Structure and Dynamics of the Hematite/Liquid Water Interface

PhD Thesis by

Guido Falk von Rudorff

Supervised by

Prof Jochen Blumberger and **Prof Alexander Shluger**

Submitted on

10 July 2017

in fulfillment of the requirements for the degree of Doctor of Philosophy from University College London (UCL)

I, Guido Falk von Rudorff confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The hematite/liquid water interface is of interest for a wide range of applications. This work investigates structural and dynamic properties of the interface for several experimentally relevant combinations of crystal cuts, termination and protonation schemes. To this end, molecular dynamics simulations at hybrid functional DFT level have been performed for both neutral and charged systems.

For bulk hematite, this work shows that the commonly used GGA+U level of theory gives good results for geometric properties, but cannot capture electronic structure equally well. However, a modified hybrid functional based on HSE06 is able to describe both geometric and electronic properties with sufficient accuracy.

The two most common crystal surfaces and their most common terminations have been investigated and compared to experiment where possible. The resulting structural information is in good agreement with experiment, but highlights the importance of dynamic equilibria for the solvation structure. This work also shows that classical force fields cannot readily describe the surface protonation structure and dynamics. Besides an atomistic view of the surface structure, protonation dynamics, surface restructuring mechanisms and interconversion of surface aquo groups to bulk solvent are discussed.

Two new methods are suggested in this work: (i) a screening method for Hartree-Fock exchange forces that significantly accelerates hybrid functional-based molecular dynamics calculation and (ii) a guided thermodynamic integration scheme for free energy estimation from short trajectories. The implementation thereof is tested and benchmarked and applied to the hematite/liquid water interface.

To accelerate the calculation of HFX forces in a molecular dynamics force screening scheme is proposed. For actual systems of interest, this can speed up the whole molecular dynamics run by a factor of three. The method is assessed on a wide range of materials and for various properties including energy conservation.

Finally, the pKa value of a surface protonation is calculated. By means of thermodynamic integration, the free energy difference between a proton at the surface of hematite and a solvated proton in bulk water is quantified. Various integration schemes are evaluated and a new analysis method is proposed to reduce human bias in the analysis and to automatically detect convergence of the vertical energy gap time series.

The work outlined above has been published:

- GF von Rudorff, R Jakobsen, KM Rosso, J Blumberger: Fast Interconversion of Hydrogen Bonding at the Hematite(001)-Liquid Water Interface, *J. Phys. Chem. Lett.*, 7(7), 1155, 2016.
- GF von Rudorff, R Jakobsen, KM Rosso, J Blumberger: Hematite (001)-liquid water interface from hybrid density functional-based molecular dynamics, *J. Phys. Condens. Matter*, 28(39), 394, 2016.
- ME McBriarty, GF von Rudorff, JE Stubbs, PJ Eng, J Blumberger, KM Rosso: Dynamic Stabilization of Metal Oxide-Water Interfaces, *J. Am. Chem. Soc.*, 139(7), 2581, 2017.
- GF von Rudorff, R Jakobsen, KM Rosso, J Blumberger: Improving the Performance of Hybrid Functional-Based Molecular Dynamics Simulation through Screening of Hartree-Fock Exchange Forces, *J. Chem. Theory Comput.*, 13(5), 2178, 2017.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Hematite (α-Fe_2O_3) is abundant,[1] cheap,[2] has low toxicity,[3] is stable under aqueous conditions[4] and has a band gap favourable for photo-oxidation.[5] These properties have made it a material of interest in a wide range of research disciplines: mineralogy,[6,7] colloid chemistry,[8,9] atmospheric science,[10] bioenergetics,[11–14] energy storage[15] and energy harvesting[16,17] (closely related to catalysis[18,19]), mercury filtering[20] and even archaeology.[21] In many of these cases (and for assessing properties like acidity[22,23] or surface reactivity[24–28]), the interface between hematite and liquid water is of crucial importance to the respective application. Despite numerous efforts both from the experiment perspective[29–31] or from theoretical simulations with various techniques (classical force fields,[32,33] ReaxFF,[34] kinetic Monte Carlo,[35] Density Functional Theory (DFT)[36–38]), discrepancies even for the equilibrium structure still exist.[17,30] These discrepancies can be partially attributed to differences and limitations of the approaches used.[39]

So far, the hematite/liquid water interface has been investigated experimentally using an X-ray diffraction method that is sensitive to the atomic configuration close to the surface of hematite. This method is powerful, but (i) cannot resolve protons due to their low scattering cross-section, (ii) cannot resolve dynamics, including dynamic equilibria and (iii) requires an detailed understanding of the atomistic structure to begin with due to technical details discussed later in this work. Therefore, it is highly desirable to complement the existing knowledge of the interface with an high-level atomistic picture such that a more detailed understanding on the surface and solvation dynamics emerges. Therefore, the aims of this thesis are

- Describe the atomistic structure of the hematite/liquid water interface for the simplest experimentally accessible case.
- Discuss the impact of different crystal surfaces and surface terminations on the solvation structure and dynamics.
- Investigate the protonation states of the solvated hematite surface and describe proton dynamics.
- Develop methods to answer the aforementioned physical questions using less human and computational resources.

To this end, the properties of bulk hematite have been studied first to identify a level of theory suitable to describe the properties at hand. Based on these results, a procedure was devised to prepare nearly equilibrated setup of the interface which then was the starting point for calculations on different crystal surfaces and surface terminations. Motivated by the high costs of these calculations, a force screening approach has been developed which allows to accelerate hybrid-functional based DFT molecular dynamics calculations. Finally, the pKa value of the surface protonation for one crystal surface has been assessed. In the following, the most important results are summarised.

In chapter 2, this work shows that the commonly used GGA+U level of theory gives good results for geometric properties, but cannot capture electronic structure equally well. In particular, the band gap is significantly underestimated. However, the hybrid functional HSE06 is able to describe both geometric and electronic properties with sufficient accuracy. The functional has been validated and shown to give good results for bulk water, as well. Being able to describe both the transition metal oxide, hematite, and the solvent, water, at the same level of theory makes it a good choice for describing interface dynamics.

In chapter 3, a reliable equilibration protocol has been developed to actually simulate the interface structure. With this method, the two most common crystal surfaces and their most common terminations have been investigated and compared to experiment where possible. The resulting structural information is in good agreement with experiment, but highlights the importance of dynamic equilibria for the solvation structure. This mechanism cannot be directly described by experiment, which is why the simulations are important. This work also shows that classical force fields cannot readily describe the surface

protonation structure and dynamics. The surface protonation OH groups form patterns that are stable on the picosecond time scale but quickly interconvert within 50 fs. This process restricts the availability of the surface protons for interaction with the solvent. Moreover, the hydrogen bonding pattern of water is disrupted at the surface which is related to the protonation dynamics including autoionisation processes. Finally, the interconversion of surface aquo groups and bulk water has been observed and described.

In chapter 4, I describe a method to accelerate the calculation of HFX forces in a molecular dynamics scheme. A detailed analysis showed that there are many small force contributions from individual orbital combinations that do not sum up to significant contributions even though there are so many of them. With a simple scheme to identify these small contributions, they can be neglected for short periods of time. This work presents a correction scheme that reduces the residual error in the force components to values smaller than convergence criteria typically employed for geometry optimisation. For actual systems of interest, this can speed up the whole molecular dynamics run by a factor of three. The method is assessed on a wide range of materials and for various properties including energy conservation.

Chapter 5 shows the calculation of the pKa value of a surface protonation that is currently debated in literature. By means of thermodynamic integration, the free energy difference between a proton at the surface of hematite and a solvated proton in bulk water is quantified. Various integration schemes are evaluated and a new analysis method is proposed to reduce human bias in the analysis and to automatically detect convergence of the vertical energy gap time series.

