dimensional, thus it requires infinite basis functions [48]. The presented methods are popular, because they allow to truncate the infinite basis without sacrificing accuracy too much [**?**]. In mathematical terms, with the right basis, ϕ_i in Eq. 2.23 has only a finite subset of χ_μ whose $c_{\mu i}$ are nonzero. This rationale becomes particularly clear in case of Gaussian basis functions. As the discussion of LCAO, Sec. 2.1.2 revealed, molecular orbitals retain a lot of characteristics for the atomic orbitals of the atoms constituting the molecule, hence this finite set is a good descriptor of a molecular orbital.

With the basis set expansion the electronic structure problem Eq. 2.14 can be remodeled into a matrix equation [56], which is given here for Gaussian orbitals

$$\sum_{\mu} \underbrace{\langle \nu | \, v_{\mathrm{ext}} + v_H - \frac{1}{2}\nabla^2 + v_{xc} \, | \mu \rangle}_{\mathbf{K}_{\mu\nu}[n(\mathbf{r})]} c_{\mu i} = \sum_{\mu} \langle \nu \, | \mu \rangle \; c_{\mu i} \varepsilon_{ii} \,, \tag{2.26}$$

which is equivalent to

$$\mathbf{K}\left[n\left(\mathbf{r}\right)\right]\mathbf{C} = \mathbf{S}\mathbf{C}\varepsilon \,. \tag{2.27}$$

Here, $\mathbf{K}$ represents the Kohn-Sham Matrix, $\mathbf{S}$ the overlap matrix and $\mathbf{C}$ is the matrix of coefficients which represents the wavefunction. This equation could be solved using standard matrix diagonalization techniques, if $\mathbf{K}$ wasn't dependent on the electron density (via v_H and v_{xc}). Instead, solving Eq. 2.26 has to be done with the Self-Consistent Field (SCF) cycle as depicted in Fig. 2.1 Professional implementations today offer a wide variety of algorithms to circumvent the SCF problem [66] or reduce the number of steps required for convergence.

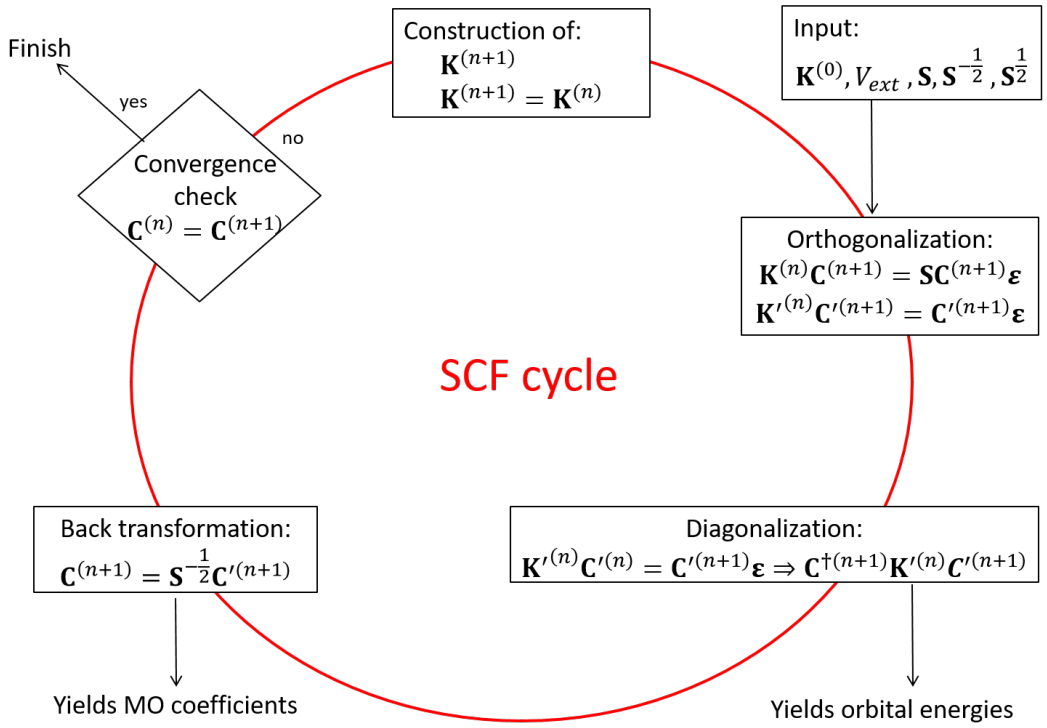


Figure 2.1: Schematic depiction of a self-consistent field cycle.

Limitations of DFT

From the existence of the universal functional F follows that DFT is mathematically exact [67]. However, actual DFT calculations suffer from a variety of well-known inaccuracies. Among those is the underestimation of the band gap, a well-known failure of current functionals, as well as the overbinding of charge-transfer complexes, the instability of vdW structures and the overestimation of reaction energies [67, 68]. These errors can be traced to three sources: The locality of functionals, erroneous delocalization, and static correlation [67, 68].

Local functionals Current exchange-correlation functional only use density values at a single point $\mathbf{r}$ to calculate E_{xc}. This approach reduces the cost of calculation but it entails the restriction that only local properties can be caputured by the functional [68]. Since popular functionals, for instance all GGA functionals are local, they cannot capture the non-local long-range electron correlation that is the basis for the van-der-Waals forces [68]. The neglect of non local correlation results, among other errors, in the instability of van-der-Waals complexes, such as benzene [68], and a gross underestimation of the density of water [69].

Electron delocalization An important property of the exact functional F is linearity, i.e.

$$F[an_a + bn_b] = aF[n_a] + bF[n_b] \,, \tag{2.28}$$

where n_a, n_b are electronic densities. It follows from this property that the energy for fractional occupations must be a straigt line (See Fig. 2.2). However, existing LDA and GGA functionals underestimate the energy of fractional orbital occupation and consequently favor those [70]. The error becomes apparent in molecular reactions, because LDA and GGA functionals are far too willing to distribute an electron fractionally in a broken bond, leading to an underestimated reaction barrier [67]. In condensed phases this behavior explains the erroneous electron absorption and ionization energies and hence the famous underestimation of the band gap. Since, in an infinite system the addition, or removal of an electron does only marginally affect the total number, the energy change can be written as

$$E(mN \pm 1) = mE(N) \pm \delta \frac{\partial E}{\partial N}|_{N+\pm\delta,\delta=1} \,, \tag{2.29}$$

where m is a large integer and N is the electron number. As is illustrated in Fig. 2.2, LDA and GGA functionals predict completely wrong energies at the next integer points, thus underestimating the band gap [71].

Static Correlation In wavefunction approaches, static correlation arises when more than one slater-determinant becomes relevant. This is often the case, when two or more states are near-degenerate. The simplest example of such a system is elongated H_2, but near-degeneracy is present in most metallo-organic complexes [67]. In the case of H_2 a realistic representation would be two determinants with all electrons at one hydrogen, H^1 and all electrons at the other hydrogen, H^2. This is equivalent to having half a spin-up and half a spin-down electron at each hydrogen [72]. The exact functional F is required to yield the same energy for all spin values for one electron between 1 and -1. [67]. After all, the Hamilton operator Eq. 2.2 does not depend on spin. It has been shown though that existing functionals violate this condition. In fact, the error calculated from placing half a spin-up and half a spin-down electron in one orbital precisely matches the dissociation error of H_2 [72, 67].

2.1.4 Molecular Mechanics

Instead of attempting to accurately describe the electronic structure of a system in order to learn about its properties, molecular mechanics (MM) aims at an efficient description of the force field in which the atoms move. The energy expression is

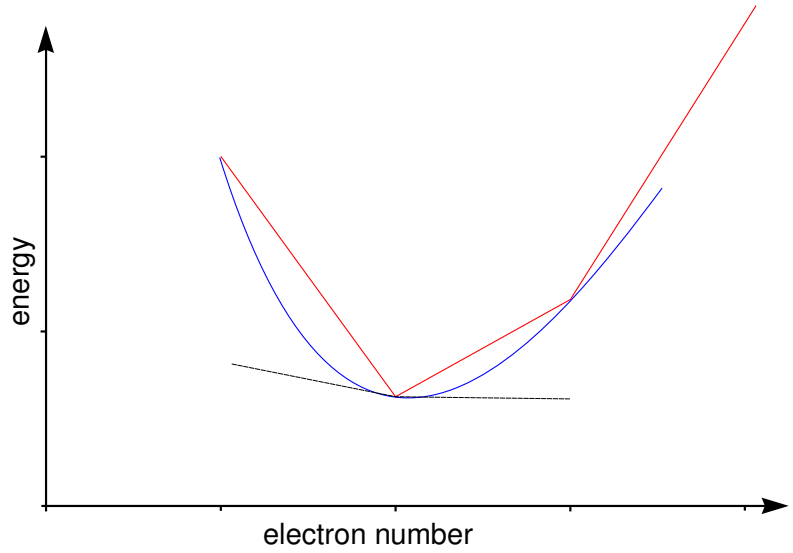


Figure 2.2: Schematic depiction of the relation between energy and number of electrons. The red line shows the exact functional, whereas the blue line shows the devolution of semilocal functionals such as LDA and GGA. The dashed black line shows the devolution for large N. Picture based on [67, 73]

comparatively simple, which makes the method suitable for the calculation of expansive systems over long timescales [74]. In contrast to other parametric methods, notably machine learning algorithms, MM force fields draw on understandable physical concepts such as Coulomb interaction and the existence of molecules to build the potential energy landscape [75]. A general form of the MM energy expression is [76, 77, 74]

$$\begin{aligned} E_{MM} = & \sum_{bonds} V(\mathbf{R}_I, \mathbf{R}_J) + \sum_{angles} \Theta(\mathbf{R}_I, \mathbf{R}_J, \mathbf{R}_K) \\ & + \sum_{dihedrals} \Phi(\mathbf{R}_I, \mathbf{R}_J, \mathbf{R}_K, \mathbf{R}_L) \\ & + \sum_{nonbonded} V_{IJ}^{\mathrm{LJ}}(\mathbf{R}_I, \mathbf{R}_J) + \frac{q_I\, q_J}{\varepsilon |\mathbf{R}_I - \mathbf{R}_J|} \,. \end{aligned} \tag{2.30}$$

Here, I, J, K, L are atomic indexes. The potentials V, Θ, Φ are constrained to atoms belonging to the same molecule. The van der Waals interaction and electrostatic interaction between molecules is represented by a Lennard-Jones and simple coulomb potential, respectively. q_I, q_J are partial charges on the respective atoms ε is the effective dielectric constant [74]. Each potential depends on parameters specific to an atomic kind, where *atomic kind* refers to an element in a specific environment. For instance, the sp^2-carbon atom in a ketone would constitute an individual

atomic kind.

The strategy on how the atomic kind parameters are obtained varies for individual force fields and is sketched out here for the CHARMM[1] method [74]. First, the intramolecular parameters are optimized such that a calculated structure matches experimentally obtained gas-phase or crystal structures. Once this process converges, additional data, notably vibrational frequencies, are taken into account. The intermolecular parameters are fixed at this stage. Second, intermolecular parameters are determined by fitting to interaction energies and structures of model dimer systems, as well as to macroscopic data of liquids using the previously optimized intramolecular parameters. This new parameter set is used to reevaluate the intramolecular ones and this process is reiterated until convergence.

2.1.5 QM/MM

Quantum mechanical methods of evaluating the force suffer from an unfavorable scaling behavior [50] with the system size which limits their applicability to comparitvely small systems. Molecular mechanics forcefields, however, are usually incapable of capturing chemical reactions [78]. Especially in biochemistry the systems of interest are large. For instance, the steric changes of ligand binding in hemoglobin have been studied extensively [79, 80, 81] The immense size of hemoglobin ($\approx 65\ \mathrm{kDa}$) [81] clearly prohibits the use of QM methods, while the need to capture a chemical reaction prohibits the use of MM methods alone.

Although linearly scaling QM methods are in development [42], which might one day be able to handle expansive systems, the best current method is to model a system as large as a protein by a QM system surrounded by a large MM system [82]. This approach, called QM/MM has been introduced by Field, Bash und Karplus in 1990 [76, 83]. Since then, the method has grown so successful and significant that the Nobel Committee chose to award Karplus in 2013 [84], along with Levitt and Warshel, who developed many of the concepts used for QM/MM [76].

One of the great challenges when setting up the QM/MM system is the question of how to model the interaction, especially the electrostatic interaction of MM and QM molecules[82]. For systems where the division between the QM subsystem and the MM subsystem is fixed, two general approaches have so far been established

[1] Chemistry at Harvard Molecular Mechanics

[76, 85]

$$\text{Additive QM/MM}: \quad H = H_{\mathrm{QM}}(QM) + H_{\mathrm{MM}}(MM) + H_{\mathrm{QM-MM}} \tag{2.31}$$

$$\text{Substractive QM/MM}: \quad H = H_{\mathrm{MM}}(all) + H_{\mathrm{QM}}(QM) - H_{\mathrm{MM}}(QM)\,. \tag{2.32}$$

Here, H is the Hamiltonian of the entire system, $H_{\mathrm{MM}}(all)$ is the MM Hamiltonian evaluated on the entire system and $H_{\mathrm{QM}}(QM)$ is the QM Hamiltonian evaluated on the QM region of the system. In the substractive scheme Eq. 2.32, the entire system is first calculated on the MM level and then a part of the system is elevated to QM level. An example of a substractive scheme is the IMMOM /ONIOM method [85]. In the additive scheme the QM and MM region are calculated separately and a special interaction term $H_{\mathrm{QM-MM}}$ added. Since the majority of implementation use an additive scheme [76], this text will focus on presenting additive QM/MM methods. There, the $H_{\mathrm{QM-MM}}$ hamiltonian has to encompass methods covering bonds crossing the QM/MM boundary, electrostatic and van der Waals coupling. This section will focus on methods that have been implemented in CP2K [86, 87], because this software package has been used predominantly in this work.

CP2K Multigrid approach

Consider the integral U_H from Eq. 2.15

$$U_H = \int \sum_i \sum_{\mu\nu} c_{\mu i} c_{\nu i}\, \chi_\nu(\mathbf{r}) \chi_\mu(\mathbf{r}) v_H(\mathbf{r})\, \mathrm{d}\mathbf{r}\,, \tag{2.33}$$

where n has been expanded into basis functions χ_μ. To evaluate U_H numerically, the χ_μ have to be expanded on a grid and evaluated at each grid point. The fineness of this grid is determined by the most fluctuating basis function. This results in slowly varying functions being evaluated unnecessarily often, leading to high computational cost. The solution to this problem is a multigrid approach [82]. Slowly varying functions are evaluated on coarse grids, quickly varying ones one fine grids. The function values on the coarse grid are then interpolated to the fine grid, where the integration is performed [82].

Electrostatic coupling

Analogous to U_H (Eq. 2.15) the electrostatic QM-MM interaction can be written as

$$E_{\mathrm{QM-MM}} = \sum_{I \in \mathrm{MM}} \int \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} v_I(\mathbf{r}) \mathrm{d}\mathbf{r} \tag{2.34}$$

Here, v_I is the potential generated by an MM-Atom's charge, $n(\mathbf{r})$ is the QM charge density. A naive evaluation of all pairs $v_I n_i$, where n_i is the density at a grid point, would make $E_{\mathrm{QM-MM}}$ the costliest term in a QM/MM calculation [82]. While a coarse graining scheme using a hirarchical set of cutoffs has been used elsewhere [88], CP2K uses its multigrid framework to reduce the computational burden. To this end the potentials of every MM atom is expanded into a set of Gaussians [82]

$$v_I(\mathbf{r}) = \frac{q_I \mathrm{erf}\left[\frac{\mathbf{r}-\mathbf{R}_I}{r_{cI}}\right]}{\mathbf{r}-\mathbf{R}_I} = \sum_{k\in N_I} A_k \exp\left[-\frac{(\mathbf{r}-\mathbf{R}_I)^2}{\sigma_k^2}\right] + R_{low}(\mathbf{r}-\mathbf{R}_I)\,. \tag{2.35}$$

Here, r_{cI} is the width of the error function, q_I the charge of atom I, A_k and σ_k^2 are the height and width of the gaussian functions, N_I is the set of gaussians for atom I and R_{low} a function that retains the long range $1/\mathbf{r}$ behavior of v_I. These Gaussians can now be collocated to the multigrid the same way the functions χ_μ are. However, the greater the distance of the MM atoms to the QM zone, the more functions will have zero value, starting from the fastest varying to the slowliest varying functions. Thus, for the majority of MM atoms only R_{low} has to be evaluated, which can be done on the coarsest grid [82].

The effects of periodic boundary conditions (PBC) are easily incorporated into this scheme [89]. Special care has to be taken for the periodic replicas of the QM subsystem, because the assumed periodicity of the QM subsystem is the length of the QM cell [89]. Hence, the QM system has to be decoupled from its periodic images and recoupled at the appropriate MM periodicity. This is achieved by using the Blöchl scheme [90]. There the QM density is represented by a set of Gaussians

$$n(\mathbf{r}) \Rightarrow \sum_\alpha q_\alpha A_\alpha \exp\left[\frac{(\mathbf{r}-\mathbf{r}_\alpha)^2}{\sigma_\alpha^2}\right]\,, \tag{2.36}$$

which are placed such that their effect on an external test charge equals that of the real density [90]. Here, ths subscript α is an arbitrary iterator. The MM charges are dealt with adaption of the Ewald-method (see 2.2). The MM-Potential v_I is split into a long range and a short range part $v_I = v_I^l + v_I^s$, where v_I^l is evaluated in reciprocal space, and v_I^s is evaluated in real space. It is easily understandable that

$$v_I^s = \sum_{k\in N_I}\sum_{\mathbf{T}_R} A_k \exp\left[\frac{\mathbf{r}-\mathbf{R}_I+\mathbf{T}_R}{\sigma_k^2}\right] \tag{2.37}$$

$$v_I^l = \sum_{\mathbf{T}_R} R_{low}(\mathbf{r}-\mathbf{R}_I+\mathbf{T}_R)\,, \tag{2.38}$$

because the quick decay of the Gaussian functions was the motivation for using them as an expansion of the MM charge [89]. The operator T_R represents a translation of an object to its periodic image.

Bonds crossing the QM/MM boundary

In many QM/MM applications, especially when studying biomolecules, chemical bonds will be cut, creating two unsaturated bridgehead atoms $\mathbf{R}_{QM}$ and $\mathbf{R}_{MM}$ at the QM/MM boundary. Senn et al. identify three main issues in that situation. First, the broken bond has to be saturated. Second, overpolarization due to a point charge close to the QM zone has to be countered. Third, double counting of forces has to be elminated [76]. They mention three types of methods to address these: Link Atoms, Frontier Atoms and Frontier Orbitals

Link Atoms are implemented into CP2K [86] as an IMOMM (integrated molecular orbital + molecular mechanics) scheme [91]. There, an H atom is placed between the $\mathbf{R}_{QM}$ and $\mathbf{R}_{MM}$ and its position $\mathbf{R}_{link} = (1+\alpha)\mathbf{R}_{QM} - \alpha\mathbf{R}_{MM}$, where α is a fixed scaling factor. This circumvents the common link atom problem of introducing additional degrees of freedom [76]. IMOMM assumes that the link atom describes the QM/MM bond in a satisfactory way. Thus the motion of $\mathbf{R}_{MM}$ with respect to $\mathbf{R}_{QM}$ is guided by the scaling relation that defines the $\mathbf{R}_{link}$ without computing any force [92, 91]. Overpolarization is avoided due to the gaussian expansion of the MM potential [82, 76]. There also exists a strategy that uses a link atom in conjunction with an optimized effective core potential (OECP). The OECP is determined by a pseudopotential generation routine, such that it reproduces the molecular property of the group replaced by the link atom [93].

Frontier Atoms which are implemented in CP2K as a generalized hybrid orbital method [94]. There, assuming the broken bond is a C-C σ bond, a set of 4 hybrid orbitals is placed at $\mathbf{R}_{MM}$. One of those, the *hybrid orbital* contains one electron and acts as a normal atomic orbital in the QM region, while the other three, the *auxilary orbitals* are strictly orthogonal to the QM zone's molecular orbitals. Instead, the latter are used to smear out the MM atom's partial charge [94, 76].

Frontier Orbitals are generated in a QM model system that incorporates the broken bond. This orbital is then transfered to the QM/MM system and reconstructed from the QM zones atomic orbitals in such a way that the frontier orbitals are orthogonal to each other and to all other MOs and AOs of the system. The frontier orbials are kept frozen during the SCF procedure and contains two electrons [95, 76].

2.2 Molecular Dynamics

Computational chemistry aims at the calculation of micro- and macroscopic properties, for instance, vibrational frequencies and dielectric constants. Their calculation for strongly interacting systems, such as crystals, glass, or single molecules, requires only their equilibrium configuration – i.e. the position of all atoms. Due to the strong interatomic interaction, thermal activation will result only in a vibration of the atoms around the equilibrium position [96] which therefore retains a statistical weight of virtually 1. As a result most properties can be be determined with sufficient accuracy from that configuration. However, a wide array of chemistry deals with weakly interacting systems, for instance solutions with water. In these cases, the depth of the potential wells of atoms can be overcome by the kinetic energy of the atoms. Thus, configurations other than the 0 K equilibrium become relevant and have to be taken into account with appropriate statistical weight [97]. The two most popular methods that accomplish this are Monte-Carlo [96, 98] and Molecular Dynamics (MD). The latter will be presented here, based on the exensive literature available on this topic [98, 99, 65, 97].