In summary, this gives a comprehensive picture of the dynamics of the surface protons from various perspectives and can help experimentalists in interpretation of their results as well as motivate further work on understanding the solvation interaction of this transition-metal oxide.

Coordination Scheme	Minimum radius ratio
Tetrahedron	0.225
Octahedron	0.414
Hematite	0.708
Cube	0.732

▲ **Table 1.1** – Comparison of coordination schemes based on purely geometric considerations.[41] According to the ratio of cation radius to anion radius, hematite is expected to fall within the octahedron regime. Value for hematite calculated from reference data.[40]

1.1 Properties of Hematite

1.1.1 Geometry

The general structure of hematite is similar to that of corundum, placing the oxygen atoms in a slightly distorted hcp where they form octahedrons containing one iron atom each, so each iron atom has six coordinated oxygen atoms. From the effective vdW radii for $Fe^{3+}\cdots O^{2-}$, which is 1.76 Å,[40] where the Fe^{3+} alone has an effective radius of 2.06 Å, a octahedron is expected as coordination polyhedron,[41] as shown in Table 1.1.

Figure 1.3 shows the supercell typically used as the basis for the simulation setups. It is a hexagonal cell with lattice angles $\alpha, \beta = 90°$ and $\gamma = 120°$. The two short lattice vectors have the same length (typically around 5.06 Å), while the other one is significantly longer (13.88 Å). The distortion of the geometry when compared to the ideal corundum structure, can be quantified by measuring specific lengths in the lattice. In Figure 1.3, these lengths are highlighted. For the iron sublattice, there is a short iron-iron distance (highlighted yellow in Figure 1.3), which coincides with an atomic bond. Parallel to this axis vector, there is a much longer distance between two iron atoms which in this case are not bonded (highlighted green in Figure 1.3). In the same way, there is a distinct nearest-neighbour distance for oxygen atoms. In Figure 1.3, this is the distance between the two red spheres without any bond. Apart from these bonds which are only formed by atoms of the same kind, there are bonds between heterogeneous atoms. In fact, there are two important nearest-neighbour distances that have to be taken into account, which are shown by the only coloured FeO_2 fragment in Figure 1.3. Typical values for the distances mentioned here can be found in Table 2.1.

For high temperatures, the antiferromagnetic ordering of hematite disappears.[42] The

E_G [eV]	E_G [eV]	E_G [eV]	E_G [eV]	E_G [eV]	E_G [eV]	E_G [eV]	E_G [eV]
1.69-1.76[47]	1.56-2.1[48]	1.95[49]	2[50]	1.94-2.15[51]	2.10[52]	2.1[53]	2.1[54]
2.14-2.15[55]	2.14-2.2[56]	2.14-2.22[57]	2.14-2.28[58]	2.18[59]	2.18[60]	2.2 ± 0.1[5]	2.2[61]
2.2[62]	2.58[63]	2.6-2.73[64]	2.67[65]	2.7[66]	2.71-2.74[67]		

▲ **Table 1.2** – Experimental direct band gaps E_G for bulk hematite.

transitions of hematite like the Morin transition at 30 kbar are well studied.[43,44] The most important feature of hematite with respect to high pressures is that the ratio of the two lattice constants c/a decreases with increased pressure, which means that the unit cell gets substantially quenched.[43,45] Both below and above the Morin transition, hematite in its antiferromagnetic state is chiral.[46]

1.1.2 Band Gap and Spin Structure

Table 1.2 shows the experimental band gaps as described in literature. For the work described here, a reference value of 2.1 eV has been used. Experimentally measuring the band gap is a challenging task. The recent publications on this topic mostly employ a single method for determining both the direct and the indirect band gap.

The main idea is to measure the absorption profile which gives the dependency of the absorption coefficient $\alpha(\lambda)$ on the wavelength λ. If this absorption spectrum is shown scaled as

$$(\alpha\omega)^n \propto \hbar\omega \tag{1.1}$$

theory predicts parts of the tail on the absorption edge. The intersections of lines fitted to these segments with the x axis then give the band gap. The difference between direct and indirect band gap lies in the exponent used for the rescaling of the measured absorption data, n, which is $1/2$ for the direct band gap and 2 for the indirect one.[68] This method is called a Tauc[69] plot. For diffuse reflection, this has to be converted using the Kubelka-Munk function[70]

$$F(R) = \frac{(1-R)^2}{2R} \tag{1.2}$$

where R is the diffuse reflectance. In this case, Eqn. 1.1 is

$$(F(R)\hbar\omega)^n \propto \hbar\omega \tag{1.3}$$

Due to the bonds between Fe and O, there is a partial charge transfer from the iron atoms to the oxygen atoms altering the electron configuration, as shown in Table 2.3. Over a broad temperature range, hematite is antiferromagnetic. Experiments have shown that the magnetic ordering is stable even for surface layers of less than 100 Å.[71] This property is essentially the argument why bulk hematite simulations and bulk properties can be used as starting point for modelling interfaces between hematite as anode and liquid water. Experimentally, it has been shown that this spin structure has a strong impact on the direction of electron transport in bulk hematite.[72]

1.1.3 Interface

Describing the interface of hematite and water is a major challenge. Partially, this is because there is little reliable information on how the interface looks like on an atomistic level, partially, because an in-depth analysis requires a high amount of computational resources. Previous studies[2] have identified the most important surfaces for typical hematite electrodes: (001), (012), (113). In this case, important surfaces are those which are most common for crystalline hematite, and therefore, contribute the most to the overall properties of anode surfaces. At the surface of the hematite slab, the regular pattern from the bulk crystal has to be terminated. Experimental evidence points towards two main terminations for the (001) surface, referred to as iron termination and oxygen termination.[73–77] As most of the evidence is gathered from dry* hematite surfaces, it has to be kept in mind that bringing the surface into contact with water may alter the order of stability of surface terminations—as it has been found to be the case for hematite surfaces other than (001).[29]

Besides geometric properties, studies with GGA functional suggest that the magnetic moment of iron atoms is heavily reduced near the surface.[81] According to these results, the reduction depends on surface termination and can be up to 50 % of the bulk value. For some terminations, the spin ordering changes from antiferromagnetic, which happens to be the bulk ordering, to ferromagnetic ordering. This may be supported by other reports that found bulk hematite to have a partially ferromagnetic behaviour.[42] However, it is likely that this effect is less pronounced than predicted by GGA, as the same authors evaluated GGA+U which estimated the decrease in spin population of surface iron atoms to be

* Dry in this context means that there was no water layer on hematite in these cases. It is typical to find water molecules adsorbed on these dry surfaces[78,79] which has been studied extensively by simulation.[80]

around 20 %.

For the water structure near the hematite-water interface, there is only computational data available.[32,82] To some extent, experimental results on the surface charge can be used to estimate the water structure.[83] Since the equilibration process of the hematite surface involves a classical MD first, it is desirable to start from a system that is not likely to show protonation or deprotonation events. This is why the simulations are started from the point of zero charge PZC that is expected to be between pH 6.7 and 9.2.[83–87]

While there are several spin pattern states that are possible for hematite,[42,88–90] it has been shown[91] experimentally, that the domain walls between areas of equal spin state are large compared to the setup that is the base of this work. Therefore, assuming an uniform spin state is well justified.

1.1.4 Water Splitting

One of the most prominent applications for the hematite/liquid water interface is water splitting. The overall water splitting process

$$2\,H_2O \overset{n\hbar\omega}{\rightleftharpoons} O_2 + 2H_2 \qquad \Delta G^\circ = 4.92\,\mathrm{eV} \tag{1.4}$$

it spatially separated in a photoelectrochemical water splitting process.[16] In fact there are two separate reactions

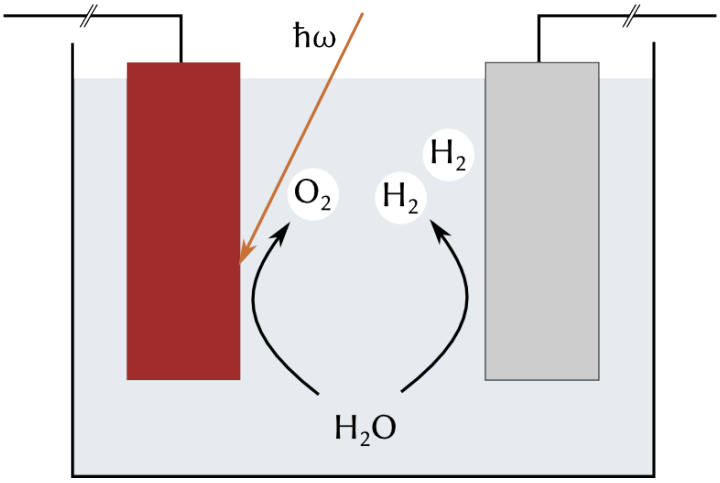
$$2\,OH^- + 2\,h^+ \rightleftharpoons H_2O + \frac{1}{2}O_2 \tag{1.5}$$

$$2\,H_2O + 2\,e^- \rightleftharpoons H_2 + 2\,OH^- \tag{1.6}$$

the first of which is happening at the anode, the second one at the cathode. In a schematic setup as shown in Figure 1.1, the incident photon provides the energy that is necessary to create an electron-hole pair. The electron is excited to the conduction band while the hole charge is transferred to the electrolyte-electrode interface at which the first part of the water splitting reaction takes place. At the cathode, the charge from the anode is used for removing the H^+ from the balance

$$H_2O \rightleftharpoons H^+ + OH^- \tag{1.7}$$

as the gaseous H_2 escapes the liquid electrolyte. The anode typically is a semiconductor, while the cathode is a metal.[16,92]

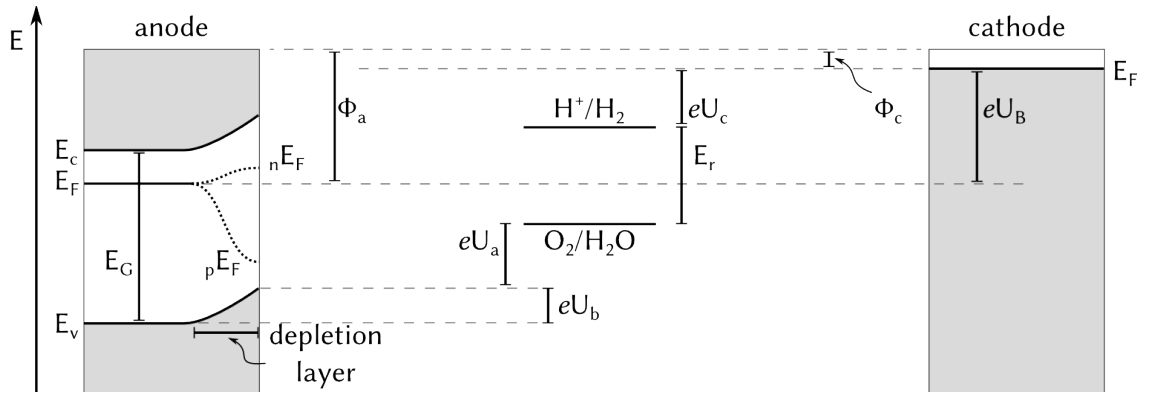


◀ **Figure 1.1** – Schematic setup for water splitting. The incident photon with energy $\hbar\omega$ provides the energy for the water splitting reaction at the anode (left). For $2\,H_2O$, the two residual H^+ propagate to the cathode (right) where they get reduced to $2\,H_2$.

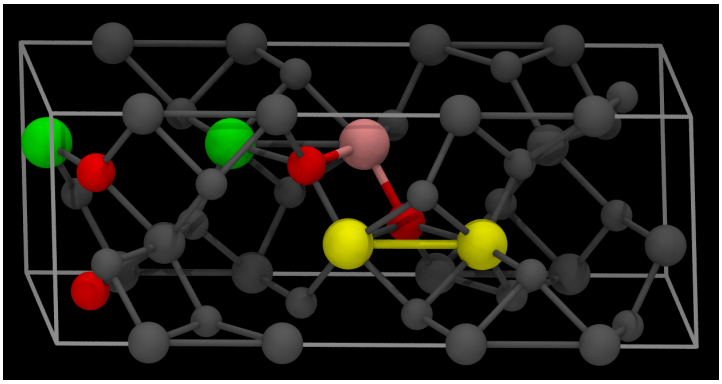
For this process to happen, there are many conditions, which can be explained by using the electronic structure of this cell, as shown in Figure 1.2. First of all, the absorption of the incident light has to be effective which means that the band gap of the semiconductor should match the AM1.5 spectrum energy.[16,92,93] As in most cases the goal is to use sunlight for energy harvesting, this band gap should be in the range of 1.0 eV to 1.4 eV, as this gives the highest conversion rate.[92] Additionally, the semiconducting anode material has to be chemically stable in the electrolyte and should be comparably inexpensive.

While ΔE gives the theoretical limit for an ideal material with no losses at all, various processes[92] have a high impact on the efficiency. Directly after the creation of the electron-hole pair due to the incident photon, this pair can recombine with the result that the energy of this photon is not available for the remaining steps of the process. While this has an impact on the electron yield, it does not affect the part of the spectrum that can be used for harvesting. Other properties like the resistance of the electrodes, the wires and contacts take energy from the photoelectron and, hence, reduce the usable spectrum range. All these losses are estimated to roughly 0.8 eV in a typical setup.[92] This leaves an optimal material band gap of 1.8 eV to 2.2 eV[92] or 1.6 eV to 1.9 eV,[93] depending on the estimation method employed.

In principle, there are two additional conditions the anode material has to fulfil, which can be relaxed under certain circumstances. There are two energy levels which are of importance for the electrolyte, namely the one of water oxidation and water reduction.[93] These two energy levels have to lie between the lower conduction band edge and the upper valence band edge of the anode material. For practical setups, there is no need to fulfil



▲ **Figure 1.2** – Energy levels for water splitting for an n-type semiconductor with incident light[16,92–95] with Fermi energies E_F, valence band edge E_v, conduction band edge E_c, band gap E_G, hole quasi Fermi level ${}_pE_F$, electron quasi Fermi level ${}_nE_F$, anode material work function Φ_a, anode over-potential U_a, bending potential U_b, water splitting energy $E_r = \Delta E$, cathode over-potential U_c, cathode work function Φ_c, and bias potential U_B.



▶ **Figure 1.3** – Supercell ($1 \times 1 \times 1$) of hematite (30 atoms). The distances between coloured atoms are used as special indicator for geometric properties. Small atoms are oxygen atoms, large spheres denote iron atoms. Despite the visual impression of forming a cuboid, the supercell is hexagonal with $\alpha, \beta = 90°$ and $\gamma = 120°$.

this condition, as this issue can be mitigated by applying an biasing potential in order to shift the energy levels of the anode material. By using a biasing potential, it is possible to overcome this limitation as long as the overall energetic difference between the valence and the conduction band, the band gap E_G, is larger than the difference in energy between the oxidation and the reduction levels ΔE.