The nuclei of molecular systems are in general assumed to behave classically. This assumption allows the use of simple Newtonian dynamics for calculation. Within this assumption a system is completely defined by the 6N - dimensional set of positions $\mathbf{R}$ and momenta $\mathbf{P}$ of the N particles. The aforementioned set further constitutes a point in the so-called phase space. The movement of the system through the phase space in time is called a trajectory. Any point in phase space completely determines the trajectory, because the equations of motion

$$\frac{\partial \mathscr{H}}{\partial \mathbf{R}} = \dot{\mathbf{P}} \tag{2.39}$$

$$\frac{\partial \mathscr{H}}{\partial \mathbf{P}} = \dot{\mathbf{R}} \tag{2.40}$$

define how to move from one point to another. Here, $\mathscr{H} = T + V$ is the Hamiltonian, or the total energy of the system, and $\dot{\mathbf{P}}$ and $\dot{\mathbf{R}}$ are the time derivatives of the momenta and positions, respectively. In reality, the analytic solution to a trajectory is inaccessible. The two main reasons behind this are: One, the quantum mechanical nature of particles. The uncertainty principle states that the precise values of $\mathbf{R}$ and $\mathbf{P}$ cannot be known simultaneously and thus a system cannot be fixed on a point in phase space. Two, the simple fact that there mathematics cannot provide an analytical solution for anything involving more than two bodies. [100]. Instead, an approximate dynamics can be derived through the expansion of $\mathbf{R}(t)$ in a Taylor

series

$$\mathbf{R}(\Delta t) = \mathbf{R}(0) + \dot{\mathbf{R}}(0)\Delta t + \frac{\mathbf{F}(0)}{2\mathbf{M}}\Delta t^2 + \frac{1}{3!}\ddot{\mathbf{R}}(0)\Delta t^3 + \cdots \tag{2.41}$$

$$\mathbf{R}(0) = \mathbf{R}(\Delta t) - \dot{\mathbf{R}}(\Delta t)\Delta t + \frac{\mathbf{F}(\Delta t)}{2\mathbf{M}}\Delta t^2 - \frac{1}{3!}\ddot{\mathbf{R}}(\Delta t)\Delta t^3 + \cdots \tag{2.42}$$

These expansions can be terminated at an arbitrary order. A very common approximation for molecular dynamics is to expand them only up to Δt^2. Through insertion of Eq. 2.42 into Eq. 2.41, a propagation scheme that calculates the velocities and the positions simultaneously, is obtained, the so-called The Velocity-Verlet algorithm [98]:

$$\begin{aligned} \mathbf{R}(\Delta t) &= \mathbf{R}(0) + \dot{\mathbf{R}}(0)\Delta t + \frac{\mathbf{F}(0)}{2\mathbf{M}}\Delta t^2 \\ \dot{\mathbf{R}}(\Delta t) &= \dot{\mathbf{R}}(0) + \frac{\mathbf{F}(0) + \mathbf{F}(\Delta t)}{2\mathbf{M}}\Delta t \ . \end{aligned} \tag{2.43}$$

Incidentally the Velocity - Verlet integrator is the simplest outcome of a scheme that systematically produces “good” integrators, which will be explained in the next section.

Systematic approach to good integrators

From considerations of symmetry and conservation laws it is known that integration algorithms for molecular dynamics should have the following features: Besides using computational resources efficiently they should be time-reversible, symplectic, as well as sufficiently accurate regarding the nuclear forces and the energy [101, 98]. An algorithm is time-reversible if moving forward and then backwards in time by Δt will result in ending at the starting point and it is symplectic if the phase space volume is preserved. A time-reversible and symplectic algorithm will also be energy conserving. Since the evaluation of the forces is the most expensive part in an MD simulation, an efficient algorithm should only require one force evaluation per time step. The accuracy of forces, as well the computational efficiency are primarily a concern of the electronic structure method, which were covered in Sec. 2.1. Symplectic and time-reversible algorithms can be obtained from a method devised by Tuckerman et al. [98, 102]. Let the system state be given by an arbitrary function f, of the positions $\mathbf{R}$ and the momenta $\mathbf{P}$ of its particles. The time derivative of f is

$$\dot{f} = \left(\dot{\mathbf{R}}\,\frac{\partial}{\partial \mathbf{R}} + \dot{\mathbf{P}}\,\frac{\partial}{\partial \mathbf{P}}\right) f \equiv \mathrm{i}\mathrm{L} f \ , \tag{2.44}$$

where the center defines the Liouville operator iL. The Liouville operator can be divided into a part that acts on the positions iL_r and one that acts on the momenta iL_p. The state of the system can be obtained by integration of Eq. 2.44.

$$f(\mathbf{R}(t),\mathbf{P}(t)) = \exp[(\mathrm{iL}_P + \mathrm{iL}_R)t] f(\mathbf{R}(t=0),\mathbf{P}(t=0)) \,. \tag{2.45}$$

The propagator $\mathrm{e}^{\mathrm{iL}t} = U(t)$ is unitary, i.e. $U^{-1}(t) = U(-t)$. The inverse propagator propagates the system back in time and it follows, that the propagator is time-reversible. It can be shown, that this remains true for all discrete approximations [102]. The Trotter identity, which is exact only in the case of $n \to \infty$, can be used to separate the exponential operator in Eq. 2.45

$$\exp[(\mathrm{iL}_R + \mathrm{iL}_P)t] \approx \left(\mathrm{e}^{\mathrm{iL}_P t/(2n)} \mathrm{e}^{\mathrm{iL}_R t/n} \mathrm{e}^{\mathrm{iL}_P t/(2n)}\right)^n = \tilde{U}(t/n) \,. \tag{2.46}$$

The formal solution, Eq. 2.45, can now be replaced with a discretized version, in which the time step Δt is $\Delta t = t/n$. In the simplest case of $n = 1$ integration of f with the discrete operator will give:

$$f(\mathbf{R}(\Delta t),\mathbf{P}(\Delta t)) = f\left(\left[\mathbf{P}(0) + \frac{\Delta t}{2}\dot{\mathbf{P}}(0) + \frac{\Delta t}{2}\dot{\mathbf{P}}(\Delta t)\right], \left[\mathbf{R}(0) + \Delta t \dot{\mathbf{R}}(\Delta t/2)\right]\right) \,, \tag{2.47}$$

where $\dot{\mathbf{P}}$ is the force acting on a particle and the velocity $\dot{\mathbf{R}}(t) = \dot{\mathbf{R}}(0) + \mathbf{F}(0)/(2\mathbf{M})$. If one uses a higher number of propagators n, more accurate integrators and hence more accurate trajectories can be obtained. This is however not required in molecular dynamics because of the so called Lyapunov instability, which will be discussed later. With only one propagator, a single application of the Liouville operator to the positions and momenta of the particles gives

$$\mathbf{P}(\Delta t) = \mathbf{P}(0) + \frac{\mathbf{F}(0) + \mathbf{F}(\Delta t)}{2}\Delta t \tag{2.48}$$

$$\mathbf{R}(\Delta t) = \mathbf{R}(0) + \dot{\mathbf{R}}(0)\Delta t + \frac{\mathbf{F}(0)}{2\mathbf{M}}\Delta t^2 \,, \tag{2.49}$$

which is precisely the Velocity-Verlet algorithm. Since $\dot{\mathbf{R}}$ is a function of $\mathbf{P}$ only and $\dot{\mathbf{P}} = \mathbf{F}(\mathbf{R})$ is a function of the positions only, the Jacobian of the transformation $\tilde{U}(\Delta t)$ is unity and the phase space volume therefore is preserved [98]. This derivation thus proves that the Velocity-Verlet algorithm is a time-reversible, symplectic algorithm that is accurate up to $\mathscr{O}(\Delta t^2)$.

Although time-reversible and symplectic, the algorithm does not strictly conserve the energy, because the finite time step of the integrator $\tilde{U}(\Delta t)$, produces erroneous forces, that result in an always positive energy drift [103]. In quantum mechanical

methods errors in the forces also arise from the fact that the wavefunction is never truly converged, which gives rise to an extra term in the derivative that is not calculated [101].

These erroneous dynamics generated by $\tilde{U}(\Delta t)$ can be expressed with a pseudo Liouville operator $\mathrm{iL}_{\mathrm{pseudo}}$ that is related to the iL by an error operator ε

$$\mathrm{iL}_{\mathrm{pseudo}} = \mathrm{iL} + \varepsilon \Delta t \ . \tag{2.50}$$

The pseudo Liouville operator corresponds to a pseudo Hamiltonian that differs from the true Hamiltonian, i.e. the energy conserving, correspondent of iL, in the order of Δt^2 [98]. However, the pseudo Hamiltonian itself is rigorously conserved by Verlet style algorithms [98, 104]. If Δt is small, the difference to the dynamics of the true Hamiltonian will be small too, and the drift smaller than inherent fluctuations of the energy.

To conclude this section, the Lyapunov instability is discussed, which will offer an explanation why only the lowest order algorithms are used. To this end, two systems are considered [98], where the initial velocity of a single particle I is slightly perturbed by $\varepsilon_{\mathbf{P}}$

$$\mathbf{R}_I(t) = f(\mathbf{R}(0), \mathbf{P}(0), t) \ , \quad \mathbf{R}'_I(t) = f(\mathbf{R}(0), \mathbf{P}(0) + \varepsilon_{\mathbf{P}}, t) \ . \tag{2.51}$$

The difference between their positions at time t, $\mathbf{R}_I$ and $\mathbf{R}'_I$ is denoted by $\Delta\mathbf{R}_I(t)$ It can be shown that their motion diverges exponentially

$$|\Delta\mathbf{R}_I| \propto \ \varepsilon_{\mathbf{P}} \mathrm{e}^{\lambda t} \ , \tag{2.52}$$

Here, λ is the largest of the 6N Lyapunov exponents; due to the exponential relation, it dominates the divergence. Frenkel and Smit illustrated the effect for two 1000 particle systems in which they perturbed the velocity of a single particle by $10^{-10}\mathbf{P}_I(0)$. To evaluate the two systems divergence, the squares of the differences of positions were summed up and monitored

$$\sum_{I=1}^{1000} |\mathbf{R}'_I(t) - \mathbf{R}_I(t)|^2 \ . \tag{2.53}$$

After only 1000 time steps this sum had increased by 14 orders of magnitude, indicating that the two systems were in no way correlated to each other any more [98]. Since typical molecular dynamics simulations run for hundreds of thousands of time

steps, it becomes obvious why there is no need for high precision integrators: The exact positions and the exact momenta cannot be known, because of the uncertainty principle. Even if they could be known, with the finite precision arithmetic used upon integration, simulated trajectories would bear no resemblance to actual trajectories long before any observables could be calculated.

The effect of the Lyapurnov instability might appear disastrous to the usefulness of molecular dynamics. However, for atoms a ”real” trajectory is not important. Instead, only the correct time averages are relevant as they link the microscopic system to macroscopic, observable properties [98]. It should be noted here that for this to be valid, the system under study has to have the ability to reach every point in phase space, i.e. be ergodic [98]. The Lyapurnov instability also easily explains the popularity of the Velocity-Verlet algorithm: It is simple and sufficient for the calculation of atomic systems.

Ab-inito MD: Propagating the wave function through time

In ab-inito molecular dynamics (AIMD) the force on the particles is evaluated quantum mechanically. In this case the question of how to recalculate the wavefunction after one MD step arises immediately. In one approach, called Born-Oppenheimer MD (BOMD), the particles traverse on the Born-Oppenheimer surface, i. e. the potential energy surface of the electronic ground state. Naturally, this requires the iterative convergence of the electronic structure at each time step, a costly procedure. Thus AIMD had only limited usefulness due to the high computational cost especially in the absence of methods to transfer electronic information between timesteps [65, 105].

In 1985 Car and Parrinello presented a way of making AIMD actually feasible by combining DFT with MD [106]. They introduced a fictitious electron dynamics:

$$\mathscr{L}(\mathbf{R},\phi) = \frac{1}{2}\sum_I M_I \dot{\mathbf{R}}_I^2 + \frac{1}{2}\sum_i \mu \frac{\partial}{\partial t}\,\langle \phi_i | \phi_i \rangle \; - V(\mathbf{R},\phi)\,, \tag{2.54}$$

where $\mathscr{L}$ is the Lagrangian of the system, ϕ an orthonormal set of molecular orbitals with elements ϕ_i, V the potential energy, M the atomic mass, and μ a fictitious electron mass. By relinquishing the requirement that electrons must be confined in their ground state (and thus have a temperature of 0) there was no need to iteratively optimize the wavefunctions. Instead the electrons were propagated through time at low electron temperature (i.e. very close to the ground state). The drawback of this approach were the introduction of high electron resonance frequencies which requires a very short timestep for successful calculations [107]. This can be alleviated to a

certain degree by tuning the ficticious mass [107].

A different approach, which allows for timesteps as large as BOMD, was introduced by Kühne et al. in 2007 [108]. It uses the always-stable predictor-corrector (ASPC) by Kolafa [109]. That method allows for the extrapolation of the wavefunction in a polynomial of order K, using the $K+1$ previous wavefunctions as support. In their adaption of the ASPC, Kühne et al. chose to extrapolate not the wavefunction $\mathbf{C}$ but the density $\mathbf{C}\mathbf{C}^T\mathbf{S}$, because the latter evolves far more smoothly [108].

$$\mathbf{C}^p\left(t_n\right) \cong \left[\sum_{m=1}^{K}\left(-1\right)^{m+1} m \frac{\binom{2K}{K-m}}{\binom{2K-2}{K-1}} \underbrace{\mathbf{C}\left(t_{n-m}\right)\mathbf{C}^T\left(t_{n-m}\right)}_{\mathbf{P}(t_{n-m})}\mathbf{S}\left(t_{n-m}\right)\right]\mathbf{C}\left(t_{n-1}\right)\,, \tag{2.55}$$

where $\mathbf{P}$ is the density kernel. Within the square brackets of Eq. 2.55 the new density $\mathbf{C}\mathbf{C}^T\mathbf{S}$ for the timestep t_n is extrapolated from the $K+1$ previous densities. The density is then back transformed to the wavefunction using $\mathbf{C}(t_{n-1})$. The extrapolated wavefunction is not idempotent. Thus the back transformation with $\mathbf{C}(t_{n-1})$ is not exact, and the error made by the back transformation adds to the error made by the extrapolation [108]. To reduce the approximation errors, the predictor step is followed by a corrector step to improve convergence towards the ground state.

$$\mathbf{C}\left(t_n\right) = \omega\ \mathrm{MIN}\left[\mathbf{C}^p\left(t_n\right)\right] + \left(1-\omega\right)\ \mathbf{C}^p\left(t_n\right)\,, \tag{2.56}$$

where

$$\omega = \frac{K}{2K-1}\ \text{ and }\ K \geq 2\,, \tag{2.57}$$

Due to the approximative extrapolation in Eq. 2.55 the orthonormality of the molecular orbitals $\phi_i = \sum_\mu \mathbf{C}_{\mu i}$ is violated. In order to restore it, the corrector step is preceded by a matrix purfication step [110]. Furthermore, iteration to convergence is avoided by considering the error resulting in the forces as a random noise Ξ_I acting upon the converged force F_I^{BO}

$$M_I\ddot{\mathbf{R}}_I = F_I^{BO} - \gamma_D M_I\dot{\mathbf{R}}_I + \Xi_I. \tag{2.58}$$

In this case, the effect of the Ξ_I can be compensated through a Langevin dynamics Eq. 2.58 on the forces, using a damping parameter γ_D. The damping parameter can be evaluated using a separate MD simulation [107]. It should be noted, that this approach turns into a BO method, when one chooses to apply the corrector step to convergence.

Periodic Boundaries

Due to the high computational cost, the simulated systems are in general far smaller than realistic systems. In these dimensions, a disproportionate amount of molecules resides at the surface rather than inside the bulk. It is easy to see that simulations of such small scale systems cannot reproduce properties found in experiment. An extended body is therefore mimicked by periodic boundary conditions (PBC). For calculations with PBC, the particles are placed in a box, which is surrounded by an infinite number of images of itself (see Fig. 2.3 for a two dimensional schematic). The vector $\mathbf{T}_R$ is the translational displacement at which any content of the original cell is repeated. A particle at distance $\mathbf{T}_R$ from a particle in the primary cell therefore has the same properties as the original particle. Thus only the particles of the original cell have to be propagated. In the event that one particle crosses the cell boundary, it is replaced by one of its images. If, for instance, a particle will leave the box on the right side, it will be replaced by a particle that moves into the box on the left side. Contrary to the isolated system, not only (N-1) interactions with other particles,

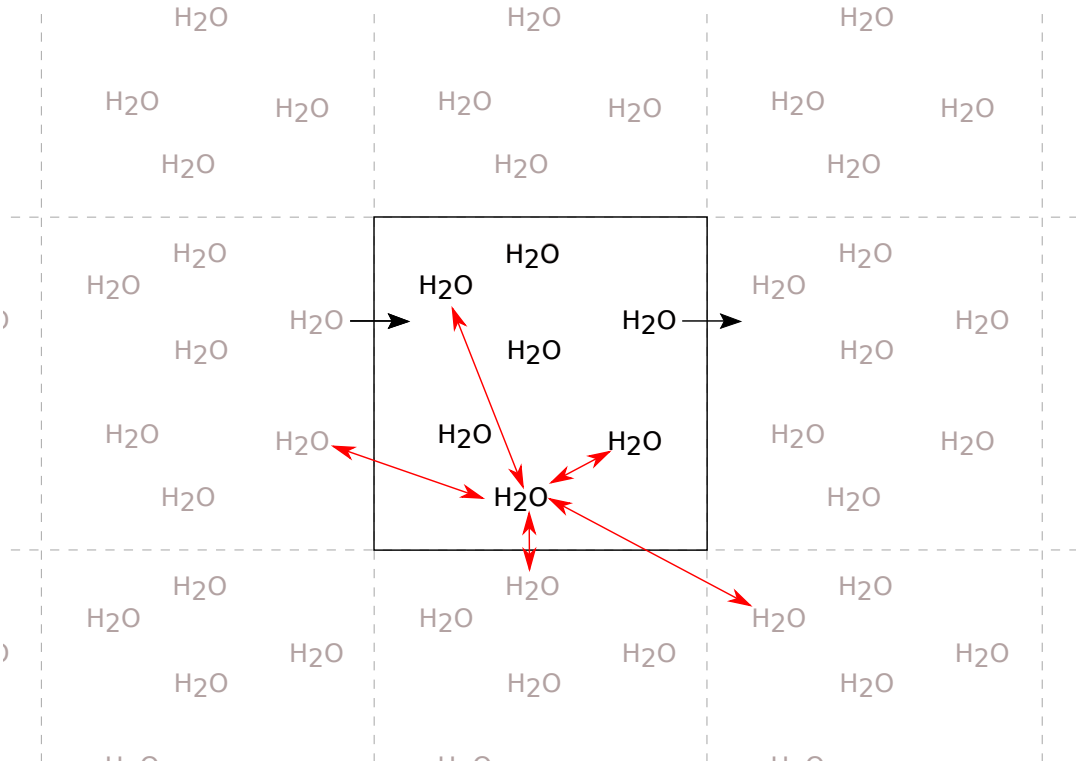


Figure 2.3: Example of a two dimensional simulation cell (black) with periodic images (gray). The black arrows depict the movement of a particle out of the simulation box. The red arrows show parts of the interactions that have to be calculated for a particle.

but also the infinitely many interactions with all the replicated particles in the images

must be calculated. Here, it is helpful to know that the potential is also periodic, because every image has exactly the same environment as the original cell:

$$V(\mathbf{r}+\mathbf{T}_R) = V(\mathbf{r}) \,. \tag{2.59}$$

However, the fact remains that, in principle, an infinite number of electrons has to be dealt with. The approaches to this problem in quantum mechanics and classical mechanics are fundamentally different. Since we used both quantum mechanical and classical methods, both approaches will be covered here.

Periodic boundary conditions for quantum mechanical forces The answer on how to approach periodic boundary conditions depends heavily on the choice of the atomic basis set. The approach presented here uses plane waves, because their regularity and infinite expansion make them naturally suited for replicated systems [50]. This suitableness stems from the fact that the wavefunction at $\mathbf{r}+\mathbf{T}_R$,

$$\Psi(\mathbf{r}+\mathbf{T}_R) = \mathrm{e}^{i\mathbf{k}\mathbf{T}_R}\, \Psi(\mathbf{r}) \,, \tag{2.60}$$

differs from the one at $\mathbf{r}$ only by a phase factor [111]. Every one-electron orbital is expanded into a basis that has the periodicity of the system.

$$\phi_{n\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{i\mathbf{k}\mathbf{r}} \sum_{\mathbf{G}} n_{\mathbf{k}}(\mathbf{G}) \mathrm{e}^{i\mathbf{G}\mathbf{r}} \,. \tag{2.61}$$

Here, n and $\mathbf{k}$ are quantum numbers. $\mathbf{k}$ is a continuous vector that lies within the first Brillouin zone and it is associated with the crystal momentum [65]. $\mathbf{G}$ is a basis vector of the reciprocal lattice with a periodicity commensurate to that of the lattice [111]. The electronic structure can be calculated from the first n functions, where n corresponds to the number of electrons, on a discrete grid of $\mathbf{k}$-points. Since ϕ converges rapidly with $\mathbf{G}$, only vectors with an energy below a cutoff energy have to be taken into account [65].

Periodic boundary conditions for classical forces In classical systems the long range terms that are affected by periodic boundary conditions are the coulomb and Lennard-Jones interaction. Both take the form of pair potentials V_{IJ} between the

atoms I and J [76, 112].

$$V = \sum_{I<J}^{\text{unit cell}} \sum_{\mathbf{T}_R} V_{IJ}(\mathbf{R}_I - \mathbf{R_J} + \mathbf{T_R}) \,. \tag{2.62}$$

In earliest simulations, all long range forces were cut-off to save computing time [113]. In order to keep either the energy or the force continuous, switching functions were employed such that $\tilde{V}_{IJ}(R_{IJ}) = S(R_{IJ}) \cdot V_{IJ}$. The switching function S could reach from a simple truncation, a shift

$$S(R_{IJ}) = \left[1 - \left(\frac{R_{IJ}}{R_c}\right)^2\right]^2 \tag{2.63}$$

where R_c is the cutoff radius, to elaborate cubic switching [114]. Improvements upon the truncation approach have provided an averaged description of the Lennard-Jones and Coulomb interaction. For instance, Coulomb interactions could be approximated by a multipole field [113]. More modern implementations have refrained from cutting off the very long ranging Coulomb interaction, and instead used sophisticated methods, in particular the Ewald summation [98]. In the following paragraph, the Ewald summation is presented for coulomb potentials.

For Coulomb interaction the term V_{IJ} of Eq. 2.62 takes the form $V_{IJ} = q_I q_J / \ \mathbf{R}_I - \mathbf{R_J} + \mathbf{T_R}$ and the charge of the atom I is $\rho_I = q_I \delta(\mathbf{R} - \mathbf{R}_I)$ [115, 116]. This charge ρ_I can be split into a long-range ρ_I^L and a short-range ρ_I^S component by adding and substracting a gaussian G function centered at I with an unspecified standard deviation [115].

$$\rho_I^L = q_I G(\mathbf{R} - \mathbf{R}_I) \tag{2.64}$$

$$\rho_I^S = q_I \delta(\mathbf{R} - \mathbf{R}_I) \tag{2.65}$$

By virtue of the poisson law, $\nabla^2 \phi = -\rho$ the potential fields experienced by I, ϕ_I, generated by I and the energy can be split into long and short-ranged terms as well. The integration of the charge density yields

$$\phi_I^S = q_I \left(1 - \operatorname{erf}\left[\frac{\mathbf{R} - \mathbf{R}_I}{\sqrt{2}\sigma}\right]\right) . \tag{2.66}$$

$$\phi_I^L = q_I \operatorname{erf}\left[\frac{\mathbf{R} - \mathbf{R}_I}{\sqrt{2}\sigma}\right] \,. \tag{2.67}$$

The usefulness of introducing the Gaussian functions now becomes clear. Upon integration, the gaussians turn into the error function. In the short range term this leads to a gradual decay of ϕ_I^S at a finite distance while the long-range part only exists at long ranges. Therefore the long range potential, contrary to the coulomb potential , does not posess a singularity [115]. Due to its finite reach the energy from the short-range potential can now be calculated directly. Due to the loss of the singularity, the long range potential can be calculated by transforming the charge density into Fourier space and calculating the Fourier transformed potential $\mathrm{F}[\phi_I^L]$, now depending on the reciprocal space vectors $\mathbf{k}$

$$\mathrm{F}[\phi_I^L](\mathbf{k}) = \sum_{J=1} q_j \mathrm{e}^{-i\mathbf{k}\mathbf{r}_j} \frac{\mathrm{e}^{-0.5\sigma^2\mathbf{k}^2}}{\mathbf{k}^2} \ . \tag{2.68}$$

Here, in the reciprocal space the potential length is truncated by the Fourier transformed gaussian $\exp(-0.5\sigma^2 k^2)$ [115]. ϕ_I^L is obtained by back transformation from the Fourier space and the energy can be calculated. Finally, the potential energy can be written as

$$V = V^S + V^L + V^{self} \tag{2.69}$$

$$\begin{aligned} &= \frac{1}{2}\sum_{\mathbf{T_R}}\sum_{I,J=1} \frac{q_i q_j}{\mathbf{R}_I - \mathbf{R_J} + \mathbf{T_R}} (1 - \mathrm{erf}\left[\frac{|\mathbf{R}_I - \mathbf{R_J} + \mathbf{T_R}|}{\sqrt{2}\sigma}\right]) \\ &+ \frac{1}{2\Omega}\sum_k \frac{\mathrm{e}^{-0.5\sigma^2\mathbf{k}^2}}{\mathbf{k}^2} - |S(\mathbf{k})|^2 \frac{1}{\sqrt{2\pi}\sigma}\sum_I q_I^2 \ , \end{aligned} \tag{2.70}$$

where V^{self} is a self introduction term introduced for formal reasons, Ω is the supercell volume, S is the structure factor, the Fourier transform of the charge density [115].