The expression for the water-splitting efficiency depends on the material used and the interface in question.[94] For hematite, it has been estimated[4] that

$$\eta = \frac{1 - \exp(-\alpha W) + L_K/(L_K + \alpha^{-1})\exp(-\alpha W)}{1 + k/k_e(L_D/\sqrt{2\nu_0} + L_K/\exp(\nu_0))} \tag{1.8}$$

with the absorption coefficient α, the depletion width W, the Debye length L_D, the recombination distance L_K, the recombination rate within the hematite electrode k, the faradaic rate constant k_e, and the degree of the band bending ν_0.

In principle, it would be desirable to use the aforementioned properties to design the simulation cell. However, the dimensions required to do so are beyond reach of an atom-

istic approach. However, the interface region itself is accessible for a DFT approach. Some of the aforementioned properties, however, are not: The absorption coefficient is not commonly calculated within the DFT framework, as it would require extensive sampling of the excited states. While this can be done using TD-DFT, the computational costs are prohibitively large to use it for parametrization.[96] The depletion width in hematite when put into contact with water is estimated to be some 20 nm.[1] Therefore, a simulation that would be able to accurately determine the depletion width would have to span across at least 20 nm to 30 nm, which makes the DFT calculation too expensive again. Moreover, there is only little experimental data available to compare to.

Unfortunately, the physical importance of getting the absorption coefficient α and the depletion width W right is rather high, as they only appear together as $\exp(-\alpha W)$ in Eqn. 1.8, and with the typical penetration depth of the incident light (roughly 50 nm to 120 nm[97]) this expression evaluates to $\exp(-\alpha W) \simeq 0.84 \cdots 0.54$.

The Debye length L_D is a property of the solid-liquid interface and, hence, requires a complete setup of the interface. Moreover, it is a dynamical property which requires long trajectories for better statistics, which also renders this property unsuitable for an atomistic approach. The same argument applies to the recombination distance and even more to the recombination rate. Also the faradaic rate constant and the band bending falls into this category.

Figure 1.2 shows the energy levels that are formed at the interface of hematite and water in case hematite is used as anode material. The bulk anode features a conduction band level E_c and a valence band level E_v, which are separated by a comparably large band gap E_G of roughly 2.0-2.2 eV. The Fermi level E_F is between the conduction band and above the valence band, while the energy difference between the Fermi level and the conduction band is lower than is the energy difference to the valence band, as it is expected for an n-type semiconductor like hematite. All bands bend in the contact region of the anode and water. This volume is called depletion layer. Physically speaking, the depletion of charge carriers is due to the induced charge in the first hydration layers of the anode. The amount of this bending depends on the anode and electrolyte material.[98] Instead of using the energy difference directly, this property is quantified using the bending potential U_b. As for the Fermi energy level, one has to consider the level splitting in the depletion layer, in case

incident light creates electron-hole pairs in the anode material. If so, the number of both electrons and holes is increased such that the Fermi level of the electrons ${}_nE_F$ increases and the Fermi level of the holes ${}_pE_F$ decreases. However, this effect is much larger for the holes as they are the minority carriers in an n-type semiconductor. As for material properties both the anode and the cathode have a work function ϕ_a and ϕ_c, respectively. The pH of the electrolyte, in this case water, is able to shift the energy levels for the water splitting reaction. The energy difference between the valence band of hematite at the surface and the lower energy level of the water-splitting reaction is called anode over-potential U_a. As expected, the energy difference between the Fermi energy of the cathode and the higher energy level of the water-splitting reaction is called cathode over-potential U_c. The difference between the two energy levels of the water-splitting reaction is $E_r = \Delta E = 1.23\,\text{eV}$. In order to ensure that the Fermi energy level of the cathode is higher than the upper energy level of the water-splitting reaction, in some cases, a bias potential U_B has to be applied.

1.2 Density Functional Theory Calculations

1.2.1 Kohn-Sham DFT

The foundation for the Kohn-Sham (KS) DFT are the two Hohenberg-Kohn (HK) theorems. The first of which shows that the external potential is uniquely determined (up to a trivial constant) by a spatial electron density which in turn defines the electronic Hamiltonian and all properties. The second HK theorem proves that there is a universal energy functional that only depends on the spatial electron density and is minimised by the electron density of the ground state.[99]

In the HK approach, the general many-electron system is treated as a system of electrons interacting with each other. To make this more accessible by calculations, the KS-DFT approach treats electrons as single particles subject to an effective field caused by all other electrons. This requires a self-consistent calculation of the Kohn-Sham potential. In this framework, the total energy is given by

$$E[\rho] = T[\rho] + \int d\mathbf{r} V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) + E_H[\rho] + E_{\text{XC}}[\rho] \tag{1.9}$$

with the kinetic energy T, the external potential V_{ext}, the Hartree energy E_H and the exchange-correlation energy E_{XC}. The latter is hiding the real challenge since the modelling ap-

proaches for the exchange-correlation energy functional are plentiful but none of them is exact.

1.2.2 GGA+U

LSDA (local spin-density approximation) has been found to have severe issues with electronic properties like the spin population in various cases,[100–102] especially in the context of transition metal oxides, where the partially filled d orbital energy levels are degenerate.[103] For these systems, DFT typically overestimates the repulsive potential for electrons on any given site, which leads to reduced spin moment on the atom sites, which in turn narrows the band gap.[104]

A particularly successful method showed that these problems can be repaired to some extent by the application of an additional potential to the (in this case iron $3d$) orbitals on a single atom.[105–107] In that case, the total energy $E[\rho]$ as a function of the electron density ρ is modified as follows

$$E^{+U}[\rho] = E[\rho] + \frac{\overline{U} - \overline{J}}{2} \sum_{\sigma} \sum_{j} \rho^{\sigma}_{jj} - \sum_{l} \rho^{\sigma}_{jl} \rho^{\sigma}_{lj} \tag{1.10}$$

where ρ^{σ}_{jl} applies to the partially filled d density matrix. σ denotes the spin component, $\overline{U}$ represents the Coulomb component of the interaction, while $\overline{J}$ gives the exchange interaction.[108] In the implementation in CP2K, only the effective potential $U_{\text{eff}} := \overline{U} - \overline{J}$ is specified. From the above equation it is immediately clear that a vanishing effective potential gives the total energy without the Hubbard-U model.

The overall idea then is to choose U_{eff} such that the desired electronic properties come out better. This typically improves the band gap—for some materials, LSDA predicts semiconductors to be metallic, which is fixed in this approach depending on the effective U value.[108] So it is a valid approach to adjust this value until the band gap is reproduced correctly. For some implementations, $\overline{J}$ has to be known explicitly. Typically, one assumes 1 V in this case,[89] although the reference cited therein states that the value is rather to be found at $0.5 - 0.75$ V.[109] $\overline{J}$ is expected to depend mostly on the main quantum number n, while the effect on the occupation number is less pronounced.[109]

While there are different schemes on how to derive the Hubbard U parameter from *ab-initio* calculations,[110–113] they may yield incompatible results.[114] Typically, the U_{eff} value is optimized for any specific setup and is hardly transferable, which is why it is

commonly given with one significant digit only. But even then the variation of the values given in literature for hematite are significantly different:[89,115–117] they range from 4 V to 6 V. This seems to be a valid range for a wide range of compounds that are similar to hematite.[118–122] A minority reports values of less than 3.5 V as they observe inverse orbital ordering at the band gap otherwise.[123] This minority result is disputed however, and may be implementation dependent.[89] In some cases, different U values have been applied to different layers of hematite at a surface[117] which is meant to deal with the differences in coordination geometry.

One can express[106] the orbital energies in terms of this effective U value

$$\varepsilon_i^{+U} = \varepsilon_i + \begin{cases} eU_{\text{eff}}\left(\frac{1}{2} - n_i\right) & d \text{ orbital} \\ 0 & \text{otherwise} \end{cases} \tag{1.11}$$

where n_i is the occupation number of the specific orbital. Essentially, this means that occupied orbitals are shifted towards lower energies by $eU_{\text{eff}}/2$ and unoccupied ones are shifted towards higher energies by the same amount. This only applies to the partially filled orbital the correlation of which may be underestimated. For hematite, this is always the iron $3d$ orbital, while this may also be an f orbital for other materials.[108]

The shifting of the energies depending on the occupation number is plausible if one considers the electronic levels near the band gap only. LSDA underestimates the band gap, which then increases with U_{eff}. The minority opinion[123] on the dangers of increasing U_{eff} in terms of orbital order around the band gap may be justified given the hematite PDOS for HSE06(12%) in Figure 2.8. The oxygen $2p$ levels are not affected by the +U term, the occupied iron orbitals would shift down (which has no effect on the orbital ordering) but the unoccupied iron orbitals may shift towards higher energies leaving the band gap framed by oxygen $2p$ levels. This would change the insulator type of hematite. Therefore, it is plausible to assume that there is an upper limit on the U_{eff} value that is suitable for hematite. However, up to 6 V seem to be acceptable.[89,115,116]

1.2.3 HSE06

HSE06[124] is based on HSE03,[125–127] which in turn interpolates between PBE[128,129] and PBE0.[130] The exchange-correlation energy expression of HSE06 is

$$\begin{aligned} E_{XC}^{HSE} =& aE_X^{HF,SR}(\omega) + (1-a)E_X^{PBE,SR}(\omega) \\ &+ E_X^{PBE,LR}(\omega) + E_C^{PBE} \end{aligned} \tag{1.12}$$

with SR denoting short-range and LR labelling the long range part under the identity

$$\frac{1}{r} = \underbrace{\frac{1-\mathrm{erf}(\omega r)}{r}}_{\mathrm{SR}} + \underbrace{\frac{\mathrm{erf}(\omega r)}{r}}_{\mathrm{LR}} \tag{1.13}$$

The screening parameter ω has been set to different values for the HF and the PBE parts of eqn. 1.12 initially for HSE03 but in HSE06, these two parameters are identical. Weighting the Hartree Fock component by 25 % can be motivated from perturbation theory considerations.[131]

$$E_{XC}^{PBE0} = \frac{1}{4}E_{XC}^{HF} + \frac{3}{4}E_X^{PBE} + E_C^{PBE} \tag{1.14}$$

$$E_{XC} = E_X + E_C = \int d\mathbf{r} f(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})) \tag{1.15}$$

where E_{XC}^{HF} is the Hartree-Fock exchange-correlation energy[132] with the density matrix ρ

$$E_X^{HF} = -\frac{1}{4}\int d\mathbf{r_1}\int d\mathbf{r_2}\frac{|\rho(\mathbf{r_1},\mathbf{r_2})|^2}{|\mathbf{r_1}-\mathbf{r_2}|} \tag{1.16}$$

while the correlation energy is given as E_C. The predecessor HSE03[127] employs a similar form

$$E_{XC}^{HSE03} = aE_X^{HF,SR}(\omega) + (1-a)E_X^{PBE0} \tag{1.17}$$

For Hematite, HF alone overestimates the band gap.[90,133]

1.2.4 ADMM

One of the methods crucial to this work is ADMM[134] which approximates the density matrix $\mathbf{P}$ with an auxiliary density matrix $\tilde{\mathbf{P}}$

$$E_X^{HFX}[\mathbf{P}] = E_X^{HFX}[\tilde{\mathbf{P}}] + E_X^{HFX}[\mathbf{P}] - E_X^{HFX}[\tilde{\mathbf{P}}] \tag{1.18}$$

$$\simeq E_X^{HFX}[\tilde{\mathbf{P}}] + E_X^{DFT}[\mathbf{P}] - E_X^{DFT}[\tilde{\mathbf{P}}] \tag{1.19}$$

This auxiliary density matrix $\tilde{\mathbf{P}}$ is either much sparser or smaller than the original density matrix. Deriving it, however, is a highly non-trivial task since the quality of the results depend on the purification method.[134] This work uses the auxiliary basis sets as available in CP2K, which have been prepared using standard approaches by S. Ling and B. Slater (both UCL).[135,136]

1.2.5 Dispersion Correction

For higher accuracy in the treatment of van der Waals contributions to the PES, dispersion corrections have been developed. In most cases, the total energy is corrected by a dispersion energy, which is calculated from the pairwise interaction distances. While this approach is similar to what is employed for classical force fields, this can be considered not to be purely *ab initio* any more, as it requires fitted parameters that introduce empirical approximations. As these fitted parameters in turn are derived from high level *ab initio* calculations, this may be acceptable.

For Grimme-D3 dispersion correction,[137] the dispersion energy contribution follows

$$E_{\text{disp}} = -\frac{1}{2}\sum_{\text{A}\neq\text{B}}\sum_{n} s_n \frac{C_n^{\text{AB}}}{R_{\text{AB}}^n} f_d(R_{\text{AB}}) \tag{1.20}$$

where A and B loop over all atoms and n iterates all exponents that are used for the correction, typically a few even integers greater than or equal to six. The overall dispersion energy can be scaled for each contribution from different exponents n separately. This is done by s_n, which in turn also depends on the functional that is used for the simulation that is to be corrected for dispersion errors. Basically, the idea is that the amount of correction that is necessary or adequate may significantly differ with the amount of (effective) dispersion that is already covered by the functional. C_n^{AB} is related to the potential depth (similar to the well-depth parameter ε in the Lennard-Jones potential) and depends on n. The value of this coefficient is determined by averaging over a set of test compounds and all spatial directions. R_{AB} is just the distance between the atoms A and B, while $f_d(R_{\text{AB}})$ is the damping function that reduces the impact of the dispersion correction for larger distances even faster than by the power expression from R_{AB}^n. The damping function f_d should only have little impact on the overall results.[137] There are many different suggestions on the exact form of this damping potential.[138–141] The authors that have suggested the energy term in eqn. 1.20, have claimed to get good results using BJ damping. The main difference between

the former versions of Grimme dispersion correction[139] and this version of the damping function resides in the behaviour for very small interatomic distances. For this damping function[139]

$$f_d(R_{AB})^{(1)} = \frac{1}{1+\exp\left(-\alpha(R_{AB}/R_0 - 1)\right)} \tag{1.21}$$

we get

$$\lim_{R_{AB}\to 0}\left(f_d(R_{AB})^{(1)}\right) := L^{(1)} = \frac{1}{1+\exp\alpha} \tag{1.22}$$

while another suggestion[138] is

$$f_d(R_{AB})^{(2)} = \frac{1}{1+\alpha(R_{AB}/R_r)^{-12}} \tag{1.23}$$

$$L^{(2)} = 0 \tag{1.24}$$

where R_r is the sum of the experimental[142] van der Waals radii of the two atoms, which also is relevant to another suggestion,

$$f_d(R_{AB})^{(3)} = \left(1-\exp\left[-c_d\left(\frac{R_{AB}}{R_r}\right)^3\right]\right)^2 \tag{1.25}$$

$$L^{(3)} = 0 \tag{1.26}$$

where c_d is a fitted constant which does not depend on the elements of atom A and B but has values between 3.54 and 7.19.[140,141] The Fermi function

$$f_d(R_{AB})^{(4)} = \left[1+\exp\left(-\beta\left(\frac{R_{AB}}{R_r}-1\right)\right)\right]^{-1} \tag{1.27}$$