In summary, PBC removes unwanted surface effects and allows to simulate bulk phases. However, only effects that exist on scales smaller than the size of the box can be observed. Likewise, only dynamic properties with a correlation time smaller than the time it takes for spacial translations to become periodic can be sampled and in any case the tails of the correlation functions will contain artifacts. Fortunately, most properties of interest, like thermodynamics and local structures have sufficiently small scales [99]. It should further be noted, that with the introduction of periodic boundaries not all conserved quantities remain conserved. The remainder of this paragraph will be used to account how periodic boundary conditions affect the mass, the energy, and the linear and angular momentum. The mass and the center of mass motion are conserved, because every particle that leaves the box is replaced by its image, which has the same velocity as the original particle. The total

energy

$$E = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + V \ , \tag{2.71}$$

is conserved too, because the kinetic energy of the particles is unaffected by the periodic boundary and the potential is evaluated over the complete real space anyway. The angular momentum

$$L = \sum_I \mathbf{R}_I \times \mathbf{P}_I \tag{2.72}$$

depends on the particles' position with respect to the box's center. If a particle leaves the box on the right side, it is replaced by one on the left side. Therefore, L cannot be conserved at that point. Luckily, L fluctuates around its supposedly constrained value, because particles leave and enter the box on all sides [99].

Sampling at Constant Temperature

Standard molecular dynamics solves Newton's equations of motion assuming constant volume (V), constant particle count (N) and constant energy (E). This is called a microcanonical or NVE ensemble. However, most expermiments are conducted under constant pressure (P) and temperature (T), i.e. in an NPT ensemble. Due to the limited system size and simulation time, thermodynamic averages obtained in a different ensemble do differ [117]. An NPT system can be simulated by adding a barostat and a thermostat to the system. Unfortunately, the density in an NPT system converges horribly slowly [118]. Simulations that include a quantum mechanical region are therefore, often only run with a thermostat attached, thus an NVT system. The simplest method of enforcing a temperature is to simply rescale the velocities of each atom to one corresponding to the target temperature at each timestep [100] This method, however, does not reproduce any known ensemble and may produce complications with various MD extensions [98, 117]. The Canonical-Sampling by Velocity-Rescaling (CSVR) is a more elaborate variant [117]. Instead of forcing a fixed target temperature T_0, the target kinetic energy K_0 is drawn from a canonical distribution.

$$\bar{P}(K_0)\mathrm{d}K_0 \propto K_0^{3N/2-1} \mathrm{e}^{-\frac{K_0}{k_B T}} \mathrm{d}K_0 \ . \tag{2.73}$$

that represents the probability P of a system of size N to have an instantaneous kinetic energy of K_0 at an average temperature T.

The dynamics of the system are smoothed by choosing K_0 from the previous target value via a stochastic dynamic, effectively distributing the regression to the average

temperature over several timesteps [117].

$$\mathrm{d}K = (\bar{K} - K)\frac{\mathrm{d}t}{\tau} + 2\sqrt{\frac{K\bar{K}}{3N}}\frac{\mathrm{d}W}{\sqrt{\tau}} \,. \tag{2.74}$$

Here, $\mathrm{d}W$ is a stochastic Wiener process, and τ a parameter that determines the time until an energy difference is thermostated. A key advantage of CSVR is that all particles are subject to the same rescaling factor. Thus, this methods preserves the total linear momentum, the total angular momentum for non-periodic systems, and constraints, for instance on bond lengths.

2.2.1 Emulating Nuclear Quantum Effects through a thermostat

Nuclei are quantum objects, yet they are simulated classically. The neglect of the quantum nature of nuclei thus entails the omission of the quantum mechanical zero point energy and tunneling effects [119]. These effects are of particular importance for the lightest atom, hydrogen, and the hydrogen bonds formed in water [119, 120, 121]. Feynman discovered the relation between a quantum partition function and path integrals [122]. This lead to the development of so-called path-integral methods, which are reviewed here [123, 124]. In these methods the quantum nucleus is replaced by N classical beads, connected via springs. These methods allow for the accurate simulation of quantum nuclei, albeit at N times the cost of a standard MD simulation. However, a shortcut method has been proposed by Michele Ceriotti et al. It aims at reproducing the quantum postition and momentum distribution through a Generalized-Langevin-Equation (GLE) thermostat [125].

A conventional Langevin equation reads

$$\ddot{\mathbf{q}}_I = \mathbf{F}_I + \Xi_I - \gamma\dot{\mathbf{q}}_I \,, \tag{2.75}$$

where $\dot{\mathbf{q}}_I$ is the momentum of particle I and $\ddot{\mathbf{q}}_I$ the second time derivative of its mass-weighted position [103, 126] and γ_I a friction parameter. In the original presentation, the exact force $\mathbf{F}_I$ would be 0, then $\gamma\dot{\mathbf{q}}_I$ would be the viscous drag experienced by a particle moving through a liquid and Ξ_I a random agitation force transmitted from the heat bath [127]. Clearly, the concept has been transposed to other areas, for instance noise-cancellation in MD forces [108, 128, 103, 126]. Another application is using Langevin dynamics to thermostat a system. Then, $\mathbf{F_I}$ is the force derived from the system's potential and, Ξ and $\gamma\dot{\mathbf{q}}_I$ model the interaction with the thermostat [129]. The formulation still assumes that the interaction with the heat bath constitues a "white noise", i.e. it is randomized and instantaneous [129].

In a GLE, the assumption that the interaction is a white noise is lifted. Thus, a GLE thermostat models an interaction between system and heat bath, where the interaction at any time t_n depends on past events. Ceriotti et al.'s formulation circumvents the ensuing complexities by replacing the non-markovian dynamics of the $3N$-dimensional system, with the markovian dynamics of a fictitious $3N+n$-dimensional system, where n represents the number of additional degrees of freedom $\mathbf{s}$. The equations of motion for this system then read:

$$\begin{aligned} \dot{\mathbf{q}} &= \mathbf{p} \\ \begin{pmatrix} \dot{\mathbf{p}} \\ \dot{\mathbf{s}} \end{pmatrix} &= \begin{pmatrix} -\mathbf{F}(\mathbf{q}) \\ \mathbf{0} \end{pmatrix} - \underbrace{\begin{pmatrix} a_{pp}\,\mathbf{a}_p^T \\ \bar{\mathbf{a}}_p\,\mathbf{A} \end{pmatrix}}_{\mathbf{A}_p} \begin{pmatrix} \mathbf{p} \\ \mathbf{s} \end{pmatrix} + \underbrace{\begin{pmatrix} b_{pp}\,\mathbf{b}_p^T \\ \mathbf{b}_p\,\mathbf{B} \end{pmatrix}}_{\mathbf{B}_p} \begin{pmatrix} \Xi \end{pmatrix}, \end{aligned} \tag{2.76}$$

where $\mathbf{q}$ are the mass-weighted position vectors, and $\mathbf{s}$ the additional degrees of freedom, a_{pp} is the Langevin friction coefficient and b_{pp} the intensity of the random force. If n is set to zero, then $a_{pp}=\gamma$ and $b_{pp}=1$ retains Eq. 2.75. The letters $\mathbf{a}$, $\mathbf{A}$ and, a represent friction, while $\mathbf{b}$, $\mathbf{B}$ and b represent diffusion. A subscript p on these matrices indicates that the respective matrix acts also on the p,s states of the state vector $\mathbf{x}=(q,p,\mathbf{s})^T$, a subscript pq indicates a matrix over the whole state vector and no subscript indicates a matrix acting only on the s states [129]. It can be shown that Eq. 2.76 holds for generic potentials. In order to determine the parameters $\mathbf{A}_p$, $\mathbf{B}_p$ a generic system is assumed to be a collection of one-dimensional harmonic oszillators. Furthermore, since the fluctuation-dissipation-theorem has to be satisfied the following relation can be set up [129].

$$k_B T(\mathbf{A}_p + \mathbf{A}_p^T) = \mathbf{B}_p \mathbf{B}_p^T \,. \tag{2.77}$$

The entries of $\mathbf{A}_p$ are then fitted to guarantee a high sampling efficiency and minimal hindrance of diffusion [129].

In order to emulate nuclear quantum behavior, the GLE thermostats specific frequencies such that the atoms' position and momentum distribution closely obey the relations $\langle \dot{\mathbf{q}} \rangle = \omega^2$ and $\langle \mathbf{q} \rangle = \frac{\hbar\omega}{2}\coth(\frac{\hbar\omega}{2k_BT})$. Since the relation Eq. 2.77 cannot be used, these distribution functions are used to determine the entries of $\mathbf{A}_p$ and $\mathbf{B}_p$ [125]. Due to the representation of a system as a collection of generic harmonic oscillators, these parameters don't need to be evaluated for each system, but can be calculated once and then reused for a wide variety of applications [129].

2.2.2 Sampling the Free Energy Surface

A vast variety of important chemical properties, such as the reaction rates and solubility are determined by the free-energy differences between chemical states, i.e. their relative position on the free energy surface (FES) [130]. Unfortunately, the FES scales with the number of atoms in a system, making unimaginably expansive in most practical cases. The sampling of any event that requires a system to pass a barrier that is greater than the kinetic energy available to the system becomes computationally unfeasible [131], because, not only is such a barrier-crossing extremely unlikely, the system is inclined to explore any direction with a lower barrier first. Thus, if one intends to study a so-called rare event in the time alloted to accomplish the average PhD or PostDoc, he is required to force the system in the direction of the rare-event of interest.

The first task is accomplished by defining arbitrary reaction coordinates ξ, which is a subset of the dimensions of the system. For instance, ξ could be the distance between two molecules, if one is interested in a binary reaction. The second task, restated, means to modify the systems probability distribution $P(\xi)$ along the reaction coordinate. A method to obtain the probability distribution is equivalent of finding the free energy, because the two are directly connected

$$F(\xi) = -k_b T \ln\left[P(\xi)\right] + f \,. \tag{2.78}$$

Here, T is the system's temperature and f is a negligible constant [132]. The respective means of how these methods modify $P(\xi)$ and then arrive at F, however, vary greatly. The simplest method, known as thermodynamic integration or blue moon method [133], is to modify the systems Lagrangian

$$\tilde{\mathscr{L}}_i = \mathscr{L} + \lambda_i(\xi - \xi_i) \,, \tag{2.79}$$

to constrain the ξ to a set of targets ξ_i, perform an MD at each constrained ξ_i and then integrate numerically over the average value of the Lagrangian multipliers λ_i [134]. However, more sophisticated approaches have been developed, two of which are Metadynamics (MTD) [131, 135] and Umbrella Sampling (US) [133]. They are advantageous over a constrained MD, because they allow to sample $P(\xi)$ at every value of ξ, not only at ξ_i.

In MTD, a biasing potential is added to the reaction coordinate space such that

$$V^b(\xi,t) = \sum_{i<t} w_i \exp\left[\frac{\xi - \xi_i}{\sqrt{2}\sigma}\right]^2 , \tag{2.80}$$

where w_i is the height of the biasing gaussian and σ is its width. The variable t indicates, that the biasing gaussians are added one by one over the course of the simulation [132]. In Well-Tempered MTD the height w_i of any new Gaussian not yet included in the bias potential V^b is calculated by

$$w_i = \omega_0 i_0 \exp\left[-\frac{V^b(\xi,t)}{k_B \Delta T}\right] , \tag{2.81}$$

where ω_0 is the initial deposition rate of the bias potential, i_0 is the time step at which gaussians are added to the bias potential and ΔT is a temperature tuning parameter that limits the exploration range of the FES [132, 135]. Thus, over time the growth of the bias potential decreases such that the system can reach an equilibrium state [135]. Then the free energy

$$F(\xi) = -\frac{T + \Delta T}{\Delta T} \lim_{t\to\infty} V^b(\xi,t) + f , \tag{2.82}$$

with f being an immaterial constant [132].

At the outset, Umbrella Sampling is similar to a constrained MD. The reaction coordinate is discretized into several bins at ξ_i The sampling in each bin is biased by a potential

$$V_i^b(\xi) = \frac{1}{2} K(\xi - \xi_i)^2 , \tag{2.83}$$

where K is the spring constant [133]. In principle potentials other than the harmonic can be also be used. For a sufficently large value of K the system is now forced to follow the bias potential instead of the FES, sampling all directions orthogonal to ξ at ξ_i. This procedure generates a set of biased probability distributions $P_i^b(\xi)$. The challenge now consists in finding a way to extract the original $P(\xi)$ from this set.

$$P_i^u(\xi) = P_i^b(\xi) \exp\left[\frac{V_i^b(\xi)}{k_B T}\right] \exp\left[-\frac{F_i}{k_B T}\right] . \tag{2.84}$$

Usually, if more than one sampling bin is used the value of F_i cannot be extracted from the simulations directly. Instead specialized method to analyze the Umbrella Sampling have to be used. Two of them, which are presented here [133], are called the weighted histogram analysis method (WHAM) and umbrella integration (UI)

The weighted histogram analysis method attempts to find the optimal global unbiased distribution, i.e. the bin weighting procedure should be constructed such that $\partial\sigma^2[P(\xi)]/\partial p_i = 0$ [136, 137, 133]:

$$P(\xi) = \sum_i p_i(\xi) P_i^u(\xi) \tag{2.85}$$

$$p_i(\xi) = \frac{N_i\,[-\beta(V_i^b(\xi)+F_i)]}{\sum_j N_j \exp[-\beta(V_j^b(\xi)+F_j)]}\,, \tag{2.86}$$

where $\beta = 1/k_B T$ is the inverse temperature and N_i is the number of samples in bin i. The F_i depend on $P(\xi)$ requiring Eq. 2.85 to be solved iteratively [133]. Since F_i is defined as

$$\exp[-\beta F_i] = \int P(\xi)\exp\left[-\beta V_i^b(\xi)\right]\mathrm{d}\xi\,, \tag{2.87}$$

P_i^u can be eliminated from Eq. 2.85 and the optimization now turns into an iterative search for the optimal F_i [137]

$$\exp[-\beta F_i] = \int \sum_i \frac{N_i \exp[-\beta V_i^b(\xi)] P_i^b(\xi)}{\sum_j N_j \exp[-\beta(V_j^b(\xi)+F_j)]}\mathrm{d}\xi\,. \tag{2.88}$$

Umbrella integration circumvents the calculation of F_i by averaging the mean force instead of P_i^u [133].

$$\frac{\partial F_i(\xi)}{\partial \xi} = -k_B T \frac{\ln[P_i^b(\xi)]}{\partial \xi} - \frac{\mathrm{d}V_i^b}{\mathrm{d}\xi} \tag{2.89}$$

The biased probability distribution is then expanded into a cumulant series that is truncated after the second term, effectively assuming a gaussian distribution for P_i^b.[133]. The global force is obtained by using a weighted sum over the bin forces

$$\frac{\partial F(\xi)}{\partial \xi} = \sum_i \frac{N_i P_i^b(\xi)}{\sum_j N_j P_j^b(\xi)} \frac{\partial F_i(\xi)}{\partial \xi}\,. \tag{2.90}$$

Here N_i is the number of samples in bin i. The unbiased force now only has to be integrated to obtain the free energy. One of the great advantages of UI is that the error of $F(\xi)$ can be calulated [133, 138]. Furthermore, this sampling technique allow the use of statistical methods that ascertain that $\bar{\xi}_i^b$ and σ_i^b don't suffer from statistical errors or unequilibrated trajectories [138].

2.3 ALMO-EDA

2.3.1 Absolutely localized MOs

Fragmentation schemes are used to divide the wavefunction into local parts that can be computed independently. The motivation for this is twofold: a) Localizing a wavefunction effectively reduces the computational cost of large systems, enabling the implementation of linearly scaling methods and b) providing an intuitive picture of the quantum system [139, 140].

One of those methods is called self-consistend-field for molecular interactions (SCF-MI) [141]. There, the system of electrons is grouped into molecular fragments x. The condition for the creation of fragments is that the interaction within a fragment is strong in comparison to the interaction between fragments [141]. In this method, molecular orbitals are forced to reside within a localization domain. The domain encompasses all atomic orbitals of fragments within a distance R_c of the fragment the electron forming the MO belongs to.

$$|\phi_{xi}\rangle = \sum_{y \in U} \sum_{\mu}^{n_y} |\chi_{y\mu}\rangle \, c_{yi}^{y\mu} \; . \tag{2.91}$$

Here, ϕ_{xi} is the MO i of fragment x, $\chi_{y\mu}$ is the AO μ in the localization domain and $c_{yi}^{y\mu}$ is the domain specific MO coefficient. U is the set of all fragments within R_c of fragment x. Obviously, the domains overlap, which leads to a slow convergence of the wavefunction [42]. As a remedy, the implementation of Khalliullin et al. [42] sets R_c to zero in a preliminary optimization step, which limits the expansion of MOs to a single fragment. These can then be optimized independently. Next, the constraint on R_c is relaxed and the wavefunction is diagonalized once. In a variant method the diagonalization can be repeated, which is useful for linearly scaling QM methods [42].

2.3.2 Energy Decompositon

The true strength of the method is however revealed, when the steps of the ALMO procedure are coupled with energy decomposition analysis, hence assigning them a physical meaning.

$$\Delta E = \Delta E_{\mathrm{FRZ}} + \Delta E_{\mathrm{POL}} + \Delta E_{\mathrm{CT}} \; , \tag{2.92}$$

where ΔE is the entire energy reduction obtained from placing a fragment in the system. It is subdivided into to the *frozen-core* energy ΔE_{FRZ}, the *polarization energy* ΔE_{POL} and the charge transfer energy ΔE_{CT} [43] ΔE_{FRZ} is the energy obtained distorting the atoms of fragment x from its isolated equilibrium position. ΔE_{POL} the energy from calculating the molecular orbitals of fragment x in the vicinity of other fragments.

The term E_{CT} is of particular interest, because, for instance the strength of hydrogen bonds can be linked to it [43]. Furthermore, E_{CT} can be split into individual donation and back-donation terms between atoms

$$\Delta E_{\mathrm{CT}} = \sum_{I,J} = E_{I\to J} + E_{J\to I} + \Delta E^{\mathrm{HO}} \,, \tag{2.93}$$

where $E_{I\to J}$ is the energy obtained from transferring charge from an occupied orbital of I to a virtual orbital of J and E^{HO} contains the difference to the converged solution of the unconstrained system. The calculation of the individual terms requires switching to a biorthogonal basis set where a molecular orbitals of x have no overlap with neither other MOs from the same fragment, nor with any MO from any other fragment [44]. The individual CT energy between two molecules can then be denoted as

$$E_{I\to J} = \sum_{i}^{o_I}\sum_{j}^{v_j} F_{Ja}^{Ii} X_{Ii}^{Ja} \,, \tag{2.94}$$

where i is an orbital from the set of occupied orbitals o_I on I , and a is an orbital from the set of virtual orbitals v_J on J. F_{Ja}^{Ii} is the biorthogonal representation of the Fock matrix and is associated with the energy obtained from transferring one electron between the two orbitals, whereas X_{Ii}^{Ja} is an amplitude that denotes how much charge is actually transferred [43]. The EDA method also accounts for the basis set superposition error (BSSE), although it has been found to be negligible for the medium to large basis sets regularly used in DFT calculations [140].

3 Computational Details

Configurations for ALMO EDA were obtained from *ab initio* molecular dynamics simulations [98] , whereas forces were calculated by DFT with a DZVP basis set. The exact simulation details may be taken from [142]. The simulations were designed to reproduce the dispersion interactions between the water molecules, which improves the local description and the overall density of water [143, 144]. The simulation of the water surface was performed at constant temperature and constant volume in a 15.604 15.604 84.000 Å box containing 384 water molecules. Periodic boundary conditions were applied in all three dimensions. ALMO EDA was carried out using the BLYP [63, 64] functional for 700 snapshots evenly distributed over a trajectory of 70 ps. A detailed description of the simulations can be found in Ref. [145].

Throughout the paper 'donor' describes a molecule *donating electrons* to another molecule, while 'acceptor' describes a molecule receiving electrons from another molecule. In this fashion $E_{D\to A}$ denotes a donor bond from a specific donor molecule to a specific acceptor molecule, whereas $E_{A\leftarrow D}$ denotes an acceptor bond between two specific molecules. In contrast, $E_{D\to J}$ and $E_{A\leftarrow J}$ describe donor, or acceptor bonds with generic partners. Furthermore E_{Don} and E_{Acc} describe all donor and acceptor interactions of a molecule

$$E_{\mathrm{Don}} = \sum_J E_{D\to J} \qquad E_{\mathrm{Acc}} = \sum_J E_{A\leftarrow J}\,, \tag{3.1}$$

with all neighboring molecules J as defined by the ALMO procedure [42, 43, 44]. Since only few neighbors, mostly 2, contribute significantly to E_{Don} or E_{Acc} [45], only the 5 strongest interactions were actually considered.