$$L^{(4)} = \frac{1}{1+\exp\beta} \tag{1.28}$$

has been evaluated as well,[140] giving inferior results. The main difference between the Fermi definition and the similar one in Eqn. 1.21 is that R_r is experimental in nature, while R_0 is subject to fitting. The most recent form, the BJ damping[143] follows

$$E_d = -\frac{1}{2}\sum_{A\neq B}\sum_n \frac{C_n^{AB}}{R_{vdw,AB}^n + R_{AB}^n} \tag{1.29}$$

$$R_{vdw,AB} = a_1 R_{c,\,AB} + a_2 \tag{1.30}$$

$$R_{c,\,AB} = \frac{1}{3}\left[\sum_{i>j}\left(\frac{C_{i,\,AB}}{C_{j,\,AB}}\right)^{\frac{1}{i-j}}\right] \tag{1.31}$$

where $i, j \in \{6, 8, 10\}$, $a_1 = 0.83$ Å and $a_2 = 1.55$ Å. This gives a damping function

$$f_d(R_{AB})^{(5)} = \frac{R_{AB}^n}{R_{vdw,AB}^n + R_{AB}^n} \tag{1.32}$$

Although the authors who suggested this form of damping also suggest a correction of the electron correlation energy,[143] this correction is not part of the BJ damping that has been used for D3 dispersion correction.[137]

Chapter 2

DFT Calculations on Bulk Materials

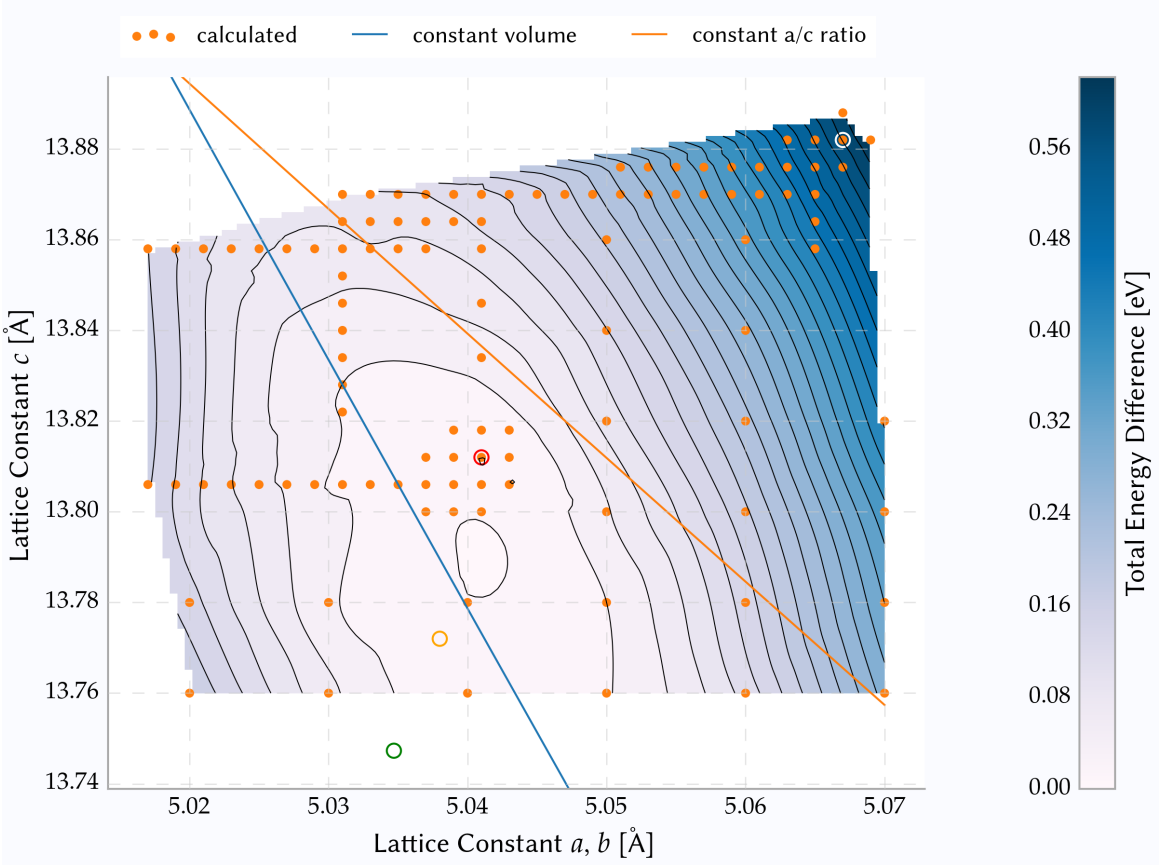
Before investigating the interface between two materials, it is important to understand the capabilities and limitations of the simulation method for the two bulk phases involved in the setup. This chapter is dedicated to growing confidence in the stability and reliability of the two outermost segments of the interface.

2.1 Hematite

As the features of hematite are much more important than those of liquid water for the interface, the optimization of the method applied has been done with hematite cells. Typically, the simulations have been carried out with integer multiples of the hexagonal supercell containing 30 atoms.

2.1.1 Geometry

Figure 2.1 shows the PES for HSE(12%)-D3, that is HSE(12%) extended by dispersion correction.[146] In this case, the total energy for a $3 \times 3 \times 1$ supercell has been calculated from a geometry optimization which in turn was initialized with the coordinates from a GGA+U cell. In all separate simulations, the three lattice constants (a, b, and c) have been kept fixed while the geometry could relax under this constraint. Furthermore, as theory predicts that a and b should be equal, they have been set to the same value. This assumption is backed by experimental data.[144, 145] The method allows for a two-dimensional sampling of the PES for this functional.



▲ **Figure 2.1** – Potential energy surface for bulk hematite and HSE(12%)-D3 cell geometry depending on the cell geometry. Filled orange circles are calculated values, the red circle denotes the minimum. White circle shows the GGA+U minimum.[81] Experimental values are shown in green[144] and orange.[145]

Calculation(s) 893

From Figure 2.1, it can be seen that the GGA+U optimized cell is comparably off from the experimental results which in turn do not completely agree with each other—not even in the statistical errors that are stated in the reference publications. For the GGA+U cell dimensions

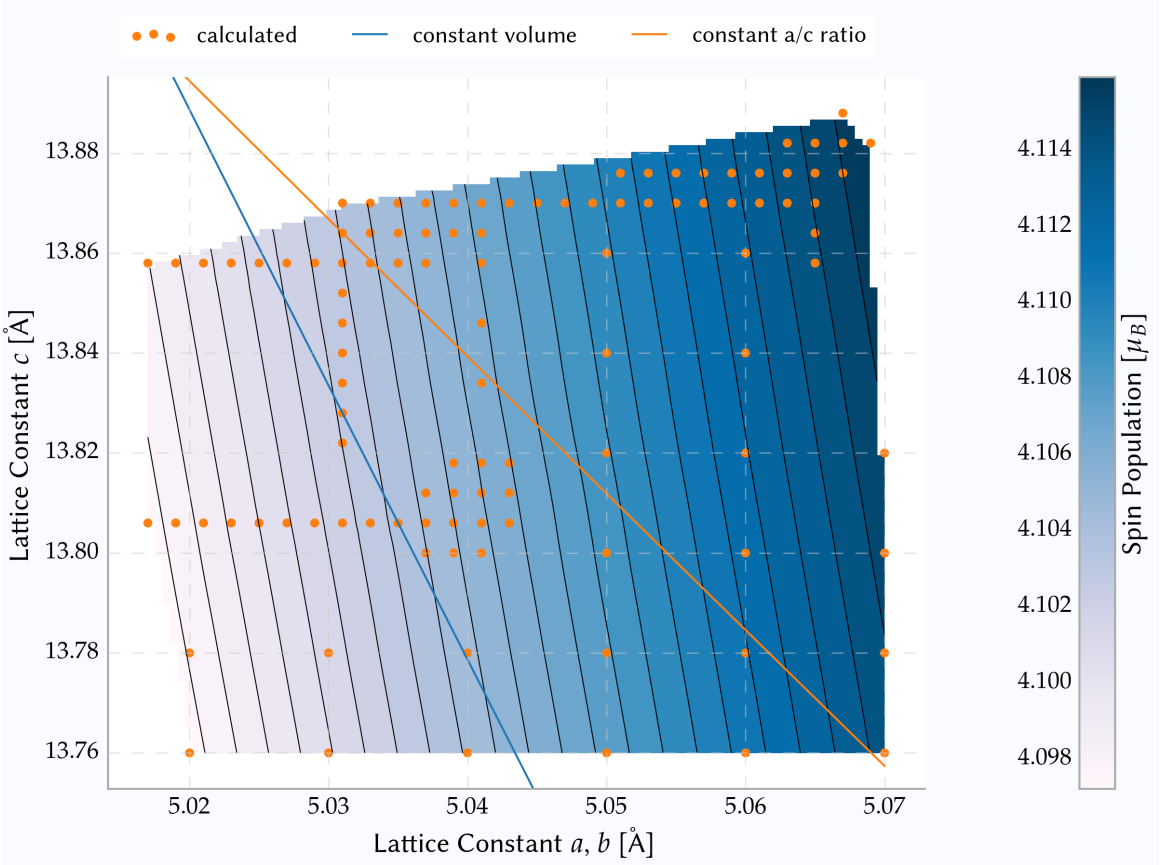
$$a, b = 5.067\,\text{Å}$$

$$c = 13.882\,\text{Å}$$

the ratio c/a is 2.7397. The experimental values mostly agree with each other: 2.7305[144] and 2.7336,[145] respectively. For HSE(12%)-D3, the optimized cell dimensions

$$a, b = 5.041\,\text{Å}$$

$$c = 13.812\,\text{Å}$$



▲ **Figure 2.2** – Average absolute spin moment for iron atoms of bulk hematite for HSE(12%)-D3 cell geometry depending on the cell geometry. Orange circles are calculated values, the remaining data has been interpolated cubically.

Calculation(s) 893

result in a ratio of 2.7399, which is practically identical to the one from GGA+U. This means that while the overall cell has shrunk under switching from GGA+U to HSE(12%)-D3, the ratio, and, therefore the internal geometric structure is largely unchanged. This is supported by the contour lines which are essentially isochores far away from the minimum on the PES. This means that the energy contribution from mechanical stress is the most important one.

For stability considerations, Figure 2.1 make clear that there is a direct pathway on the PES of HSE that connects the GGA+U geometry with the local minimum for the cell geometry. Moreover, the simulation results suggest, that this pathway is well within the quadratic region, which would assert that a simulation that would start from the GGA+U cell setup would find this minimum easily. The relaxation frees roughly 0.6 eV for the whole $3 \times 3 \times 1$ setup of 270 atoms. This excess energy could be handled by a thermostat, but this would

mean that this heat would dissipate into the water of any interface simulation, where the additional heat would change dynamics. This is particular important as thermodynamic properties of liquid ab initio MD simulations are highly sensitive to unstable temperature conditions.[147]

As it has been shown that the cell geometry of bulk TiO_2 does react considerably to dispersion correction,[148] the cell geometry has to be re-evaluated after changing the DFT method.

2.1.2 Electronic Properties

Table 2.1 shows a comparison of the results of typically used functionals when applied to a hematite super cell. Besides the different approaches in terms of functionals, the methods give quite similar qualitative results, as they all yield similar bond lengths and spin populations. However, from the data it is immediately apparent that the finite size effects for a single super cell of 30 atoms are quite strong, as they even affect geometric properties which typically are less sensitive. For example, PBE both over- and underestimates bond lengths by roughly 10 %. Surprisingly, this only affects bonds that involve at least one iron atom and this effect is the strongest for coordination distances between iron atoms. As the basis set for the data in Table 2.1 is identical for all PBE simulations except for the reference work, it is most likely that this is the influence of the small box, as the lattice constants have not been optimized for PBE but rather for GGA+U (see section for bulk systems).

For HSE(12%), coordination distances are mostly slightly overestimated when compared to literature data. The only exception is the shorter distance between iron atoms and oxygen atoms, which is slightly underestimated. This points towards a generally distorted structure, which is most likely due to the fact that the box dimensions themselves have been optimised for GGA+U instead of HSE at this stage. However, the results are compatible to computational reference data with the same functional but with plane waves instead of atomic orbitals.

The picture is a bit different for the electronic properties, where the chosen basis set and functional are expected to be more important.[154] While the experimental band gap is reproduced by HSE(12 %), PBE significantly underestimates the band gap, and HF drastically overestimates the same property. To some extent, this is expected, as HSE is a mixture be-

	Fe · · Fe (A)	Fe · · Fe (B)	Fe · · O (A)	Fe · · O (B)	O · · O	Volume [Å^3]
PBE $1 \times 1 \times 1$	3.319 (+12%)	3.623 (−9.2%)	1.868 (−3.8%)	2.280 (+6.0%)	2.680	308.7* ± 0.1*
PBE $2 \times 2 \times 1$	2.953 (±0%)	3.987 (±0%)	1.945 (±0%)	2.149 (±0%)	2.682	308.7* ± 0.1*
PBE $3 \times 3 \times 1$	2.950 (±0%)	3.994 (+0.1%)	1.937 (−0.2%)	2.123 (−1.3%)	2.683	308.7* ± 0.1*
PBE[149] $2 \times 2 \times 1$	2.95	3.99	1.94	2.15	–	308.7* ± 0.1*
GGA[a][89]	2.941	4.006	–	–	–	301.8[b]
GGA[c][81]	2.929	3.998	–	–	–	300.2* ± 0.1*
GGA+U[81]	2.896	4.044	–	–	–	301.4* ± 0.1*
HSE06[149]	2.91	4.03	1.96	2.13	–	308.7* ± 0.1*
HF[149]	2.88	4.06	1.97	2.11	–	308.7* ± 0.1*
HSE(12%) $2 \times 2 \times 1$	2.939 (−0.1%)	4.002 (+0.1%)	1.961 (+0.6%)	2.137 (−0.1%)	2.685	308.7* ± 0.1*
HSE(12%) $3 \times 3 \times 1$	2.938 (−0.1%)	4.005 (+0.1%)	1.965 (+0.8%)	2.157 (+0.8%)	2.696	308.7* ± 0.1*
HSE(12%)[149]	2.94	4.00	1.95	2.14	–	308.7* ± 0.1*
Experimental	2.883[150]	3.98[149]	1.985 ± 0.025[150]	2.060 ± 0.035[150]	2.545 ± 0.060*[150]	300.7 ± 2.6*[150]
Experimental[144]	2.896 ± 0.002	–	1.944 ± 0.002	2.113 ± 0.003	2.665 ± 0.007	–[d]

[a] Note that their labelling of the electromagnetic ordering is different from the notation used here. They use + + - -.
[b] Constrained volume and cell vector ratio.
[c] Note that in this paper, the magnetic ground state is + + - -.
[d] The publication only gives the octahedral volume, i.e. the volume of the polyhedron formed by all oxygen atoms next to an iron atom. This is not directly convertible to the unit cell volume.

▲ **Table 2.1** – Atom distances and volumetric properties in the super cell of hematite (see Figure 1.3) from this work (Γ point only) compared to plane wave calculations[149] ($5 \times 5 \times 5$ Monkhorst-Pack k-point mesh[151]) and experimental values.[150] Differences to calculations with the same functional are shown as relative errors. Data derived from data in cited sources is marked with an asterisk.