The definition of bond asymmetry by Kühne and coworker was used and is given here for the reader's convenience

$$\gamma_{\mathrm{Don}} = 1 - \frac{\Delta E^{1\mathrm{st}}_{D\to A}}{\Delta E^{2\mathrm{nd}}_{D\to A}}\,, \qquad \gamma_{\mathrm{Acc}} = 1 - \frac{\Delta E^{1\mathrm{st}}_{A\leftarrow D}}{\Delta E^{2\mathrm{nd}}_{A\leftarrow D}}\,, \tag{3.2}$$

where $\Delta E^{1\mathrm{st}}$ and $\Delta E^{2\mathrm{nd}}$are the energies of the strongest and second strongest donor, or acceptor bond respectively [45]. In case of geometric hydrogen bonds, these were

identified by utilizing the concept of Luzar and Chandler [146] with the enhanced cut-off parameters based on the PMF of hydrogen bonds [145]. The mass-density ρ was used to define the positon of the dividing surface plane, such that $\rho_{\text{surf}} = \frac{1}{2}\rho_{\text{bulk}}$. The actual surface positions were calculated with the instantaneous surface definition of Willard and Chandler [147]. The surface volume was divided into indiviual layers following the categorization of Kessler et al. [119], where the first, second and third surface layer ended at 0.5 Å, 3.5 Å, and 6.5 Å below the dividing surface, respectively. All water molecules farther than 6.5 Å from the surface were considered bulk like.

X-ray absorption spectra were obtained with CP2K [87, 86], using the GAPW approach [148, 149], within CP2Ks QM/MM framework. The MM atoms of the slab were represented by the flexible TIP3P [150] model and the interaction was modeled using the multigrid approach by Laino et al. [82, 89], expanding the interaction into 12 gaussians. The MM-radii of O and H were set in accordance with the findings in [82]. The QM region was selected by including every molecule within a 6.0 Å distance of the excited molecule and simulated using unrestricted Kohn-Sham DFT. The wavefunction was expanded in the IGLO-III basis set [151], while the auxiliary plain-wave grid was cut off at 320 Ry. The QM zone was decoupled from its periodic images using the the multipole decoupling scheme [90]. Calculations were carried out for 700 of the 3501 snapshots available.

The x-ray photon acting on the 1s orbital of oxygen was emulated with the half core-hole transition potential [152, 153]. In this method the XAS signal is calculated by removing a fraction of an electron – exactly half an electron for the half-core-hole potential – from the orbital absorbing the exciting photon. This electron is delocalized in the conduction band, i.e. it is not promoted into a specific orbital. All orbitals are optimized in presence of the core-hole and the absorption frequencies are calculated from the energy differences between the core orbital and all virtual orbitals. The absorption intensity can be calculated from the dipole transition element

$$I \propto |\left\langle \psi'_f \right| \mathbf{E}\hat{\mathbf{D}} \left| \psi'_i \right\rangle|^2 , \tag{3.3}$$

where I is the absorption intensity, $\mathbf{E}$ is the electric field that represents the incoming photon, $\hat{\mathbf{D}}$ is the dipole operator [41, 153]. Line broadening is manually applied afterwards [154] The transition potential method is known to overestimate the post-edge and underestimate the pre-edge peak [45, 154], but despite this these methods have been found to semiquantitavely reproduce the XAS spectrum of water. [154, 39].

Error values and their corresponding error bars were determined by integrating the corresponding student t-function for the sample's degrees of freedom with standard numerical techniques [155]. The errors reported below thus represent the 95% confidence interval for the mean [156, 157]. All graphs were produced with matplotlib [158] and fitting procedures were conducted with scipy [159, 160]. In all fits data points with a confidence interval greater than 2 times the respective error of the mean were omitted. Data points in plots, except for the two dimensional distributions, were assumed to be averages over normal distributions. The validity of the assumption was tested by fitting the distribution over each fixed value to a gaussian with scipy and assuring the significance of the fit with an F-test [157] using $\alpha = 0.05$. Fitting procedures were performed using the Levenberg-Marquardt algorithm [161] implemented in SciPy and the statistical computations were accomplished using SciPy [159, 160].

4 Energetic Cooperation in Hydrogen Bond Formation

Cooperation describes the interrelation of hydrogen bonds. It has been known to exist for decades but its extent has never been measured. Here, we use a novel approach. Instead of relying on geometric definitions, we relate the interaction energies of a water molecule. A first insight into cooperative behavior is given by the joint distribution function $p(E_{\mathrm{Don}}|E_{\mathrm{Acc}})$ in Fig. 4.1.

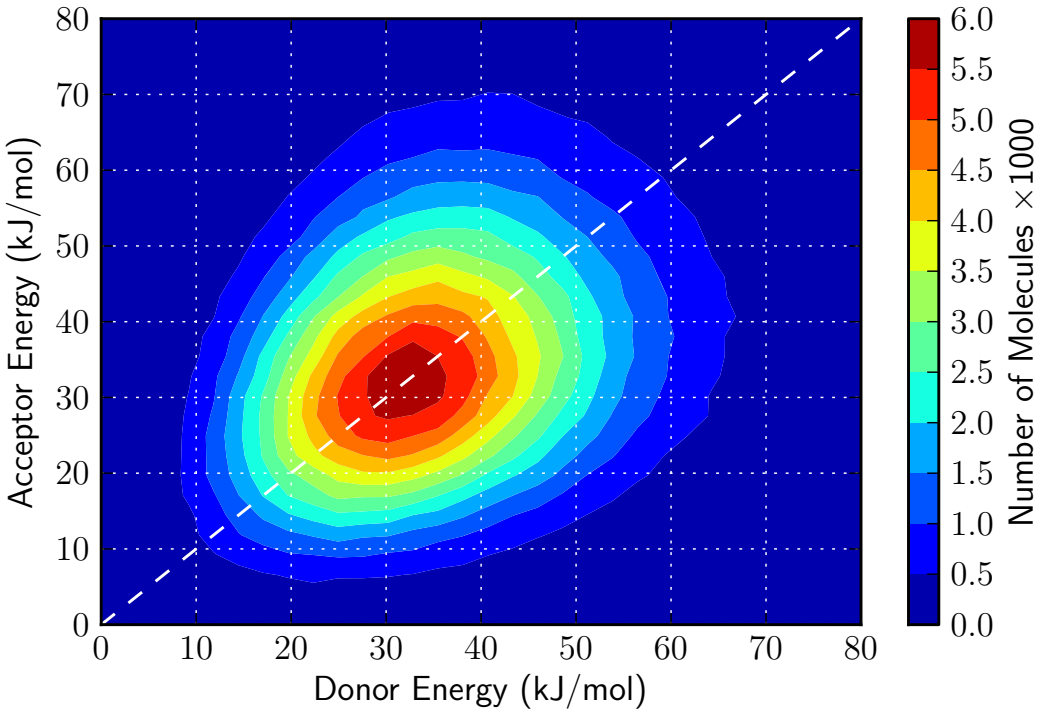


Figure 4.1: Distribution of the E_{Don} versus E_{Acc} of a bulk molecule. Its symmetry with respect to the isoenergetic line as well as the elongation along it, suggests an interrelation between the two quantities. Figure gives the raw molecule count over all snapshots.

Clearly, the distribution is symmetric with regards to the iso-energetic line with the maximum being placed directly on that line. Furthermore, the distribution is slightly elongated regarding said iso-energetic line. Even though the approach is rather sim-

		acceptors		
	D\A	1	2	3
donors	1	23.4(1)	25.09(5)	24.9(4)
	2	17.32(3)	18.99(1)	18.8(1)
	3	11.71(8)	12.87(2)	12.3(3)

Table 4.1: Change of $\langle E_{\mathrm{D}\to\mathrm{J}} \rangle$ with the number of cooperative and anti-cooperative bonds. Values are given in $\frac{\mathrm{kJ}}{\mathrm{mol}}$.

ple, this zeroth-order approach to cooperation so to speak, it already indicates an interrelation between E_{Don} and E_{Acc}.

As hydrogen bonds are the carriers of molecular water interaction each bond should carry a part of the total energy depicted in Fig. 4.1. One may presume that molecules with few bonds tend to be located in the lower left of that figure, i.e. they have a low E_{Acc}. Since acceptor and donor energy are related, these molecules can be expected to have a low E_{Don} as well. This in turn entails a low number of donor bonds. The resulting relation between the number of acceptor and the number of donor bonds is already a well known fact [23]. Instead, the relation between the number of bonds and their energy will be investigated. The corresponding data is presented in Tab. 4.1 which gives the donor energies with respect to the number of bonds attached to a molecule. The bond energy reduces roughly by half between one and three donor bonds, indicating a fair amount of competition between donor bonds. The number of acceptors affects the donor bond energy as well. In contrast to the linear progression found for donor bonds, here the highest bond energy is found at two acceptors; an observation that is in contradiction with the assumption presented above.

These observations suggest the following model to describe hydrogen bonds. In this model, a water molecule attempts to balance the hydrogen bond energies. In order to get acceptor and donor energies balanced, energy received by a molecule through its acceptor bonds is pushed to its donor bonds. In turn, acceptor bonds pull energy from the donor bonds. If a molecule has many acceptor and only few donor bonds then the energy of two bonds is redistributed to a single bond, which consequently becomes stronger. In contrast, if a molecule has many donor and few acceptor bonds, then the energy of the donor bonds have to be redistributed to two bonds, leaving fewer energy for each of them. Furthermore, addition of a cooperative bond will increase the bond strength, while addition of an anti-cooperative bond will diminish it. The model presented here neatly explains the reduction of the en-

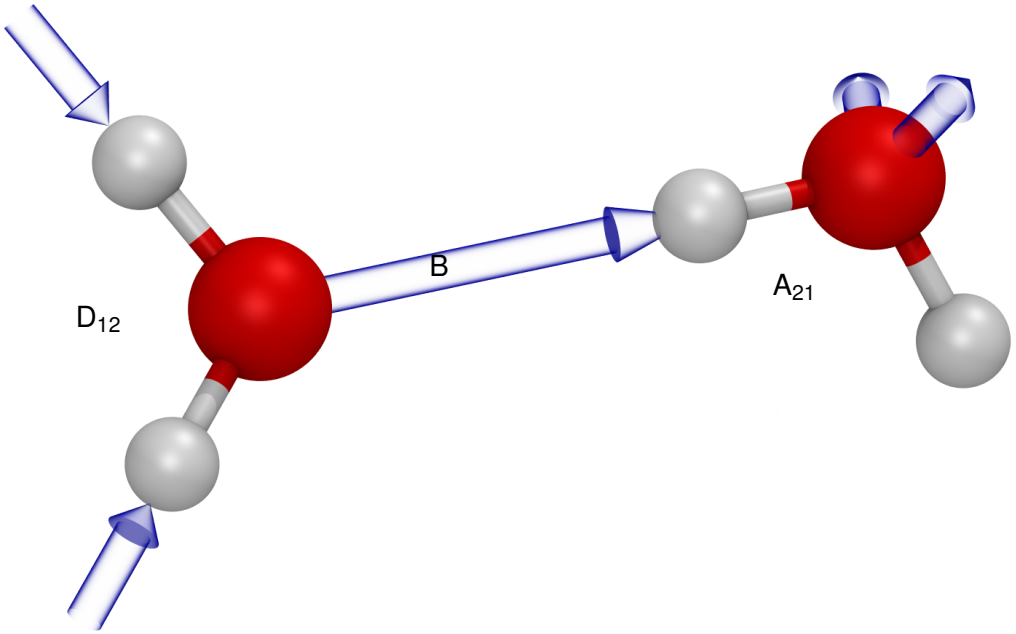


Figure 4.2: Schematic of a molecule pair connected by a bond 'B' whose donor 'D_{12}' has 1 donor and 2 acceptor bonds and whose acceptor 'A_{21}' has 2 donor and 1 acceptor bond. Picture rendered with vmd [162].

ergy in Tab 4.1, but does not explain the loss of energy in the third column. However, as any hydrogen bond connects two molecules, its strength is determined by all the bonds these two molecules engage in. Therefore, in order to properly test how a hydrogen bond reacts to other bonds in its vicinity, molecule pairs have to be observed.

The bond energies of $E_{D\to A}$ of pairs of water molecules with the most common numbers of donor and acceptor bonds are presented in Tab. 4.2. The entries are ordered in such a way that $E_{D\to A}$, when moving from top left to bottom right increases in Tab. 4.2a and decreases in Tab. 4.2b. In Tab. 4.2a we see that $E_{D\to A}$ increases between the first and second column. Pairs in the second column have an additional acceptor bond at 'D'. This extra bond transfers extra energy to 'D' which gets pushed into 'B'. Likewise, the transition from A_{22} to A_{21} strengthens 'B' as well, because it becomes the only bond on which the pull from A's two donor bonds is exerted. When these observations are generalized, one finds that every addition of an acceptor bond to 'D' and every addition of a donor bond to 'A' increases $E_{D\to A}$. These bonds will be henceforth called *cooperative* bonds. Likewise, every addition of a donor bond to 'D' or every addition of an acceptor bond to 'A' decreases $E_{D\to A}$,

	D_{21}	D_{22}	D_{11}
A_{11}	17.0(2)	18(2)	22.8(7)
A_{22}	17.65(5)	19.30(2)	23.3(1)
A_{21}	19.0(2)	20.68(8)	25.4(4)

(a)

	D_{12}	D_{11}	D_{22}
A_{22}	25.10(6)	23.3(1)	19.30(2)
A_{11}	23.9(4)	22.8(7)	18(2)
A_{12}	22.4(1)	21.2(2)	17.16(3)

(b)

Table 4.2: Variation of $\langle E_{D\to A}\rangle$ between the most common pair configurations. Values are given in $\frac{\text{kJ}}{\text{mol}}$. An illustration of the labels' meaning can be found in Fig. 4.2.

and these bonds will be called *anti-cooperative*. A more intricate behavior is obtained when one looks at the simultaneous addition of a cooperative and anti-cooperative bond. In cases involving only acceptor bonds the energy remains unchanged, as one would expect. In cases involving donor bonds one finds an increase and a decrease of the bond energy. For instance, at the transition from A_{11} to A_{22} the added pull from the donor bond supersedes the anti-cooperative effect from the acceptor bond. The addition of two bonds to 'D' shows the opposite effect. It appears as if the addition of an anti-cooperative bond to 'D' interferes with the orbitals participating in 'B', thus weakening the bond. This kind of orbital interference appears to be of minor significance in the case of anti-cooperative acceptor bonds. Thus the cooperative behavior observed, i.e. the change of a bonds energy with respect to the change number of donor /acceptor bonds appears to be modulated by steric hindrance when the additional bond is added to the oxygen where 'B' resides. Overall, these data clearly show the increase in a bond's energy upon addition of a cooperative bond in the most common molecules.

In order to ease quantification of this cooperative effect and incorporate all possible pairs in an easily readable manner we use a measure introduced by Ohno et. al [14]. The hydrogen bond offset

$$o_H = D_{\text{Acc}} + A_{\text{Don}} - D_{\text{Don}} - A_{\text{Acc}}\ , \tag{4.1}$$

gives the surplus of energy increasing bonds over energy reducing bonds. Here, D_{Acc} is the number of acceptor bonds connected to molecule 'D' and A_{Don} is the number of donor bonds connected to 'A'. The negative terms represent anti-cooperative contributions, the positive represent cooperative bonds.

Within this picture, the example schematic Fig. 4.2 represents a pair denoted '$D_{21}A_{12}$', which is a strong pair with an offset $o_H = +2$. In fact, the average $E_{D\to A} =$

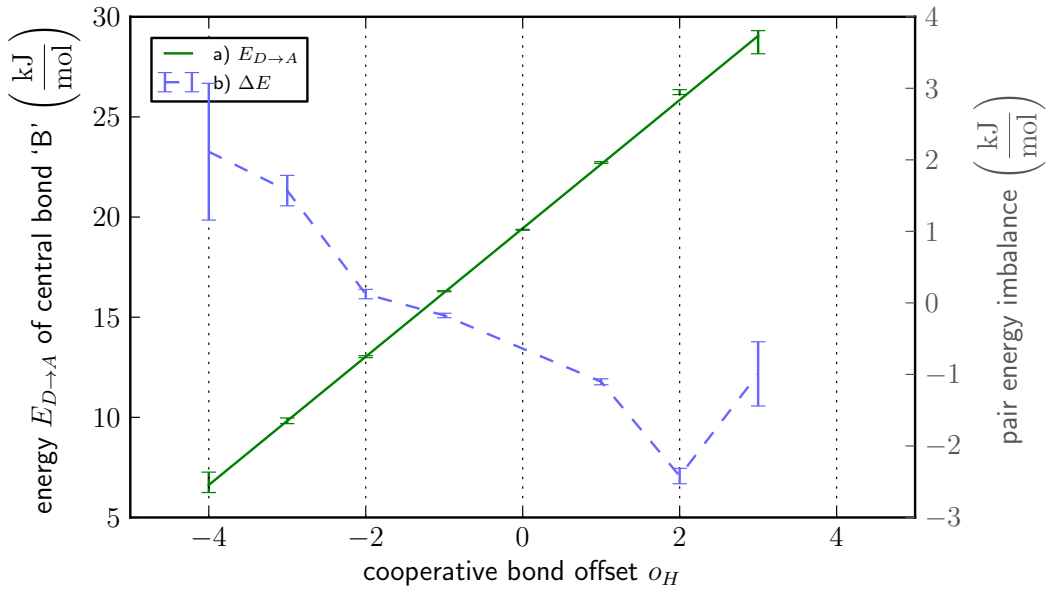


Figure 4.3: a) effect of cooperation on the strength of a bond. b) average amount of energy on a molecule not balanced by hydrogen bonds.

$26.9(1)\frac{\text{kJ}}{\text{mol}}$, a value well above any of the values presented in Tab. 4.2. The most common pair '$D_{22}A_{22}$' has an offset $o_H = 0$. With a lower offset the average bond energy of the latter pair is well below that of '$D_{21}A_{12}$', as '$D_{22}A_{22}$ has two additional energy reducing bonds. o_H is depicted in Fig. 4.3a. As can be seen, the increase of 'B's strength is perfectly linear over the entire range of offsets. On a side note, a slight asymmetry in Fig. 4.3a exists in the form of the presence of an offset of -5 but the absence of an offset of +5. This is an artifact of the definition, because the mere presence of 'B' produces an offset of -2. Due to this shift fewer combinations for high positive offsets exist. It should be noted that $E_{D\to A}$ continues to increase for molecules with high offsets, in contrast the donor energies of single molecules in Tab. 4.1. Offsets are pair quantities and do not apply to single molecules. In order to enable a comparison, the average offset of all pairs containing a donor D_{13} and D_{12} has been calculated. The average offset of D_{13} is $o_H = 2.09$, compared to $o_H = 1.06$ for D_{12}, suggesting a higher value for $E_{D\to J}$ for the former. Tab 4.1 however reads $E(D_{13} = 24.9\frac{\text{kJ}}{\text{mol}}$, while $E(D_{12} = 25.1\frac{\text{kJ}}{\text{mol}}$. Furthermore the reduction of $E_{D\to A}$ when moving from D_{12} to D_{13} can also be observed for specific pairs, i.e. when the number of bonds at 'A' is controlled. From a physical perspective, three acceptor bonds require two of them to be bound to the same hydrogen atom. Due to its size, both bonds will not be positioned optimally, thus possibly reducing the amount of charge actually transferred, resulting in weaker bonds. From the fact that this does not show in Fig. 4.3a, it can be inferred that these pairs are comparatively rare.

Despite the fact that o_H can only have integer values, a regression through the data points can be made to obtain information about the devolution of the energy of 'B' upon addition or removal of a bond. A linear fit $f(x) = a \cdot x + b$, is produced for the graph Fig. 4.3a, which returns the following values: $a = 3.272 \pm 2 \cdot 10^{-7}$ and $b = 19.437 \pm 1 \cdot 10^{-6}$. These numbers imply the following physical picture of a hydrogen bond energy; namely that it consists of a cooperative part and an inherent energy. The cooperative part, represented by the incline a, adds $\approx 3 \frac{\text{kJ}}{\text{mol}}$ to the hydrogen bond energy per cooperative bond and reduces the energy by the same amount per anti-cooperative bond. The inherent hydrogen bond energy would then be the value b. However, with the current definition, $o_H = 0$, encompasses predominantly the pair '$D_{22}A_{22}$' From the amount of bonds, and the fact that this is the default configuration in water, one can induce that cooperative effects make up a considerable portion of $E_{\text{D}\to\text{A}}$. In contrast, the only pair that is with absolute certainty not influenced by cooperation is the water dimer '$D_{10}A_{01}$'. Unfortunately, even the long trajectory used here only has four of such pairs, all in the second surface layer. Instead of pairs without cooperation we look for pairs in which the cooperative and the anti-cooperative effect even out, i.e. where the number of energy increasing bonds equals the number of energy reducing bonds. These pairs can be found at $o_H = -2$. Because 'B' alone contributes -2 to o_H, the net effect of all other bonds is zero. The energy of pairs with this offset is $12.99(5) \frac{\text{kJ}}{\text{mol}}$ for the bulk phase; a value which is reasonably close to to the binding energy of the water dimer which is given as $13.0(4) \frac{\text{kJ}}{\text{mol}}$ [13, 163, 164].

The reason why $E_{D\to A}$ depends on other bonds was that a water molecule seeks to balance its donor and acceptor energies. If this assumption is true then for all values of o_H, the values of E_{Don} and E_{Acc} must even out at each molecule constituting a pair. A glance at Fig. 4.1 indicates that this is not perfectly true at all times. But that distribution is centered at equal values of E_{Don} and E_{Acc}, hence the magnitude of the imbalance is expected to be small for the majority of molecules. This imbalance is measured by

$$\Delta E = \frac{1}{2}\left(\sum_J E_{D\to J} - \sum_J E_{D\leftarrow J} + \sum_J E_{A\to J} - \sum_J E_{A\leftarrow J}\right) , \tag{4.2}$$

where the first tho terms give the difference between donor and acceptor energies at 'D' and the second two terms give the difference between donor and acceptor energies at 'A'. As Fig. 4.3b reveals, the imbalance between donor and acceptor energies on a molecule has no clear relation to the energy of the central bond or the offset of the pairs. For most offsets the imbalance is small compared to the bond

strengths and it is virtually zero for the most common pairs at $o_H = 0$. The fact that ΔE is low for most pairs is interpreted as a result of the fact that molecules want to remain neutral, and it can be concluded that energy balancing is indeed the motivator for cooperation.

The previous discussion established a quantitative measure for using the geometric definition of hydrogen bonds as a basis. However, this geometric definition has two shortcomings. One, it does not account for the individual hydrogen bond strengths, and two it is inherently arbitrary. The uniquely determined ALMO energies can be used to overcome these two issues [44]. In order to incorporate these advantages into our offset, an alternative measure to o_H is proposed.

In the definition of o_H, Eq. 4.1, four bond types were identified and the number of bonds of each type was summed up to give the offset value. The improved description o_E replaces the amount of each bond type by their respective energetic contribution. Hence,

$$o_E = \overbrace{\sum_J E_{D\leftarrow J}}^{\equiv D_{\mathrm{Acc}}} + \overbrace{\sum_J E_{A\rightarrow J}}^{\equiv A_{\mathrm{Don}}} - \overbrace{\sum_J E_{D\rightarrow J}}^{\equiv D_{\mathrm{Don}}} - \overbrace{\sum_J E_{A\leftarrow J}}^{\equiv A_{\mathrm{Acc}}} + 2E_{D\rightarrow A}\,, \tag{4.3}$$

where, $E_{D\leftarrow J}$ gives the energy of an acceptor bond at 'D', $E_{A\rightarrow J}$ that of a donor bond at 'A', $E_{D\rightarrow J}$ gives the energy of an donor bond at 'D' and $E_{A\leftarrow J}$ that of an acceptor bond at 'A'. The set of partners J is different for each sum, but always includes all partners of the respective type. The braces give the equivalent contribution in Eq. 4.1. The term $2E_{D\rightarrow A}$ removes the energy of 'B' from the equation, to avoid the trivial dependence of $E_{\mathrm{D}\rightarrow\mathrm{A}}$ on o_E. Given that the desired behavior of o_E is equal to that of o_H, i.e. $o_E \propto E_{D\rightarrow A}$, inclusion of $E_{D\rightarrow A}$ must be avoided.

The desired properties stated at the beginning of this section require the inclusion of non-bonded water interactions into the calculation of o_E. Non-bonded interactions are ALMO-EDA charge transfer energies, that do not correspond to a geometric hydrogen bond. As such, their energy is significantly below that of hydrogen bonds. In the resulting figure, Fig. 4.4a the monotonous increase of $E_{D\rightarrow A}$ is retained, however, the incline is decidedly non-linear. A comparison of $E_{\mathrm{D}\rightarrow\mathrm{A}}$ at $o_H = 0$ and $o_E = 0$ reveals that the interaction energies at $o_E = 0$ are only about a third of the bond energy observed in Fig. 4.3a. Furthermore, the interaction energies for $o_E < 0$ are virtually zero. These properties indicate that o_E is to be interpreted differently than o_H (Eq. 4.1). Where o_H gave the cooperative enhancement of a bond, starting from a base value, o_E cannot be interpreted in the same manner.

Before the discussion of Fig. 4.4 is continued, we make a few remarks on the scale of o_E. First, given that o_E has the dimension $\frac{\text{kJ}}{\text{mol}}$, values of $o_E > 100$ appear to be rather large. However, Fig. 4.1 shows that the difference between a molecule's donor and acceptor energy and acceptor energies of a molecule can easily amount to approximately 30 kJ/mol. Since such a difference can compound in a pair, and the energy of 'B' is taken out of the equation, values that large can easily occur.

Fig. 4.4a is easiest explained in terms of transferred charge, even though energies have been plotted. This is possible because of the proportionality between charges and energy [165]. At large positive offsets o_E the cooperative bonds are significantly stronger then the anti-cooperative bonds, leading to the accumulation of negative charge at 'D' and positive charge at 'A' In turn, values of $o_E < 0$ indicate that for this pair the anti-cooperative bonds dominate the o_E, leading to the accumulation of negative charge on 'A' and positive charge on 'D'. The formation of a bond would only transport more electrons – negative charge – to 'A'. That negative charge is consequently a strong disincentive towards the formation of a hydrogen bond in those pairs. Instead of interpreting o_E as cooperative enhancement, like o_H, it is better to interpret o_E as an incentive to a pair to form a hydrogen bond between them. At very low values of o_E, the charges on donor 'D' and acceptor 'A' are such that a hydrogen bond would do nothing to reduce the charge on either molecule, and no bond is formed. At high o_E the opposite is the case, and the hydrogen bonds are consequently strong. Thus, o_E provides a quantified description of the hydrogen bond cooperation in water.

This idea is corroborated by the energy imbalance residual in a molecule pair, given in Fig. 4.4b. At variance to the definition, Eq. 4.2, all interactions have been included in the calculation. Contrary to Fig. 4.3b, ΔE shows a clear trend here. Comparison of Eq. 4.3 and Eq. 4.2 shows that this can only be the case if $\sum_J E_{D \leftarrow J}$ grows faster than $\sum_J E_{D \rightarrow J}$. The negative value for high offsets may be explained by the high volatitlity of the donor position [46]. It further can be expected that the amount of energy carried is not infinite. The imbalance is thus the result of two acceptor bonds simply being able to deliver more energy than can be funneled by the donor bond, which additionally blocks formation of another donor bond. Given that at low offsets $o_E \leq 0$ a bond is rarely formed, the reason for the imbalance there is more difficult to grasp. It may possibly be the result of a strong bond between another pair involving of the molecules. Despite the constant drift it is worth noticing, that the value of ΔE over the entire range of o_E is small compared to the bond strengths

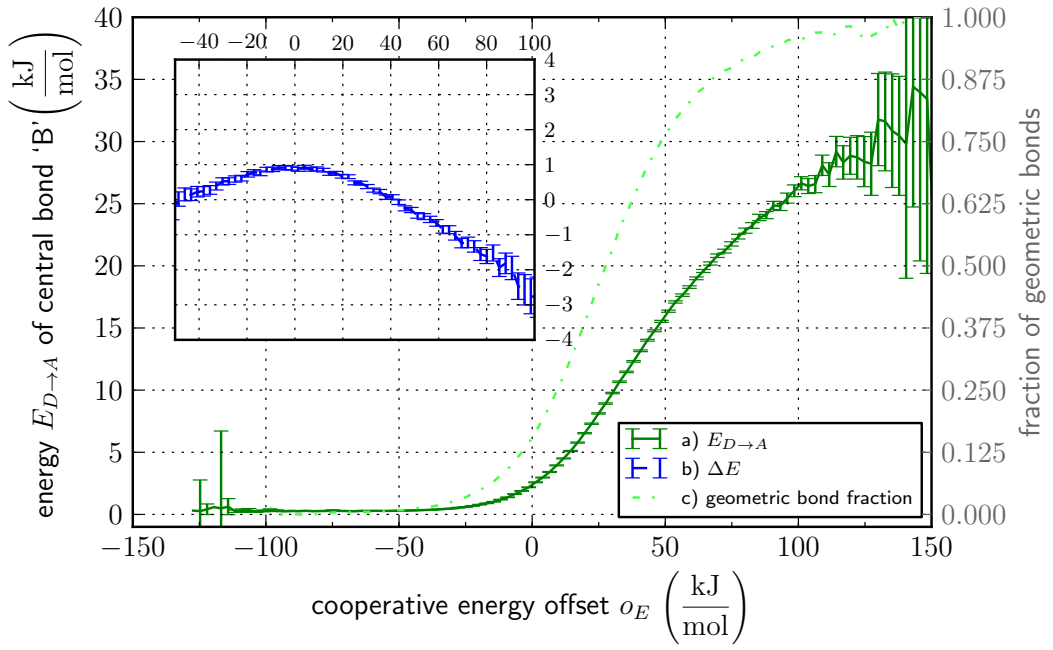


Figure 4.4: a) The increase of $E_{D\to A}$ upon heightened cooperative demand; b) The capability of 'B' to cancel E_{Don} and E_{Acc} on the connected molecules. c) The likelihood of a bond being present between two molecules.

indicating that the assumption of a charge compensation mechanism remains valid. A few notes should be added on the nature of the non-bonding interactions. One, the average oxygen-oxygen distance of $4.94(1)$ Å between the interacting molecules, suggests that many of these interactions occur with second shell neighbors. While this explains the weakness of the interaction, further conclusions cannot be drawn at this point. Two, earlier versions of ALMO-EDA did not cut off interaction at vanishing overlap. In such a case a phenomenal amount of non-bonded interactions would be added, whose physical significance is questionable. However, their contribution to o_E is marginal. It is hence to be expected that the large cooperative and anti-cooperative contributions in 'D' and 'A' even out. Therefore, it is unlikely that such an interaction would reach large positive values of o_E, just as the distribution of the now recorded interactions already indicates. Therefore, our conclusions are not affected by the number of non-bonded interactions.

In going from o_H to o_E, the measure for cooperative enhancement of bonds is lost. However, relating the $E_{D\to A}$ to energies rather than bond numbers is certainly desirable, as the energy, not the bond number determines how much charge is neutralized. If the evaluation of Eq. 4.3 is restricted to geometric hydrogen bonds, Fig. 4.5a is obtained. The graph retains the linear dependence of Fig. 4.3a, as well as the

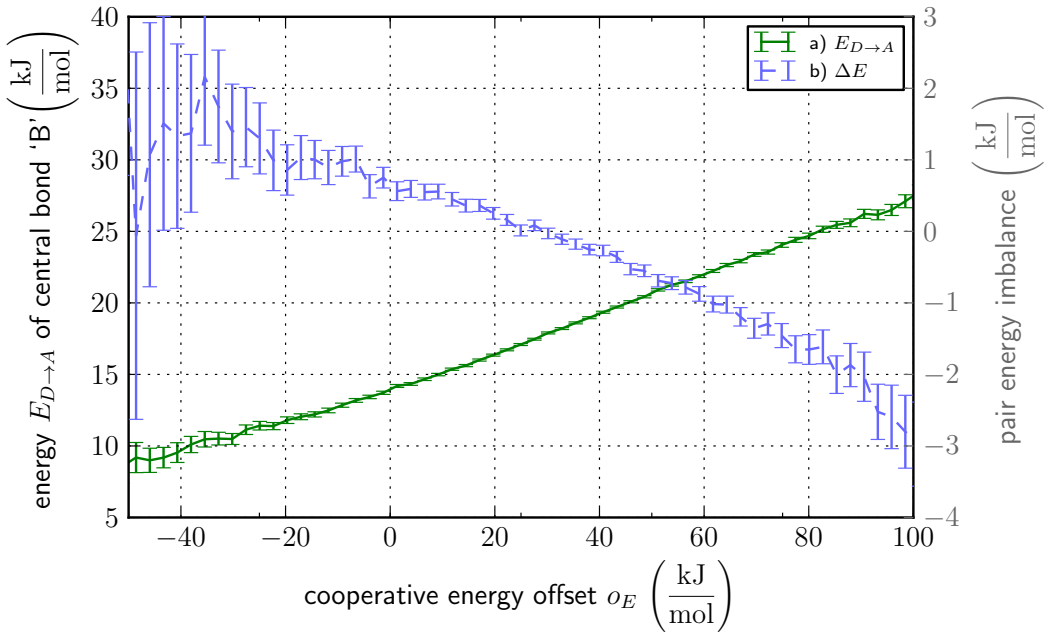


Figure 4.5: a) average strength of a geometric bond 'B' with respect to the offset o_E. b) pair energy imbalance of geometric bonds.

small incline $E_{D\to A}/o_E$. The question remains, whether Fig. 4.5a should be interpreted in terms of Fig. 4.4a, or in terms of Fig. 4.3a. In essence, should o_E be interpreted as cooperative enhancement like o_H, or should it be interpreted as the incentive to form bonds. Since $E_{D\to A}$ has a significant value at $o_E = 0$ and the energy does not go to zero even for relatively large negative values of o_E, it is apparent, that it is more readily interpreted as a measure for the cooperative enhancement of bonds. As such, a link between the values of o_E and o_H can be established. With an energy of $\approx 13\frac{\text{kJ}}{\text{mol}}$ the value $o_E = 0$ is energetically as well as conceptually equivalent to $o_H = -2$. Furthermore $E_{D\to A}$ at $o_E = 94{,}51\frac{\text{kJ}}{\text{mol}}$ corresponds best to the value at $o_H = 2$. The resulting link between the two offsets then reads

$$o_H = 23.6295 \cdot o_E - 2 \tag{4.4}$$

In short, the restriction of o_E on geometric bond offer sort of a compromise definition between o_H and the o_E of Fig. 4.4a. It gives a cooperative enhancement to bonds, but contrary to o_H it also takes the energy of the hydrogen bonds into account and is thus, in principle, more precise.

In this chapter, cooperation in water is elucidated. Cooperative bonds are always defined relative to a hydrogen bond. They deliver charge to a donor or take charge

from an acceptor, while anti-cooperative bonds do the opposite. Molecules want to balance the charges that are received and given via hydrogen bonds. If all relevant hydrogen bonds are accounted for, then the strength a hydrogen bond can be directly determined from the difference between the energies of all cooperative bonds and the energies of all anti-cooperative bonds.

5 Cooperative Effects at the Water Surface

In the previous chapter, the interrelation of hydrogen bond energies in bulk has been elucidated. Hydrogen bond energies are proportional to charges transported from a donor to an acceptor, and molecules prefer to have zero charge. This tendency of molecules is the cause for the interrelation of hydrogen bond energies. Molecule pairs were studied to obtain a measure o_E (Eq. 4.3) for how much a bond is affected by cooperation. A molecule pair (see illustration Fig. 4.2), consisting of a donor and an acceptor, are connected by a central hydrogen bond. With respect to that central bond, cooperative bonds are defined as bonds that fulfill the opposite role. If the central bond is a donor bond on one of the molecules, then a cooperative bond is an acceptor bond on that molecule. Likewise, a bond that is anti-cooperative with respect to the central bond fulfills the same role, i.e. if the central bond is a donor bond, then the anti-cooperative bond is also a donor bond.

It turned out that bulk water is highly cooperative, with cooperation accounting for up to 50% of a bond's strength. As one approaches the surface of water, the hydrogen bond network becomes increasingly disturbed. This shows, for instance, in the coordination number of water [119]. The number of hydrogen bonds a molecule is coordinated to drops from an average of 3.70 hydrogen bonds in bulk, to 1.86 in the 1^{st} surface layer. Furthermore a preferential ordering with respect to the surface normal emerges. The incidence of hydrogen bonds fluctuates with increasing intensity for all orientations upon approaching the surface from the bulk. The fluctuation of the density here is consistent with density fluctuations at the surface that have been reported for water [119], as well as other liquid surfaces [166, 167]. A graphical depiction of the occurrence of various hydrogen bond orientations can be found in the appendix, Fig. 5.1. Moreover, it shows that the molecules which reside very close

	1	2	3	4
$E_{D\to A}$	15.25	18.65	19.42	19.58

Table 5.1: reduction of average bond energy towards the surface

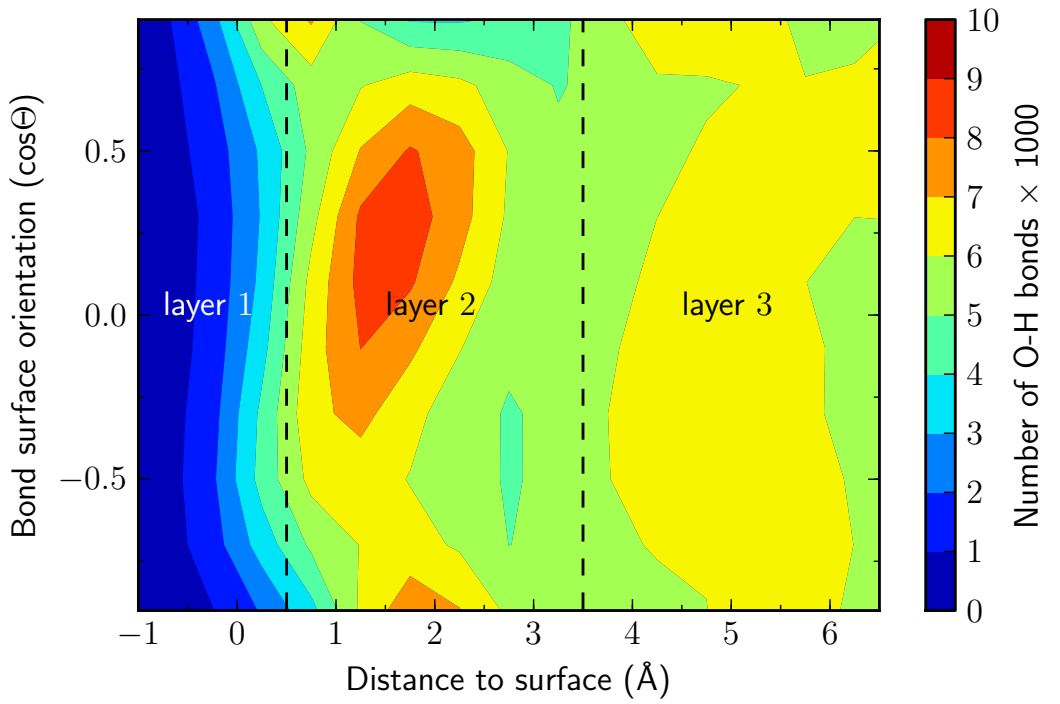


Figure 5.1: Propensity of O-H bond orientation at the surface. Molecules aligned with surface normal ($\cos\Theta \approx 1$) are preferentially found in the first surface layer. O-H bond facing the bulk ($\cos\Theta \approx -1$) are found below them.

to the dividing surface, i.e. at a distance value of 0, they preferably have an O-H bond aligned with the surface normal ($\cos\Theta \approx 1$). In contrast to these molecules, those that are situated below them, i.e. at a distance value of 2, have one O-H preferentially oriented towards the bulk ($\cos\Theta \approx -1$). Despite the severe disorder of the hydrogen bond network, cooperation remains a relevant factor to the bond energy. Fig. 5.2 shows that the bond energy in bulk is directly dependent on the cooperative energy offset o_E, where ,the stronger the offset, the stronger the bond energy. This is as true for the bulk phase as for the 2$^{\text{nd}}$ surface layer, which is evident from their virtually overlapping graphs. The similarity of the graphs ascertain that the physics of cooperation have not changed. In essence, water molecules at the surface are motivated to have neutral charge as the molecules in bulk. Furthermore, they pass on received charges to the same extent through hydrogen bonds as molecules in the bulk do.

The inset of Fig. 5.2 shows that the distribution of bonds over o_E averages at 0 for the 1$^{\text{st}}$ layer. This is a considerable reduction in comparison to the bulk value of 15 indicating that bonds in this layer are significantly less enhanced by coopera-

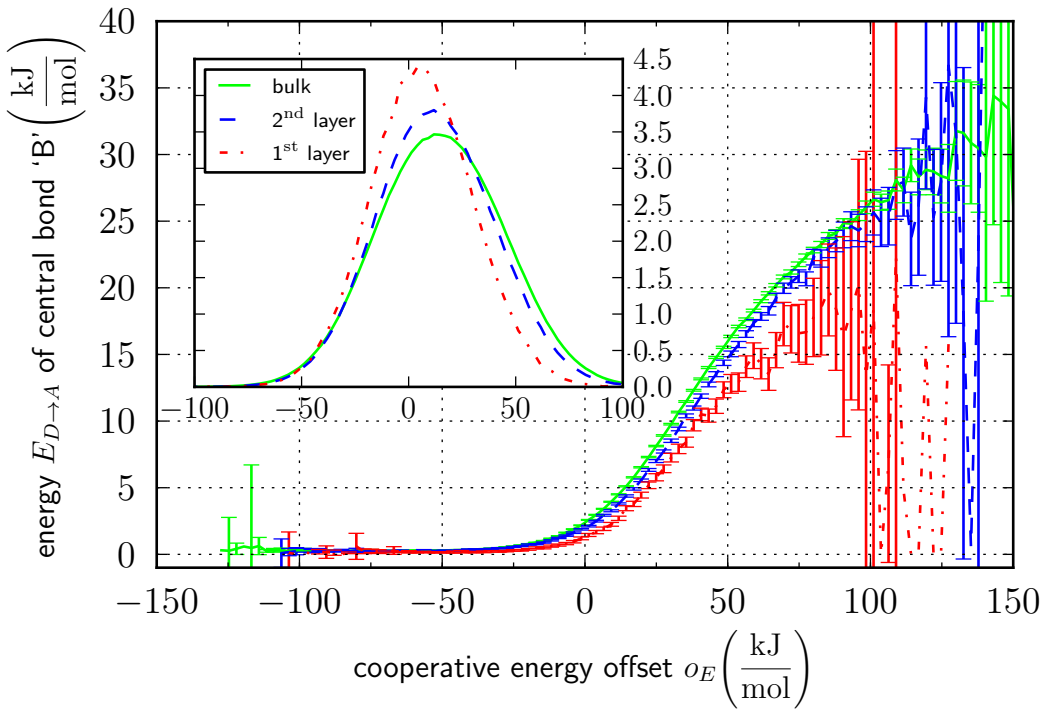


Figure 5.2: similarity in the dependence of $E_{D\to A}$ on o_E for different surface layers. Inset shows the distribution of bonds in each layer normalized to 1.

tive interaction. In contrast, the 2nd surface layer shows only little deviation from the bulk distribution. This suggests that the cooperative enhancement here is almost as strong as in bulk, despite the aforementioned reduction of the number of bonds. As the inset shows, the average distribution of o_E does not change much from bulk to the 2nd surface layer, corroborating the surprisingly strong bond energies. These results are consistent with Tab. 5.1, as the bonds' energy is only significantly reduced directly in the first layer. Up to then, despite the lower number of bonds, the bond energy holds up quite well. In the 2nd surface layer the number of bonds has reduced by $\approx 14\%$ per molecule, whereas the energy has only reduced by $\approx 5\%$. To explain why the bond energy holds up rather well, despite the reduction of the bonds, o_E is split up into its cooperative and anti-cooperative contributions. Then, the distribution of pairs with respect to these two contributions is studied. The inset of Fig. 5.3 shows that in bulk, the pairs are distributed elliptically around the maximum. The maximum of the distribution is thus considered to be an appropriate stand-in for the entire distribution, and the discussion is continued with the maxima alone.

The comparison of the cooperative and anti-cooperative energies of the bulk maximum and layer 2 maximum in Fig. 5.3 give the following energy changes. Upon

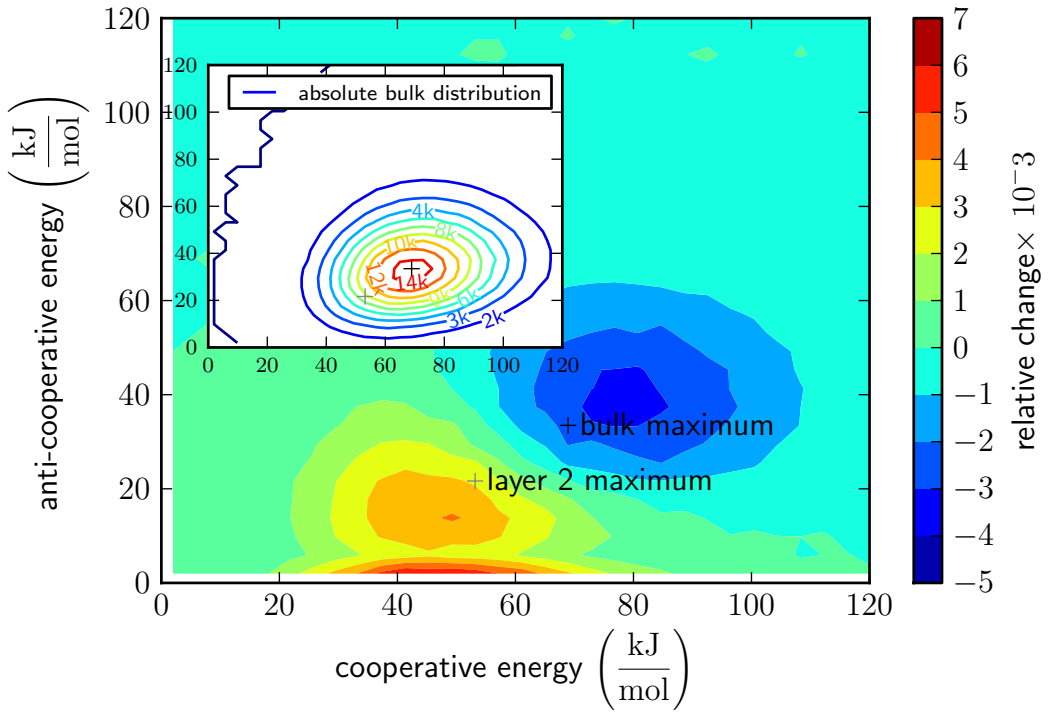


Figure 5.3: concerted shift to lower cooperative and anti-cooperative energies in the 2nd surface layer. Inset shows bulk distribution for comparison.

transition from bulk to the 2nd layer the cooperative energy reduces by $15.75\frac{\text{kJ}}{\text{mol}}$, the anti-cooperative energy reduces by $11.82\frac{\text{kJ}}{\text{mol}}$. Thus, relative to their values in bulk, the anti-cooperative energy loses more. Moreover, one can see that in the second surface layer there exists a significant amount of molecule pairs where the contribution of anti-cooperative interactions to the offset o_E is virtually zero. In terms of bonds this corresponds at least to a molecule pair that has been depicted in Fig. 4.2, i.e. a pair with a central bond only enhanced by cooperative bonds. However, given the generally lower amount of cooperative energy of this group, these pairs likely miss cooperative bonds as well. This indicates that at the surface the hydrogen bond network is reduced to fewer pathways with less branches through which charge is transported from molecule to molecule. Hence, the charge that is transported through hydrogen bonds does not have to be shared that often between two anti-cooperative bonds, which is why the energy of hydrogen bonds is affected to a rather small extent in the 2nd surface layer.

The sudden competitiveness of molecules without anti-cooperative can be traced to the a shift in the energy balance landscape. Fig. 5.4 shows the average pair energy imbalance, i.e. how well the molecules in a pair neutralize the charges they

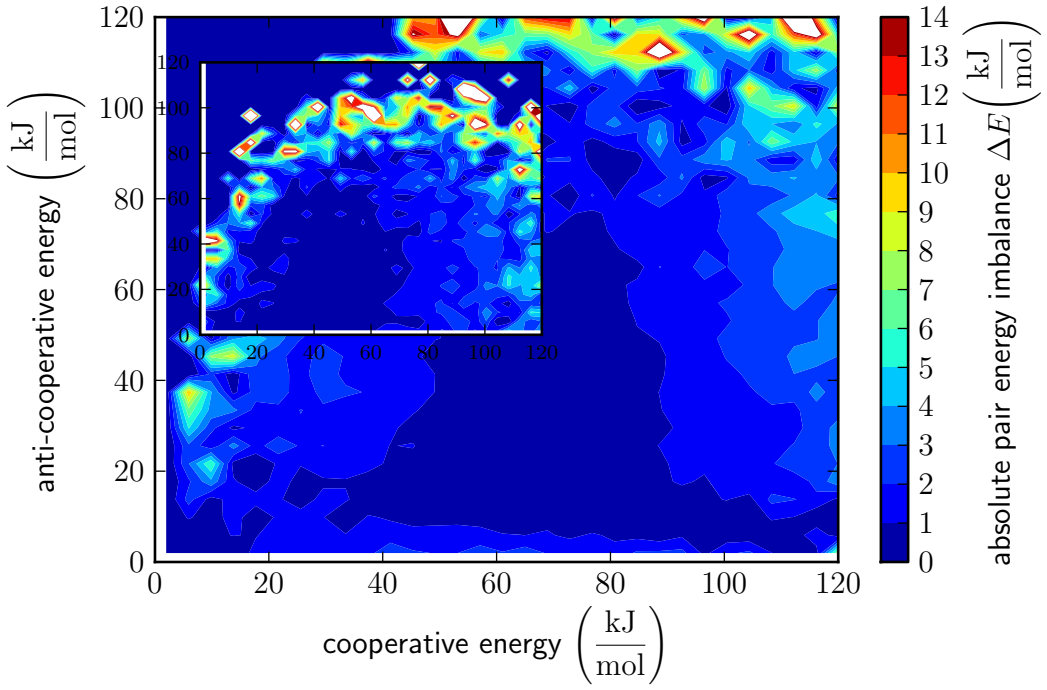


Figure 5.4: Difference between the efficacy of balancing donor and acceptor energies in bulk (main) and 2nd surface layer (inset).

receive from hydrogen bonds. A value of 0 thereby signifies that a molecule pair is well able to do so and thus, that the bonds it maintains constitute a desirable state. The higher the value, the less desirable the state is. The highest values of ΔE are a result of the low molecule count at the fringes of the distribution. Comparison of the bulk distribution with the surface distribution reveals that the energies at which ΔE is zero, shift towards lower values when going to the surface. In the bulk phase, the area where no anti-cooperative interaction is measurable has a high imbalance $1 - 3\frac{\text{kJ}}{\text{mol}}$. At the surface however, this imbalance reduces to $< 1\frac{\text{kJ}}{\text{mol}}$ indicating, that these configurations have become far more viable. The existence of these pairs that have no anti-cooperative bonds is related to the change of a different parameter characterizing water, the bond asymmetry γ.

The study of the surface revealed another effect, namely that the strength of a hydrogen bond becomes dependent on the orientation of the associated O-H bond. Fig. 5.5 shows that beyond the surface, at distances < 0, the bond energy is low for all orientations. There, as evidenced by Fig. 5.1 the density is very low, meaning that barely any molecule is present. Hydrogen bonds cannot experience cooperative enhancement, because there are only few partners with which a hydrogen bonded pair could form additional bonds. However, hydrogen bonds of molecules with O-H

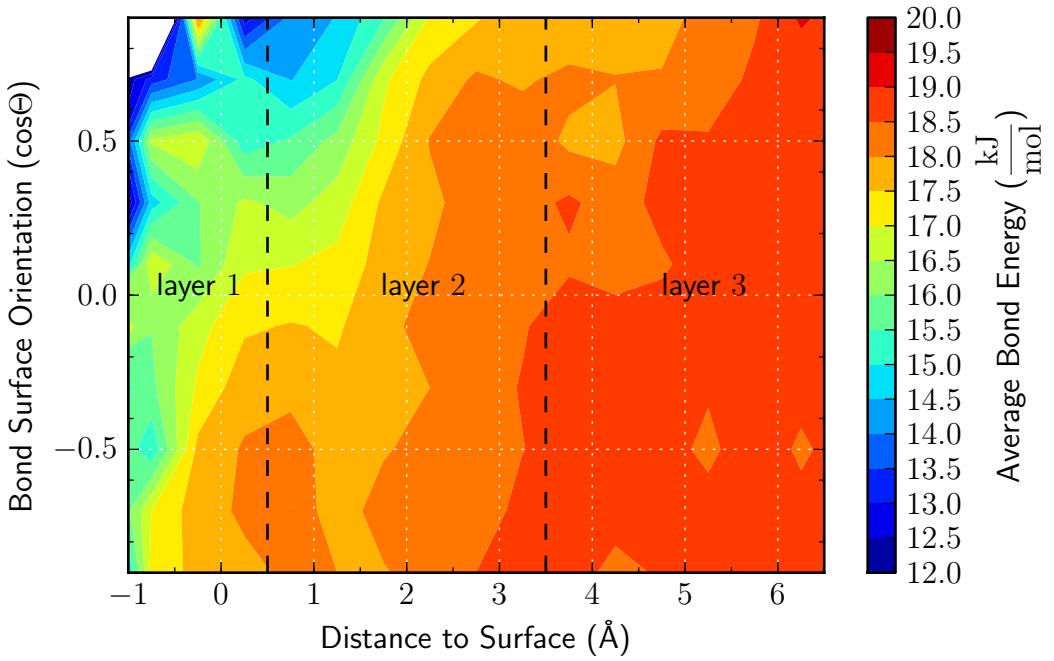


Figure 5.5: Average hydrogen bond energy with respect the surface distance of the acceptor and the surface orientation of the corresponding O-H bond. In the white area bonds are weaker than $12\frac{\text{kJ}}{\text{mol}}$.

oriented towards are weak at far greater depths than molecules where the corresponding O-H points in the opposite direction. This weakness extends well into the 2nd surface layer if their corresponding O-H group is oriented towards the surface. In order to investigate this feature we use a common [119, 168] model of how water molecules are arranged at the surface, that is based on preferred surface orientations, as discussed in Fig. 5.1. In this picture, the topmost water molecules have one O-H bond facing the surface. The molecules form hydrogen bonds into the water slab, to molecules that have one O-H bond pointing towards the bulk. A graphical depiction can be found in Fig. 5.6. The two bond types in this model can be distinguished by their angles. The angle corresponding to the bond $E_{D\to A}$ is $\cos\Theta_D = 0.52$, whereas the angle corresponding to $E_{A\leftarrow}$ is $\cos\Theta_A = -0.44$. These orientations connect our model to the continuous distribu-

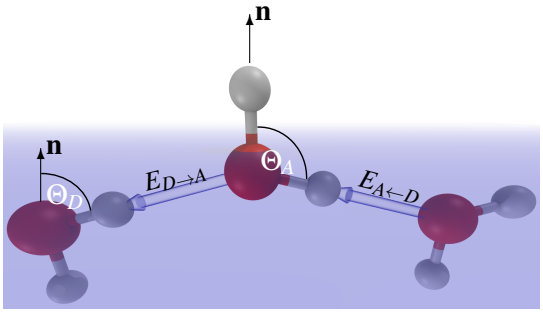


Figure 5.6: Model of hydrogen bonding between molecules at the water surface. $\mathbf{n}$ is the surface normal and Θ are angles to the surface normal.

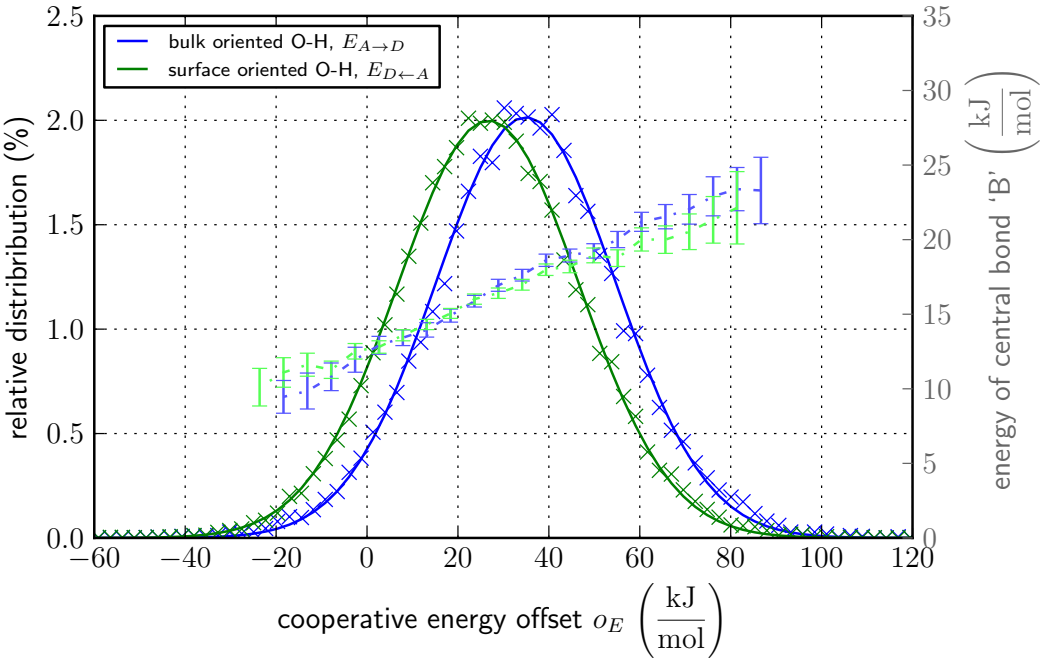


Figure 5.7: Depiction of the lower cooperative offset of bonds $E_{D\to A}$ compared to $E_{A\leftarrow D}$, which entails a lower average energy (light colors).

tion in Fig. 5.5. A look at the bonding pattern reveals that molecules in the first surface layer form 1.021(5) donor and 0.840(5) acceptor bonds, revealing a great imbalance between the number of donor and acceptor bonds. Of those bonds, 80% are bonds between the 1st and the 2nd surface layer [119]. In molecules that form bonds between layers, this imbalance is retained. The 1st layer molecule donates 1.239(5) bonds while accepting 1.102(4). This imbalance is easily understandable in our model. With one O-H bond pointing towards the vacuum, a 1st layer molecule has can still donate two bonds, but accept only one.

A hydrogen bond's strength relates to the amount of excess charge on the respective donor and acceptor molecule, as has been shown in Sec. 4. A hydrogen bond was strongest when the donor molecule had accepted many bonds and the acceptor had donated many bonds. As discussed in Fig. 5.2 the energy of surface hydrogen bonds depends in the same way on the cooperative offset o_E as in bulk. Therefore, we expect that one bond, $E_{A\leftarrow D}$, see Fig. 5.6, is stronger than the other, $E_{D\to A}$. We find this to be the case, with $E_{D\to A} = 16.1(1)\frac{\text{kJ}}{\text{mol}}$ and $E_{A\leftarrow D} = 17.5(1)\frac{\text{kJ}}{\text{mol}}$. Using our measure for the extent of cooperation to plot the bond energy, Fig. 5.7, we see that the graphs for $E_{D\to A}$ and $E_{A\leftarrow D}$ have virtually equal inclines and effectively overlap over entire range of o_E. It can thus be stated, not unexpectedly, that $E_{D\to A}$ and $E_{A\leftarrow D}$

depend in the same way on o_E, because the two graphs overlap. The property in which they differ, is their Gaussian distribution. The molecules molecules connected via $E_{D\rightarrow A}$ average at lower values of o_E. At those values the need of molecules to neutralize charge via that bond is reduced and the bond's energy is consequently lower. Thus, the energy difference between $E_{D\rightarrow A}$ and $E_{A\leftarrow D}$ can entirely be explained by cooperation.

Overall, the study of hydrogen bonds at the surface brought to major results. First, maintaining many bonds at the surface does not balance the energy of a molecule. Instead, these molecules are connected by fewer bonds. These bonds are to a higher degree cooperative to one another. Second, cooperation at the surface leads to an orientation dependence of hydrogen bond energies. This is the result of the preferancial surface orientation of molecules and cooperation. The topmost molecules, with one O-H pointing to the vacuum, must have fewer acceptor bonds, therefore they have weaker donors than acceptors.

6 Relating X-Ray Absorption to Hydrogen Bonds

At the surface of water, molecules exhibit a preferential orientation of their O-H bonds [119]. The phenomenon of cooperation, as shown in chapter 5, makes the energy of hydrogen bonds become dependent on the orientation of the molecules O-H bond. Here, we discuss how surface orientation, and the electronic effects linked to it affect the X-Ray absorption (XA) spectrum of water. In order to remain clear and concise, we will incorporate the molecules of the first 1st and 2st surface layer under the label surface.

One aspect of the aforementioned preferential surface orientation is that the molecules in the 1st surface layer tend to have one O-H bond aligned with the surface normal ($\cos\Theta \approx 1$). This means they point into the vacuum and as such lack the potential partners to form hydrogen bonds with. As we have shown previously, the preferential surface orientation brings about an orientation dependent hydrogen bond energy. As XAS is sensitive towards the local environment of a molecule, we find that also the XA response is sensitive to the orientation of a molecule, Fig 6.1.

Most notable is the decline of the pre-edge peak at 536 eV with increasing surface angle. It is prominent when the O-H bond is aligned with the surface normal. As has been stated above, these molecules exist predominantly in the first surface layer. Thus, the existence of the peak at $\cos\Theta \approx 1$ corresponds well to the association of the peak with broken hydrogen bonds. On the other hand, at angles where both O-H face downward ($\cos\Theta < 0$), a strong signal at 542 eV, which will be called post-edge signal, appears. The post-edge signal is associated with the strong bonds of ice. The average bond in ice is significantly stronger than even the bond of bulk water [46], which in turn have a higher energy ($19.58\ \frac{\text{kJ}}{\text{mol}}$), than downward facing O-H bonds at the surface. The hydrogen bonds associated with downward facing O-H bonds have an energy below $19\ \frac{\text{kJ}}{\text{mol}}$ and are the strongest bonds found at the surface; as a glance at Fig. 5.5 reveals. Independently of the XA spectra, another property of a molecule, the hydrogen bond asymmetry γ also changes with molecule orientation at the surface. The asymmetry describes the relative strength of the two

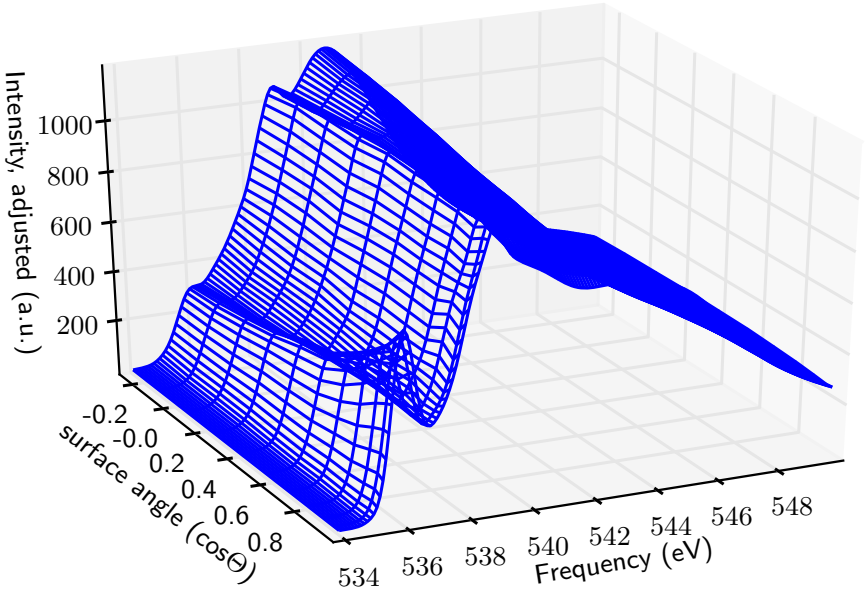


Figure 6.1: The intensity of pre-, mid- and post-edge peak depends on the orientation of the absorbing molecule. Spectra are normalized to the number of molecules, cosΘ describes a molecules smallest angle.

strongest donor or acceptor bonds, see Eq. 3.2 for the exact definition. Of the two possible asymmetry values, only γ_{Acc} is used here, because changes of the electron density at the hydrogen atoms have a greater effect on the XA spectrum [38]. The asymmetry is the highest for molecules with O-H aligned with the surface normal and the lowest at the opposite end. A graphical representation of the relation between asymmetry and surface angles is given in Fig 6.2. Molecules with O-H aligned to the surface normal ($\cos\Theta = 1$) point to the vacuum and therefore lack potential bonding partners, intuitively explaining their high asymmetry γ. For molecules where both O-H point away from the surface ($\cos\Theta \approx -0.4$), γ is low, simply by virtue of the molecules being most able to form two bonds. While the molecules where both surface angles are virtually equal, appear to have high asymmetry, this feature can be attributed to a low molecule count at the fringes of the molecule distribution.

The asymmetry increases as $\cos\Theta$ of the surface angle gets larger – meaning smaller surface angles – and at large $\cos\Theta$, an intense pre-edge peak is expected. Therefore, it is well possible that the orientation dependence of the XA signal is a result of the underlying changes in the asymmetry γ of the molecules. Consequently, instead of continuing the analysis of the XA spectrum's orientation dependence, the spectrum's dependence on γ is now investigated. Using asymmetry to describe the

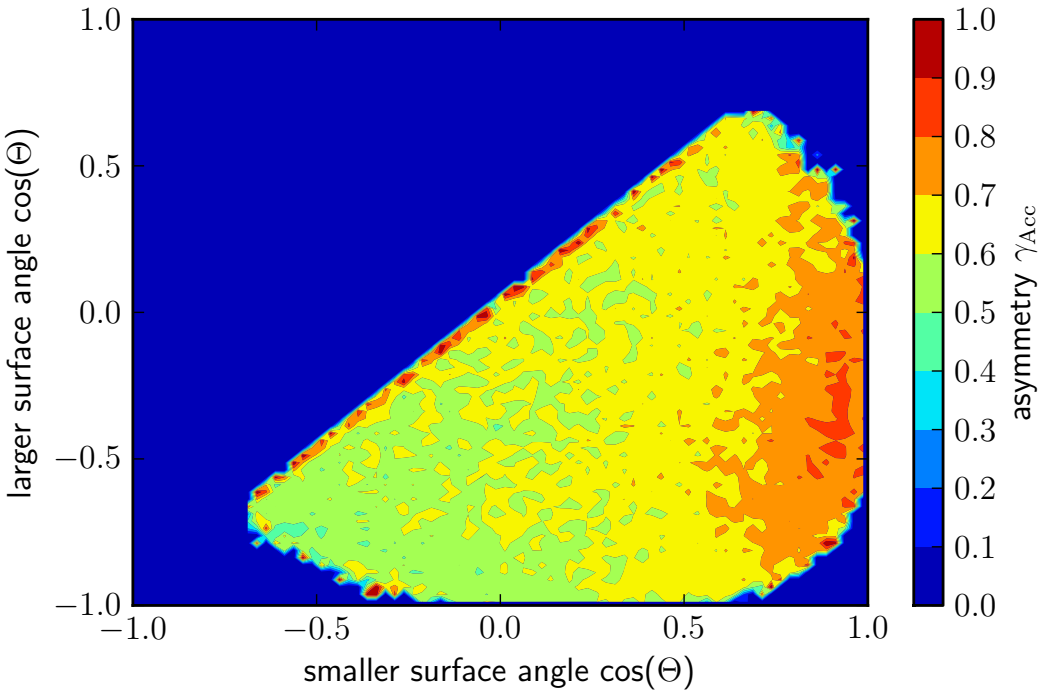


Figure 6.2: The asymmetry γ_{Acc} of molecules coincides with their alignment to the surface vector.

spectra has the advantage that, contrary to the surface orientation, the asymmetry is a relevant descriptor in the bulk as well. The surface molecules are divided according to their asymmetry from most symmetric ($\gamma_{\mathrm{Acc}} = 0$) to most asymmetric ($\gamma_{\mathrm{Acc}} = 1$). Fig. 6.3 depicts the XA spectra for each group. For the most asymmetric molecules, we find, in accordance with Fig. 6.1 and Fig. 6.2 a dominant pre-edge peak at 536 eV. The spectrum of the most symmetric molecules, is dominated by post-edge peak at 542 eV. In addition to these features, the prominence of the mid-edge peak at 539 eV appears to also vary with the asymmetry. Indeed, Fig 6.3 suggests that the post edge peak gains intensity at the cost of the mid-edge peak and the signal at 542 eV. The first five graphs, which are in the range of $\gamma_{\mathrm{Acc}} = 0$ to 0.83 show a continuous decrease of the post-edge signal, as well as slight increases of the mid-edge and pre-edge signals, as well as heightened intensity beyond 541 eV. With increasing asymmetry, a dip between the mid-edge and the post-edge signal is formed. In addition, an increasing signal intensity beyond 542 eV is observed.

In going from the group with $\gamma_{\mathrm{Acc}} \leq 0.8\bar{3}$ to the group with the highest asymmetry, drastic changes in the spectrum are observed. The intensity of mid- and post-edge is reduced, a pre-edge peak emerges and intensity beyond 542 eV is greatly increased. However, despite the overall reduced intensity, the mid-edge peak con-

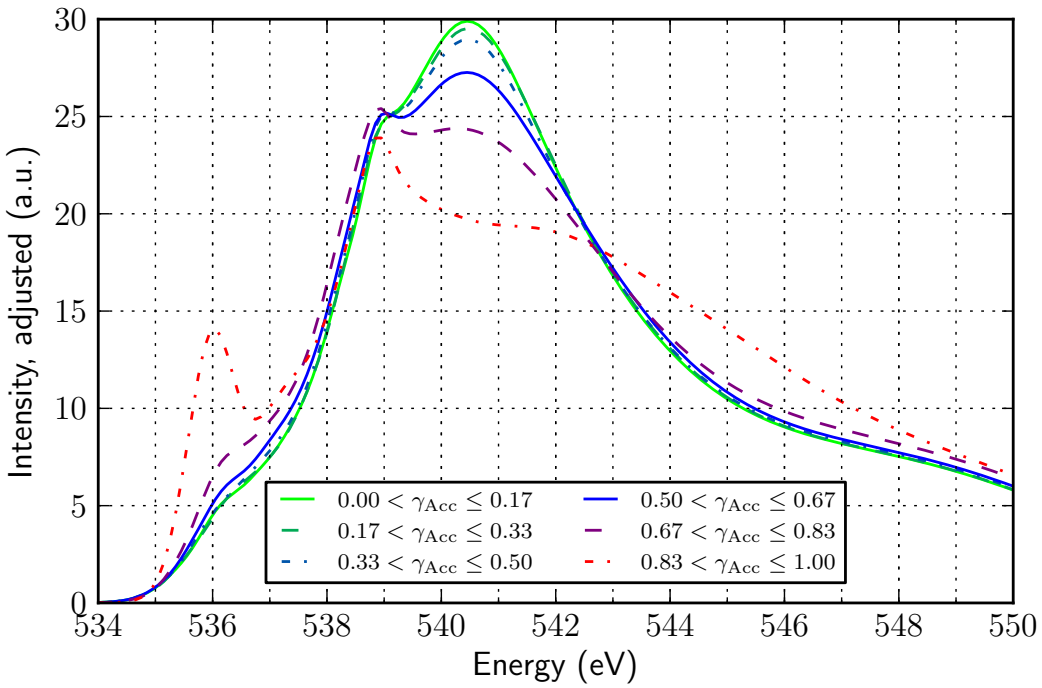


Figure 6.3: Devolution of the XA-Spectra of the 1st and 2nd layer with γ_{Acc}. Increasing γ causes the mid-egde peak to rise at the cost of the post-edge peak.

tinues to increase relative to the post-edge intensity. The spectrum of the most asymmetric molecules is thus considered to follow the same pattern as all the other spectra. The relative intensity changes of mid- and post-edge are interpreted as the redistribution of intensity from the signal at 541 eV the signals at 539 eV and 542 eV. Where the intensity of the pre-edge peaks goes in case of symmetric molecules is not entirely clear, it is possibly buried beneath the main and post-edge signal. The signals in XAS stem the promotion of a 1s electron into a virtual orbital, and the intensity redistribution is interpreted as two distinct virtual orbitals, whose energy difference is mirrored by the electronic asymmetry: For high asymmetry values two separate maxima are found, reflecting a strongly and a weakly interacting bond. Invoking LCAO[1] splitting, the virtual orbital corresponding to the strong hydrogen bond, has an incredibly high energy, whereas the virtual orbital to the weakly interacting bond is considerably lower. When the asymmetry is reduced, the two states move together. This reduces the intensity beyond 542 eV and the intensity of the mid-edge peak, until only the singular post-edge peak is produced by two bonds of similar strength.

[1] Linear Combination of Atomic Orbitals, see Sec. 2.1.2

γ_{Acc}	$0.1\bar{6}$	$0.\bar{3}$	0.5	$0.\bar{6}$	$0.8\bar{3}$	1.0
$E^{1\mathrm{st}}_{A\leftarrow D}$	17.1(2)	18.9(2)	20.6(2)	22.6(3)	22.0(3)	20.1(1)
$E^{2\mathrm{nd}}_{A\leftarrow D}$	15.6(2)	14.1(1)	12.0(1)	9.4(1)	5.53(9)	0.72(1)
$\sum_{i=1}^{2} E^{i-\mathrm{th}}_{A\leftarrow D}$	32.7(3)	33.0(3)	32.7(3)	32.0(3)	27.5(3)	20.8(1)

Table 6.1: Changes of the strongest ($E^{1\mathrm{st}}_{A\rightarrow D}$) and second strongest ($E^{2\mathrm{nd}}_{A\leftarrow D}$) bond at the surface. All energies are given in $\frac{\mathrm{kJ}}{\mathrm{mol}}$.

At variance to our interpretation based on asymmetry, the XA-signal has been linked to the energy of hydrogen bonds in previous publications [169, 29]. The continuous change of the water XA spectrum with asymmetry could be the result of continuous change of the underlying bond energies. Indeed, Tab. 6.1 shows that both bond energies, as well as their sum devolve in a largely monotonous fashion. However, the fact remains that molecules with a strong post-edge feature and relatively weak bonds exist; this has been observed for downward facing bonds in the discussion of Fig. 6.1. Furthermore, the sum of bond energies in Tab. 6.1 is equal for two different values of the asymmetry. The XA spectra, follow the asymmetry uniformly, and they show different intensities for equal sums of bond energies. Therefore, it can be stated with confidence that the asymmetry is a better descriptor of surface XA spectra, than bond energy.

The propsed dependence between XA signals and the γ is qualitatively applied to other spectra, by comparing the height of mid- and post-edge signals of ice and bulk water. In ice $\mathrm{I_h}$, the spectrum features a prominent post-edge and a semi-merged mid edge peak [29, 169]. Hydrogen bonds in ice $\mathrm{I_h}$ are far more symmetric than those in water, where the average acceptor asymmetry $\gamma_{\mathrm{Acc}} \approx 0.34$ and it is molecules with asymmetry values $> 0.\bar{6}$ are extremely rare [46]. The XA spectrum of bulk water exhibits a greater intensity for the post-edge signal than for the mid-edge signal. The average asymmetry $\gamma_{\mathrm{Acc}} = 0.50$ is also lower than the average asymmetry of the 2$^{\mathrm{nd}}$ surface layer $\gamma_{\mathrm{Acc}} = 0.61$. At the surface, the intensity of mid-edge signal exceeds the intensity of the post-edge signal. Despite the preceeding considerations being cursory to some extent the following trend can be seen. The relative intensities of mid- and post edge peaks follows the magnitude of the average asymmetry γ_{Acc} in all three environments. This implies that indeed a low asymmetry value causes a strong symmetric peak, whereas a high asymmetry value causes a strong mid-edge peak.

The surface of water not only has a wide array of hydrogen bonds, but also of asymmetries. Especially high asymmetry species are prevalent here, while they are harder to come by in bulk. This wide array of hydrogen bond energies and asymmetries allows to distinguish the possible causes for the presence of a peak in an XA spectrum. The new explanation of the XA signals interprets the relative intensities of mid-edge and post-edge signals as the result of two virtual states whose energy difference increases with asymmetry. It is argued that γ_{Acc} offers a more consistent description of water's XAS peaks that does not only work at the surface but appears to be applicable to bulk water and ice as well.

7 Conclusions

In this work, we have investigated hydrogen bond energies. The study encompassed their interrelation in bulk and at the surface as well as the question how hydrogen bonds relate to X-Ray Absorption.

The first major conclusion that can be drawn is that this work presents for the first time a quantitative description of the hydrogen bond cooperation in water. We found that a bond's energy linearly depends on the amount of energy received by and given to other bonds. We have further shown that the origin of cooperative behavior lies in the need of water molecules to reduce the charge imbalance that is introduced by the formation of hydrogen bonds. An additional investigation of the surface revealed that cooperative behavior is unchanged by the distortions caused to hydrogen bond network in close proximity to the surface. We further found that the distortion of the hydrogen bond network does not affect o_E much, an effect that could be traced to the occurrence of pairs without anti-cooperative bonds. It thus appears that at the surface less charge has to be neutralized, consequently there are fewer bonds.

With this knowledge, the many observations attributed to cooperation can now be backed by relative bond strengths. For instance, the relationship between the number of donor and acceptor bonds is now a direct result of the energetic gain a bond receives in the presence of another hydrogen bond. Previous approaches to cooperation that relied on hydrogen bonds were not easily applied to the surface due to disturbed hydrogen bond network. However, as the physics behind cooperation can be expected to be equal, our model based on the bond energies of pairs can be anticipated to provide compatible results at the surface. There the method provides an avenue for the investigation of the effect of cooperation on surface properties. The idea of energy balancing futhermore raises questions about the time correlation of the energy of cooperative bonds. This may also shed more light on the role of water wires, which were identified by Hassanali et al. [170]. We have further demonstrated that the orientational dependence of hydrogen bond energies to be a cooperative effect driven is by the fact that the first layer can donate more bonds into the second layer than it can accept. The orientational dependence of the hydrogen bond energy entails that molecules of different orientation absorb at different frequencies when

probed with SFG spectra. Thus, with a map of hydrogen bond strengths, and the known connection between frequency and energy [35, 36], it must be possible to identify the number and location of all group that give rise to the SFG spectrum of the water surface. Additionally, as hydrogen bond formation is essential to waters reorientation time [171, 172], the noticeable difference of bond strengths for various orientations implies different correlation lengths with respect to the starting orientation.

Our work shows that three signals of the water XA signals consistently follow the change of the asymmetry parameter. Contrary to interpretation using hydrogen bond strength, the asymmetry parameter can correctly predict the relative heights of the mid-edge and post-edge feature, whether it is bulk water, water surface or ice. At variance to the usage of bond numbers, γ_{Acc} provides a continuous measure on which the XA result depends. Moreover, as this dependence includes also the disputed mid-edge peak, γ_{Acc} might offer the avenue to a comprehensive interpretation of the water XA spectrum. As the current excitation method is semi-quantitative, future work should include all slabs of water and ice, as well as more sophiticated methods to model the exitation, such as time-dependent density functional theory or approximations based on Green's functions [29]. Furthermore, an explicit investigation of the orbitals involved in the absorption process can verify the role virtual orbitals in the dependency of XA spectra on the asymmetry. From our findings a unified interpretation of water XA spectrum then becomes available, that enables researchers to link XA spectra of certain asymmetry groups to other water properties.

Bibliography

[1] Nico F. A. van der Vegt, Kristoffer Haldrup, Sylvie Roke, Junrong Zheng, Mikael Lund, and Huib J. Bakker. Water-Mediated Ion Pairing: Occurrence and Relevance. *Chem. Rev.*, **116**:7626 – 7641, 2016.

[2] Katrin Amann-Winkel, Marie-Claire Bellissent-Funel, Livia E. Bove, Thomas Loerting, Anders Nilsson, Alessandro Paciaroni, Daniel Schlesinger, and Lawrie Skinner. X-ray and Neutron Scattering of Water. *Chem. Rev.*, **116**:7570 – 7570, 2016.

[3] Frank N. Keuthsch, Jeffery D. Cruzan, and Richard J. Saykally. The Water Trimer. *Chem. Rev.*, **103**:2533 – 2577, 2003.

[4] Richard N. Butler and Anthony G. coyne. Water: Nature's Reaction Enforcer – Comparative Effects for Organic Synthesis "In-Water" and "On-Water". *Chem. Rev.*, **110**:6302, 2010.

[5] Philip Ball. Water as an Active Constituent in Cell Biology. *Chem. Rev.*, **108**(1):74, 2008.

[6] Noam Agmon, Huib J. Bakker, R. Kramer Campen, Richard H. Henchman, Peter Pohl, Sylvie Roke, Martin Thämer, and Ali Hassanali. Protons and Hydroxide Ions in Aqueous Systems. *Chem. Rev.*, **116**:7642 – 7672, 2016.

[7] Olle Björneholm, Martin H. Hansen, Andrew Hodgson, Li-Min Liu, David T. Limmer, Angelos Michaelides, Philipp Pedevilla, Jan Rossmeisl, Huaze Shen, Gabriele Tocci, Eric Tyrode, Marie-Madeleine Walz, Josephina Werner, and Hendrik Bluhm. Water at Interfaces. *Chem. Rev.*, **116**:7698 – 7726, 2016.

[8] Imre Bakó and István Mayer. Hierarchy of the Collective Effects in Water Clusters. *J. Phys. Chem. A*, **120**:631 – 638, 2016.

[9] Henry S. Frank and Wen-Yang Wen. Ion-Solvent Interaction. *Disc. Farad. Soc.*, **24**:133, 1957.

[10] D. Hankins and J. W. Moskowitz. Water Molecule Interactions. *J. Chem. Phys.*, **53**:4454, 1970.

[11] Frank H. Stillinger. Water Revisited. *Science*, **209**:451 – 457, 1980.

[12] David P. Shelton. Slow polarization relaxation in water observed by hyper-rayleigh scattering. *Phys. Rev. B*, **72**(2):020201, 2005.

[13] Frank N. Keuthsch and Richard J. Saykally. Water clusters: Untangling the mysteries of the liquid, one molecule at a time. *Proc. Natl. Acad. Sci*, **98**(19):10533 – 10540, 2011.

[14] Keiichi Ohno, Mari Okimura, Nobuyuki Akai, and Yukiteru Katsumoto. The effect of cooperative hydrogen bonding on the oh stretching-band shift for water clusters studied by matrix-isolation infrared spectroscopy and density functional theory. *Phys. Chem. Chem. Phys.*, **7**:3005 – 3014, 2005.

[15] Kevin Stokely, Marco G. Mazza, H. Eugene Stanley, and Giancarlo Franzese. Effect of Hydrogen Bond Cooperativity on the Behaviour of Water. *Proc. Natl. Acad. Sci.*, **107**:1301, 2010.

[16] Pradeep Kumar, Giancarlo Franzese, Sergey V. Buldyrev, and H. Eugene Stanley. Molecular dynamics study of orientational cooperativity in water. *Phys. Rev. E*, **73**(4):041505, 2006.

[17] Michel Masella and Jean Pierre Flament. A pairwise and two many-body models for water: Influence of nonpairwise effects upon the stability and geometry of (h2o)n cyclic (n=3–6) and cagelike (n=6–20) clusters. *J. Chem. Phys.*, **107**(21):9105 – 9116, 1997.

[18] Ping Qian, Wei Song, Linan Lu, and Zhongzhi Yang. Ab initio investigation of water clusters (h2o)n (n = 2–34). *Intl. J. of Quant. Chem.*, **110**(10):1923 – 1937, 2010.

[19] Sotiris S. Xantheas and Thom H. Dunning Jr. Ab initio studies of cyclic water clusters (H2O)n, n=1–6. i. optimal structures and vibrational spectra. *J. Chem. Phys.*, **99**(11):8774 – 8792, 1993.

[20] Suehiro Iwata, Pradipta Bandyopadhyay, and Sotiris S. Xantheas. Cooperative Roles of Charge Transfer and Dispersion Terms in Hydrogen-Bonded Networks of $(h_2o)_n$, n=6,11, and 16. *J. Phys. Chem. A*, **117**:6641 – 6651, 2013.

[21] José Manuel Guevara-Vela, Rodrigo Chávez-Calvillo, Marco García-Revilla, Jesús Hernández-Trujillo, Ove Christiansen, Evelio Francisco, Ángel Martín Pendás, and Tomás Rocha-Rinza. Hydrogen-Bond Cooperative Effects in

Small Cyclic Water Clusters as Revealed by the Interacting Quantum Atoms Approach. *Chem Eur. J.*, **19**:14304 – 14315, 2013.

[22] Pengyu Ren and Jay W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem B*, **107**(24):5933 – 5947, 2003.

[23] Noam Agmon. Liquid Water: From Symmetry Distortions to Diffusive Motion. *Acc. Chem. Res.*, **45**:63–73, 2012.

[24] Omer Markovitch and Noam Agmon. The distribution of acceptor and donor hydrogen-bonds in bulk liquid water. *Mol. Phys.*, **106**(2 – 4):485 – 495, 2008.

[25] Sridhar Narayan, John Muldoon, M. G. Ginn, Valery V. Fokin, Hartmuth C. Kolb, and K. Barry Sharpless. "On Water": Unique Reactivity of Organic Compounds in Aqueous Suspension. *Angew. Chem. Int. Ed.*, **44**:3275, 2005.

[26] Kristof Karhan, Rustam Z. Khaliullin, and Thomas D. Kühne. On the Role of Interfacial Hydrogen Bonds in "On-Water" Catalysis. *J. Chem. Phys.*, **141**(22):22D528, 2014.

[27] Yuki Nagata and Mischa Bonn. Untersuchung der Struktur und Dynamik von Wasser an der Wasser-Luft-Grenzfläche mittels oberflächenspezifischer Schwingungsspektroskopie. *Angew. Chem.*, **127**:5652 – 5669, 2015.

[28] Akihiro Morita, Takako Imamura, and Tatsuja Ishiyama. Theoretical Studies of Structures and Vibrational Sum Frequency Generation Spectra at Aqueous Interfaces. *Chem. Rev.*, **114**:8447 – 8470, 2014.

[29] Thomas Fransson, Yoshihisa Harada, Nobuhiro Kosugi, Nicholas A. Beley, Bernd Winter, John J. Rehr, Lars G. M. Pettersson, and Anders Nilsson. X-ray and Electron Spectroscopy of Water. *Chem. Rev.*, **116**:7551 – 7569, 2016.

[30] Quan Du, Eric Freysz, and Ron Shen. Surface Vibrational Spectroscopic Studies of Hydrogen Bonding and Hydrophobicity. *Science*, **264**:826 – 828, 1994.

[31] Satoshi Nihonyanagi, Tatsuya Ishiyama, Touk-kwan Lee, Shoichi Yamaguchi, Mischa Bonn, Akihiro Morita, and Tahei Tahara. Unified Molecular View of the Air/Water Interface Based on Experimental and Theoretical $\chi^{(2)}$ Spectra of an Isotopically Diluted Water Surface. *J. Am. Chem. Soc.*, **133**:16875 –16880, 2011.

[32] Maria Sovago, Campen R. Kramer, George W. H. Wurpel, Michiel Müller, Bakker Huib J., and Mischa Bonn. Vibrational Response of Hydrogen-Bonded

Interfacial Water is Dominated by Intramolecular Coupling. *Phys. Rev. Lett.*, **100**:173901, 2008.

[33] Maria Sovago, Campen R. Kramer, George W. H. Wurpel, Michiel Müller, Huib J. Bakker, and Mischa Bonn. Sovago *et al.* Reply:. *Phys. Rev. Lett.*, **101**:139402, 2008.

[34] Jan Schäfer, Ellen H. G. Backus, Yuki Nagata, and Mischa Bonn. Both inter- and intramolecular coupling of o–h groups determine the vibrational response of the water/air interface. *J. Phys. Chem. Lett.*, **7**(22):4591 – 4595, 2016.

[35] Rossend Rey, Klaus B. Moller, and James T. Hynes. Hydrogen Bond Dynamics in Water and Ultrafast Infrared Spectroscopy. *J. Phys. Chem. A*, **106**(50):11993 – 11996, 2002.

[36] Hosssam Elgabarty, Rustam Z. Khaliullin, and Thomas D. Kühne. Covalency of hydrogen bonds in liquid water can be probed by proton nuclear magnetic resonance. *Nat. Commun.*, page 8318, 2015.

[37] Lingzhu Kong, Xifan Wu, and Roberto Car. Roles of quantum nuclei and inhomogeneous screening in the x-ray absorption spectra of water and ice. *Phys. Rev. B*, **86**:134203 – 5, 2012.

[38] R. L. C. Wang, Kreuzer H. J., and Grunze M. Theoretical modeling and interpretation of X-ray absorption spectra of liquid water. *Phys. Chem. Chem. Phys.*, **8**:4744 –4751, 2006.

[39] A. Nilsson and L. G. M. Pettersson. Perspective on the structure of liquid water. *Chem. Phys.*, **389**:1 – 34, 2011.

[40] M. Cavlleri, H. Ogasawara, L. G. M. Pettersson, and A. Nilsson. The interpretation of X-ray absorption spectra of water and ice. *Chem. Phys. Lett.*, **364**:363 – 370, 2002.

[41] Marcella Iannuzzi and Jörg Hutter. Inner-shell spectroscopy by the Gaussian and augmented plane wave method. *Phys. Chem. Chem. Phys.*, **9**:1599 – 1610, 2007.

[42] Rustam Z. Khaliullin, Joost VandeVondele, and Jürg Hutter. Efficient Linear-Scaling Density Functional Theory for Molecular Systems. *J. Chem. Theor. Commun.*, **9**:4421 – 4427, 2013.

[43] Rustam Z. Khaliullin and Thomas D. Kühne. Microscopic Properties of Liquid Water From Combinded *Ab-initio* Molecular Dynamics and Energy Decomposition Studies. *Phys. Chem. Chem. Phys.*, **15**:15746, 2013.

[44] Rustam Z. Khaliullin, Alexis T. Bell, and Martin Head-Gordon. Analysis of Charge Transfer Effects in Molecular Complexes Based On Absolutely Localized Molecular Orbitals. *J. Chem. Phys.*, **128**:184112, 2008.

[45] Thomas D. Kühne and Rustam Z. Khaliullin. Electronic Signature of the Instantaneous Asymmetry in the First Coordiantion Shell of Liquid Water. *Nat. Comm.*, **4**:1450, 2013.

[46] Thomas D. Kühne and Rustam Z. Khaliullin. Nature of the Asymmetry in the Hydrogen-Bond Networks of Hexagonal Ice and Liquid Water. *J. Am. Chem. Soc*, **136**:3395, 2014.

[47] W. Kohn. Nobel Lecture: Electronic structure of matter – wave functions and density functionals. *Rev. Mod. Phys.*, 1999.

[48] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry*. Dover Publications Inc., 1. edition, 1989.

[49] Michael D. Fayer. *Elements of quantum mechanics*. Oxford University Press, 1. edition, 2000.

[50] David S. Sholl and Janice A Steckel. *Density Functional Theory: A practical introduction*. John Wiley and Sons, 2009.

[51] Arnold Frederik Hollemann and Nils Wieberg. *Anorganische Chemie*, chapter X, pages 347 – 371. Walter de Gruyter, 102 edition, 2007.

[52] Roald Hoffmann. Die Begegnung von Chemie und Physik im Festkörper. *Angew. Chem.*, **99**:871 – 906, 1987.

[53] W. Kohn and P. Hohenberg. Inhomogeneous Electron Gas. *Phys. Rev.*, **136**(3B):B864 – B871, 1964.

[54] Kurt Lejaeghre et *al.* Reproducibility in density functional theory calculations of solids. *Science*, **351**(6280):aad3000 – 7, 2016.

[55] Robert G. Parr and Weitao Yang. *Density-Functional Theory of Atoms and Molecules*, volume 16 of *International series of monographs in chemistry*. Oxford Science Publications, 1. edition, 1989.

[56] Eberhard Engel and Reiner M. Dreizler. *Density Functional Theory An Advanced Course*. Theoretical and mathematical Physics. Springer, 2011.

[57] Mel Levy. Electron densities in search of Hamiltonians. *Phys. Rev. A*, **26**(3):1200 – 1209, 1982.

[58] T. L. Gilbert. Hohenberg-Kohn theorem for nonlocal external potentials. *Phys. Rev. B*, **12**(6):2111 – 2120, 1975.

[59] Lucas O. Wagner and Kieron Burke. DFT in a nutshell. *Int. J. Quant. Chem.*, **113**:96 – 101, 2013.

[60] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, **140**(4A):A1133 – A1138, 1965.

[61] Klaus Capelle. A bird's-eye view of density-functional theory, 2006.

[62] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, **77**:3865 – 3868, Oct 1996.

[63] A. D. Becke. Density-Functional Exchange-Energy Apphysrevoximation with Correct Asymptotic Behavior. *Phys. Rev. A*, **38**:3098, 1988.

[64] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B*, **37**:785, 1988.

[65] Dominik Marx and Jakob Hutter. *Ab Initio Molecular Dynamics*. Cambridge University Press, 1. edition, 2009.

[66] Joost VandeVondele and Jürg Hutter. An Efficient Orbital Transformation Method for Electronic Structure Calculations. *J. Chem. Phys.*, **118**:4365, 2003.

[67] Aron J. Cohen, Paula Mori-Sánches, and Weitao Yang. Insights into Current Limitations of Density Functional Theory. *Science*, **321**:792 – 794, 2008.

[68] Stefan Grimme. Density functional theory with London dispersion corrections. *Comp. Mol. Sci.*, **1**:211 – 228, 2011.

[69] Jue Wang, G. Román-Pérez, Jose M. Soler, Emilio Artacho, and M.-V. Fernández-Serra. Density, structure, and dynamics of water: The effect of van der Waals interactions. *TJ. Chem. Phys.*, **134**(2):024516, 2011.

[70] Weitao Yang, Yingkai Zhang, and Paul W. Ayers. Degenerate Ground States and a Fractional Number of Electrons in Density and Reduced Density Matrix Functional Theory. *Phys. Rev. Lett.*, **84**(22):5172 – 5175, 2000.

[71] Paula Mori-Sánchez, Aron J. Cohen, and Weitao Yang. Localization and Delocalization Errors in Density Functional Theory and Implications for Band-Gap Prediction. *Phys. Rev. Lett.*, **100**(14):146401, 2008.

[72] Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang. Fractional Spins and Static Correlation Error in Density Functional Theory, 2008.

[73] Matteo Cococcioni. A LDA+U Study of Selected Iron Compounds, 2002.

[74] A. D. et al. MacKerell Jr. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, **102**:3586 – 3616, 1998.

[75] Jörg Behler. Constructing high-dimensional neural network potentials: A tutorial review. *Intl. J. Quant. Chem.*, **115**(16):1032 – 1050, 2015.

[76] Hans Martin Senn and Walter Thiel. QM/MM-Methoden für biomolekulare Systeme. *Angew. Chem.*, **121**(7):1220 – 1254, 2009.

[77] Thom Vreven, Suzie K. Byun, István Komáomi, Dapprich Stefan, John A. Montgomery, Keiji Morokuma, and Michael J. Frisch. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theo. Comput.*, **2**:815 – 826, 2006.

[78] Adri C. T. van Duin, Siddharth Dasgupta, Francois Lorant, and William A. Goddard. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A*, **105**(41):9396 – 9409, 2001.

[79] Victor Guallar, Andrzej A. Jarzecki, Richard A. Friesner, and Thomas G. Spiro. Modeling of Ligation-Induced Helix/Loop Displacements in Myoglobin: Toward an Understanding of Hemoglobin Allostery. *J. Am. Chem. Soc.*, **128**(16):5427 – 5435, 2006. PMID: 16620114.

[80] R. E. Alcantara, C. Xu, T. G. Spiro, and V. Guallar. A quantum-chemical picture of hemoglobin affinity. *Proc. Natl. Acad. Sci.*, **104**(47):18451 – 18455, 2007.

[81] Yue Yuan, Ming F. Tam, Virgil Simplaceanu, and Chien Ho. New Look at Hemoglobin Allostery. *Chemical Reviews*, **115**(4):1702 – 1724, 2015. PMID: 25607981.

[82] Teodoro Laino, Fawzi Mohamed, Alessandro Laio, and Michele Parrinello. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *J. Chem. Theo. Comput.*, **1**(6):1176 – 1184, 2005.

[83] Martin J. Field, Paul A. Bash, and Martin Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, **11**(6):700– – 733, 1990.

[84] The Nobel Prize in Chemistry 2013, 2013.

[85] Feliu Maseras and Keiji Morokuma. IMMOM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.*, **16**(9):1170 – 1179, 1995.

[86] http://www.cp2k.org.

[87] Jürg Hutter, Marcella Iannuzzi, Florian Schiffmann, and Joost VandeVondele. cp2k: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, **4**(1):15–25, 2014.

[88] Alessandro Laio, Joost VandeVondele, and Ursula Rothlisberger. A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.*, **116**(16):6941 – 6947, 2002.

[89] Teodoro Laino, Fawzi Mohamed, Alessandro Laio, and Michele Parrinello. An Efficient Linear-Scaling Electrostatic Coupling for Treating Periodic Boundary Conditions in QM/MM Simulations. *J. Chem. Theo. Comput.*, **2**(5):1370 – 1378, 2006.

[90] P. E. Blöchl. Electrostatic decoupoling of periodic images of plane-wave expanded densities and derived atomic point charges. *J. Chem. Phys.*, **103**(17):7422 – 7428, 1995.

[91] Tom K. Woo, Luigi Cavallo, and Tom Ziegler. Implementation of the IMOMM methodology for performing combined QM/MM molecular dynamics simulations and frequency calculations. *Theor. Chem. Acc.*, 1998.

[92] Feliu Maseras and Keiji Morokuma. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.*, **16**(9):1170 – 1179, 1995.

[93] O. Anatole von Lilienfeld, Ivano Tavernelli, Ursula Rothlisberger, and Daniel Sebastiani. Variational optimization of effective atom centered potentials for molecular properties. *J. Chem. Phys.*, **122**(1):014113, 2005.

[94] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J. Field. A Generalized Hybrid Obital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A*, **102**(24):4714 – 4721, 1998.

[95] Xavier Assfeld and Jean-Louis Rivail. Quantum chemical computations on parts of large molecules: the ab initio local self consistent field method. *Chem. Phys. Lett.*, **263**(1):100 – 106, 1996.

[96] Stephen R. Ellliot. *The Physics and Chemistry of Solids*. John Wiley and Sons, 2000.

[97] Dennis C. Rappaport. *The Art of Molecular Dynamics*. Cambridge University Press, 2004.

[98] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation*, volume 1 of *Computational Science Series*. Academic Press, 2. edition, 2002.

[99] J. M. Haile. *Molecular Dynamics Simulation; Elementary Methods*. John Wiley & Sons, Inc., 1. edition, 1997.

[100] Ronald E. Miller and Ellad B Tadmor. *Modeling Materials – Continuum, Atomistic, and Multiscale Techniques*. Cambridge University Press, 1. edition, 2011.

[101] Thomas D. Kühne. *Ab-Initio* Molecular Dynamics. *arXiv.1201.5945v2*, 2013.

[102] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.*, **97**:1990–2001, 1992.

[103] Dorothee Richters, Michele Ceriotti, and Thomas D. Kühne. A Field-Theoretic Aproach To Linear Scaling *Ab-Initio* Molecular Dynamics. *in PhysRevepara-tion*, 2013.

[104] Søren Toxvaerd. Hamiltonians for Discrete Dynamics. *Phys. Rev. E*, **50**:2271, 1994.

[105] John M. Herbert and Martin Head-Gordon. Accelerated, energy-conserving Born-Oppenheimer molecular dynamics via Fock matrix extrapolation. *Phys. Chem. Chem. Phys.*, **7**:3269 – 3275, 2005.

[106] R. Car and M. Parrinello. Unified Apphysrevoach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, **55**:2471, 1985.

[107] Thomas D. Kühne. Second Generation Car–Parrinello Molecular Dynamics. *Comp. Mol. Sci.*, **4**:391, 2014.

[108] T. D. Kühne, M. Krack, F. R. Mohamed, and M. Parrinello. Efficient and Accurate Car-Parrinello-like Apphysrevoach to Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett.*, **98**:066401, 2007.

[109] Jiří Kolafa. Gear Formalism of the Always Stable Predictor-corrector Molecular Dynamics of Polarizable Molecules. *J. Chem. Phys.*, **122**:164105, 2005.

[110] Adam H. R. Palser and David E. Manolopoulos. Canonical Purification of the Density Matrix in Electronic-Structure Theory. *Phys. Rev. B*, **58**:12704, 1998.

[111] Neil W. Ashcroft and N. David Mermin. *Festkörperphysik*. Oldenbourg, 1. edition, 2001.

[112] Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A Progam for Macromolecular Energy, Minimization, and Dynaics Calculations. *J. Comput. Chem*, **4**:187 – 217, 1983.

[113] Alexander D. Mackerell Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem*, **25**:1584 – 1604, 2004.

[114] Peter J. Steinbach and Bernard R. Brooks. New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *J. Comput. Chem.*, **15**(7):667 – 683, 1994.

[115] Hark Lee and Wei Cai. Ewald summation for coulomb interactions in a periodic supercell, 2009.

[116] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys*, **64**:253, 1921.

[117] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys*, **126**:014101, 2007.

[118] William L. Jorgensen. Convergence of monte carlo simulations of liquid water in the npt ensemble. *Chem. Phys. Lett.*, 92(4):405 – 410, 1982.

[119] Jan Kessler, Thomas Spura, Kristof Karhan, Pouya Partovi-Azar, Ali A. Hassanali, and Thomas D. Kühne. Structure and Dynamics of the Instantaneous Water/Wapor Interface Revisited by Path-Integral and Ab-Initio Molecular DynamicsSimulations. *J. Phys. Chem. B*, **119**:10079, 2015.

[120] Dominik Marx, Mark E. Tuckerman, Jürg Hutter, and Michele Parrinello. The nature of the hydrated excess proton in water. *Nat.*, **397**(6720):601 – 604, 1999.

[121] Mark E. Tuckerman, Dominik Marx, Michael L. Klein, and Michele Parrinello. On the Quantum Nature of the Shared Proton in Hydrogen Bonds. *Science*, **275**(5301):817 – 820, 1997.

[122] Richard Phillips Feynman, Albert R. Hibbs, and Daniel F. Styer. *Quantum Mechanics and Path Integrals Emended Edition*. Dover Publications, 2005.

[123] Scott Habershon, Manolopoulos, Thomas E. Markland, and Thomas F. Miller III. Ring-Polymer Molecular Dynamics: Quantum Effects in Chemical Dynamics from Classical Trajectories in an Extended Phase space. *Annu. Rev. Phys. Chem.*, **64**:387 – 413, 2013.

[124] Alexander Witt, Sergei D. Ivanov, Motoyuki Shiga, Harald Forbert, and Dominik Marx. On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy. *J. Chem. Phys.*, **130**:194510, 2009.

[125] Michele Ceriotti, Giovanni Bussi, and Michele Parrinello. Nclear Quantum Effects in Solids Using a Colored-Noise Thermostat. *Phys. Rev. Lett.*, **103**:030603, 2009.

[126] F.R. Krajewski and M. Parrinello. Linear Scaling Electronic Structure Calculations and Accurate Statisitcal Mechanics Sampling with Noisy Forces. *Phys. Rev. B*, **73**:041105, 2006.

[127] Paul Langevin, Don S. Lemosn, and Anthony Gythiel. Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien", C. R. Acad. Sci (Paris) 146, 530 – 533 (1908)]. *Am. J. Phys.*, **65**(11):1079 – 1081, 1997.

[128] Jiayu Dai and Jianmin Yuan. Large-scale Efficient Langevin Dynamics, and Why It Works. *Europhys. Lett.*, **88**:20001, 2009.

[129] Michele Ceriotti, Giovanni Bussi, and Michele Parrinello. Colored-Noise Thermostats à la Carte. *J. Chem. Theo. Comput.*, **6**(4):1170 – 1180, 2010.

[130] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B*, **114**(32):10235 – 10253, 2010.

[131] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci.*, **99**(20):12562 – 12566, 2002.

[132] Shalini Awasthi, Kapil Venkat, and Nisanth N. Nair. Sampling Free Energy Surfaces as Slices by Combining Umbrella Sampling and Metadynamics. *J. Comput. Chem.*, **37**:1413 – 1424, 2016.

[133] Johannes Kästner. Umbrella Sampling. *Comp. Mol. Sci*, **1**:932, 2011.

[134] Michiel Sprik and Giovanni Ciccotti. Free Energy From Constrained Molecular Dynamics. *J. Chem. Phys*, **109**:7737, 1998.

[135] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.*, **100**:020603, Jan 2008.

[136] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, and P.A. Kollman. The wieghted histogram anlaysis method for free-energy calculations on biomolecules. I. the mehtod. *J. Comput. Chem.*, **13**:1011 – 1021, 1992.

[137] M. Souaille and B. Roux. Extension to the weighted histogram analysis method:Combining umbrella sampling with free energy calculations. *Comput. Phys. Comm.*, **135**, 2001.

[138] Johannes Kästner and Walter Thiel. Analysis of the statistical error in umbrella sampling simulations by umbrella integration. *J. Chem. Phys.*, **124**:234106, 2006.

[139] Mark S. Gordon, Fedorov Dmitir G., Spencer R. Pruitt, and Lyudmila V. Slipchenko. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.*, **112**:632 – 672, 2012.

[140] Rustam Z. Khaliullin, Erika A. Cobar, Rohini C. Lochan, Alexis T. Bell, and Martin Head-Gordon. Unravelling the Origin of Intermolecular Interactions Using Absolutely Localized Molecular Orbitals. *J. Phys. Chem. A*, **111**:8753, 2007.

[141] Rustam Z. Khaliullin, Martin Head-Gordon, and Alexis T. Bell. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.*, **124**:204105, 2006.

[142] Thomas Spura, Christopher John, Scott Habershon, and Thomas D. Kühne. Nuclear quantum effects in liquid water from path-integral simulations using an ab initio force-matching approach. *Mol. Phys.*, **113**(8):808–822, 2015.

[143] Joseph A. Morrone and Roberto Car. Nuclear quantum effects in water. *Phys. Rev. Lett.*, **101**:017801, 2008.

[144] Jochen Schmidt, Joost VandeVondele, I. F. William Kuo, Daniel Sebastiani, J. Ilja Siepmann, Jürg Hutter, and Christopher J. Mundy. Isobaric–iIsothermal Molecular Dynamics Simulations Utilizing Density Functional Theory: An Assessment of the Structure and Density of Water at Near-Ambient Conditions. *Jour. Phys. Chem. B*, **113**(35):11959 – 11964, 2009.

[145] T. D. Kühne, T. A. Pascal, E. Kaxiras, and Y. Jung. New Insights into the Structure of the Vapor/Water Interface from Large-Scale First-Principles Simulations. *J. Phys. Chem. Lett.*, **2**:105, 2011.

[146] Alenka Luzar and David Chandler. Hydrogen-bond Kinetics in Liquid Water. *Nature*, **379**:55, 1996.

[147] Adam P. Willard and David Chandler. Instantaneous Liquid Interfaces. *J. Phys. Chem. B.*, **114**:1954, 2010.

[148] Gerald Lippert, Jürg Hutter, and Parrinello Michele. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor. Chem. Acc.*, **103**:124 – 140, 1999.

[149] Matthias Krack and Michele Parrinello. All-electron ab-initio molecular dynamics. *Phys. Chem. Chem. Phys.*, **2**:2105 – 2112, 2000.

[150] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**(2):926 – 935, 1983.

[151] Werner Kutzelnigg, Ulrich Fleischer, and Michael Schindler. *The IGLO-Method: Ab-initio Calculation and Interpretation of NMR Chemical Shifts and Magnetic Susceptibilities*, pages 165 – 262. Springer Berlin Heidelberg, Berlin, Heidelberg, 1991.

[152] Nelson H. F. Beebe. ON THE TRANSITION STATE IN THE $X\alpha$ METHOD. *Chem. Phys. Lett.*, **19**:290 – 294, 2973.

[153] J. C. Slater and Johnson K. H. Self-Consistent-Field $x\alpha$ Cluster Method for Polyatomix Molecules and Solids. *Phys. Rev. B*, **5**:844 – 853, 1972.

[154] Marcella Iannuzzi. X-ray absorption spectra of hexagonal ice and liquid water by all-electron Gaussian and augmented palne wave calculations. *J. Chem. Phy.*, **128**, 2008.

[155] William H. Press, Saul A. Teukkolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipies in* FORTRAN*: The Art of Scientific Computing*. Cambridge University Press, 2. edition, 1992.

[156] Norbert Henze. *Stochastik für Einsteiger*. Vieweg + Teubner, 9. edition, 2012.

[157] Herman J. C. Berendsen. *Data and Error Analysis*. Cambridge University Press, 1. edition, 2011.

[158] J. D. Hunter. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.*, **9**(3):90 – 95, 2007.

[159] S. van der Walt, S. C. Colbert, and G. Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Comp. Sci. Eng.*, **13**(2):22 – 30, March 2011.

[160] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001 – 2013. [Online; accessed ¡today¿].

[161] D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Indust. Appl. Math.*, **11**:431, 1963.

[162] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, **14**:33 – 38, 1996.

[163] N. A. Benedek, I. K. Snook, M. D. Towler, and R. J. Needs. Quantum Monte Carlo caluclations of the dissociation energy of the water dimer. *J. Chem. Phys.*, **125**(10):104302, 2006.

[164] Martin W. Feyereisen, David Feller, and David A. Dixon. Hydrogen Bond Energy of the Water Dimer. *J. Phys. Chem.*, **100**(8):2993 – 2997, 1996.

[165] Rustam Z. Khaliullin, Alexis T. Bell, and Martin Head-Gordon. Electron Donation in the Water-Water Hydrogen Bond. *Chem. Eur. J.*, **15**:851–853, 2009.

[166] Mirijam Zobel, Reinhard B. Neder, and Simon A. J. Kimber. Universal solvent restructuring induced by colloidal nanoparticles. *Science*, **347**(6219):292 – 294, 2015.

[167] H. K. Christenson, D. W. R. Gruen, R. G. Horn, and J. N. Israelachvili. Structuring in liquid alkanes between solid surfaces: Force measurements and mean-field theory. *J. Chem. Phys.*, **87**(3):1834 – 1841, 1987.

[168] Yuobo Fan, Xin Chen, Lijaiang Yang, Paul S. Cremer, and Yi Qin Gao. On th Structure of Water at the Aqueous/Air Interface. *J. Phys. Chem. B*, **113**(34):11672 – 11679, 2009.

[169] Wei Chen, Wu Xifan, and Roberto Car. X-Ray Absorption Signatures of the Molecular Environment in Water and Ice. *Phys. Rev. Lett.*, **105**:017802 – 4, 2010.

[170] Ali Hassanali, Federico Giberti, Jérôme Cuny, and Thomas D. Kühne. Proton transfer through the water gossamer. *Proc. Natl. Acad. Sci.*, **110**(34):13723 – 13728, 2013.

[171] Meng Chen, Xiancai Lu, Xiandong Liu, Qingfeng Hou, Youyi Zhu, and Huiqun Zhou. Retardation of Water Reorientation at the Oil/Water Interface. *J. Phys. Chem. C*, **119**(29):16639 – 16648, 2015.

[172] Damien Laage and James T. Hynes. On the Molecular Mechanism of Water Reorientation. *J. Phys. Chem. B*, **112**(45):14230 – 14242, 2008.

Appendix

A.1 Publications

- Kristof Karhan, Rustam Z. Khaliullin, and Thomas D. Kühne. On the Role of Interfacial Hydrogen Bonds in “On-Water” Catalysis. *J. Chem. Phys.*, **141**(22):22D528, 2014.
- Jan Kessler, Thomas Spura, Kristof Karhan, Pouya Partovi-Azar, Ali A. Hassanali, and Thomas D. Kühne. Structure and Dynamics of the Instantaneous Water/Wapor Interface Revisited by Path-Integral and Ab-Initio Molecular Dynamics Simulations. *J. Phys. Chem. B*, **119**:10079, 2015.
- Kristof Karhan and Thomas D. Kühne. Energetic Cooperation in Hydrogen Bond Formation. *in preparation*, 2017
- Kristof Karhan and Thomas D. Kühne. Water Cooperation Revisited: The Case of the Surface. *in preparation*, 2017
- Kristof Karhan and Thomas D. Kühne. Water X-Ray Absorption Can Be Explained by Hydrogen Bond Asymmetry. *in preparation*, 2017
- Tobias Schneider, Kristof Karhan, Boray Torun, Guido Grundmeier and Thomas D. Kühne. Molecular vibrational decomposition of the IR-Spectrum of water absorbed on the anatase(101) surface . *in preparation*, 2017
- Julian Heske, Kristof Karhan, Boray Torun, Guido Grundmeier and Thomas D. Kühne. Effects of Anatase(101) Binding to the Surface of Water. *in preparation*, 2017
- Patrick Müller, Kristof Karhan, Matthias Bauer and Thomas D. Kühne. Impact of condensed-phase and finite-temperature effects on X-ray absorption spectroscopy simulations. *in preparation*, 2017

A.2 Thanks and Acknowledgements

First and foremost I would like to thank Thomas Kühne for being a calm and helpful advisor, who always had time for my questions and could provide helpful insights and comments during the years workin.

I would also thank my colleagues Frederik and Jure, who were amicable colleagues and greatly contributed to my enjoyment of being at the office. I'd also want thank them for having time for a chat so often.

To people who took their time to proof-read my work I also like to say thank you. These are, in no particular order, Frederik Zysk, Julian Heske, Claudia Schröder, Marta Gebremedhin, Markus Dörr, Jure Gujt, Melanie Köcher, and Andres Henao.

Finally I'd like to thank my family for supporting me throughout the entire duration of my PhD thesis.