Calculation(s) 3dd, 689, 42e, af9, d9b, b42

	$\lvert\mu\rvert$ on Fe[a]	$\lvert\mu\rvert$ on Fe[b]	E_g
PBE $1 \times 1 \times 1$	3.77		–[c]
PBE $2 \times 2 \times 1$	3.68 ± 0.05	3.85	0.92
PBE $3 \times 3 \times 1$	3.67 ± 0.05	3.83	0.95
PBE[149] $2 \times 2 \times 1$	3.60		1.17
PBE[36]		3.6	
GGA[d][89]	3.44		0.3
GGA[e][81]	3.45		0.5
GGA+U[81]	4.11		2.0
PBE+U[80]			2.0
HSE06[149]	4.16		4.02
HF[149]	4.60		15.7
HSE(12%) $2 \times 2 \times 1$	3.96 ± 0.05	4.12	2.20
HSE(12%) $3 \times 3 \times 1$	3.97 ± 0.05	4.11	2.16
HSE(12%)[149]	4.16		2.56
SCF-Xα-SW[152][f]	4.69		
Experimental[145]	4.64		2.7
Experimental[153]	4.9		

[a] Wigner-Seitz population analysis unless experimental result
[b] Mulliken population analysis unless experimental result
[c] Calculation of the projected density of states not possible due to numerical instability.
[d] Note that their labelling of the electromagnetic ordering is different from the notation used here. They use ++--.
[e] Note that in this paper, the magnetic ground state is ++--.
[f] Simulation, but without population analysis.

▲ **Table 2.2** – Electronic and magnetic properties of hematite from this work compared to plane wave calculations[149] and experimental values.[150] Differences to calculations with the same functional are shown as relative errors.

Calculation(s) 3dd, 689, 42e, af9, d9b, b42

tween those two. Less surprisingly,[155] the spin moments depend on the population analysis method. Natively, CP2K supports Mulliken charge fitting[156] amongst others, which typically attributes higher absolute spin moments to the iron atoms than Wigner-Seitz population analysis[157] would do. This can be directly seen from the data in Table 2.2. Moreover, PBE again underestimates the spin moments much more severely than HSE(12 %) does.

Upon changing the cell dimensions, a relaxation of the internal geometry of bulk hematite is observed. As this slightly changes the atom distances, the overlap of delocalised electrons may change as well, resulting in a different spin distribution over the neighbouring iron atoms. This is expected to have some impact on the band structure, and, hence on

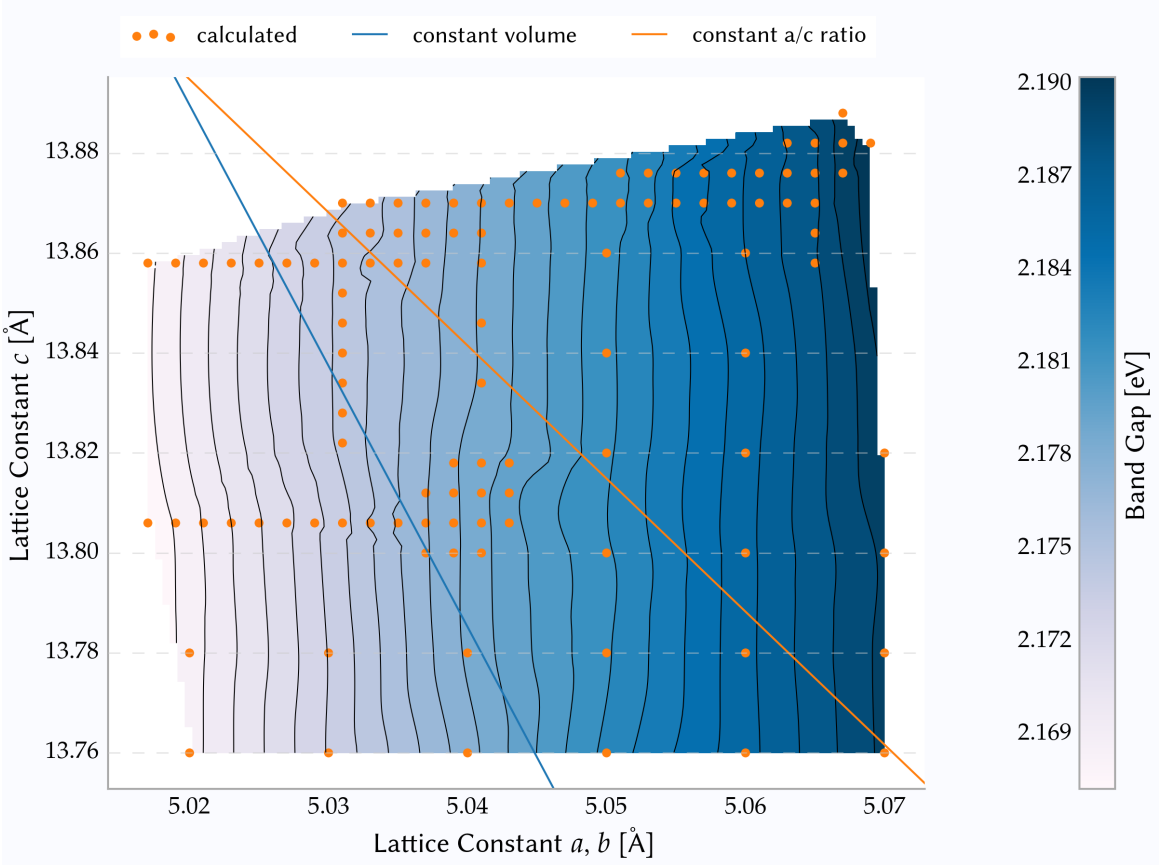
the band gap. For all the simulations shown in Figure 2.1, the average absolute spin population of the iron atoms has been determined (see Figure 2.2) and the band gap has been investigated (see Figure 2.3).

Interestingly, the spin population and the band gap exhibit a different dependency on the box geometry. Neither of these two properties solely depends on the ratio of the lattice constants or the volume of the simulation cell which means that the geometry is distorted in a non-linear fashion. The spin moment is less sensitive to changes in the length of the lattice vector **c**, that is the vector pointing in direction of the iron atoms in the unit cell containing one formula unit. The distances between iron atoms along this vector are alternating between two values, the shorter of which denotes bond between the irons in question. This is a hint that the delocalisation of the spin moment perpendicular to this axis is more important than the one along this axis. The reduction of the spin population with decreasing volume of the cell is somewhat expected, but the highly linear decrease itself is surprising, as the sampling efforts cover a number of cell dimensions which are well apart from the minimum on the PES.

The overall variance in the spin population is small. The relative difference in the spin population is one tenth of the relative difference of the lattice vectors.

For the band gap, Figure 2.3 shows that around the minimum on the PES, the band gap mainly depends on the lengths of the two short lattice vectors, while there is no clear evidence for any dependence on the long lattice vector. Some variations in the contour lines are due to artefacts resulting from the cubic interpolation, which can be easily seen in areas in lattice vector space which are sampled with a high density. In these areas, the band gap is nearly constant with respect to changes in the length of the longest lattice vector. While the linearity of the data is not as pronounced as it is for the spin population, again, the nearly constant decrease is a striking feature.

The variation of the band gap with scaling of the lattice vector lengths is comparable to the one of the lattice vectors. In contrast to the spin population, this means that the band gap is more heavily influenced by changes of the geometry as performed in this simulation. In any case, the band gap value is within the range of typical experimental results, so one can safely assume that small changes in the cell geometry due to slight changes in the dispersion correction are less likely to have a significant impact on the band gap, which is



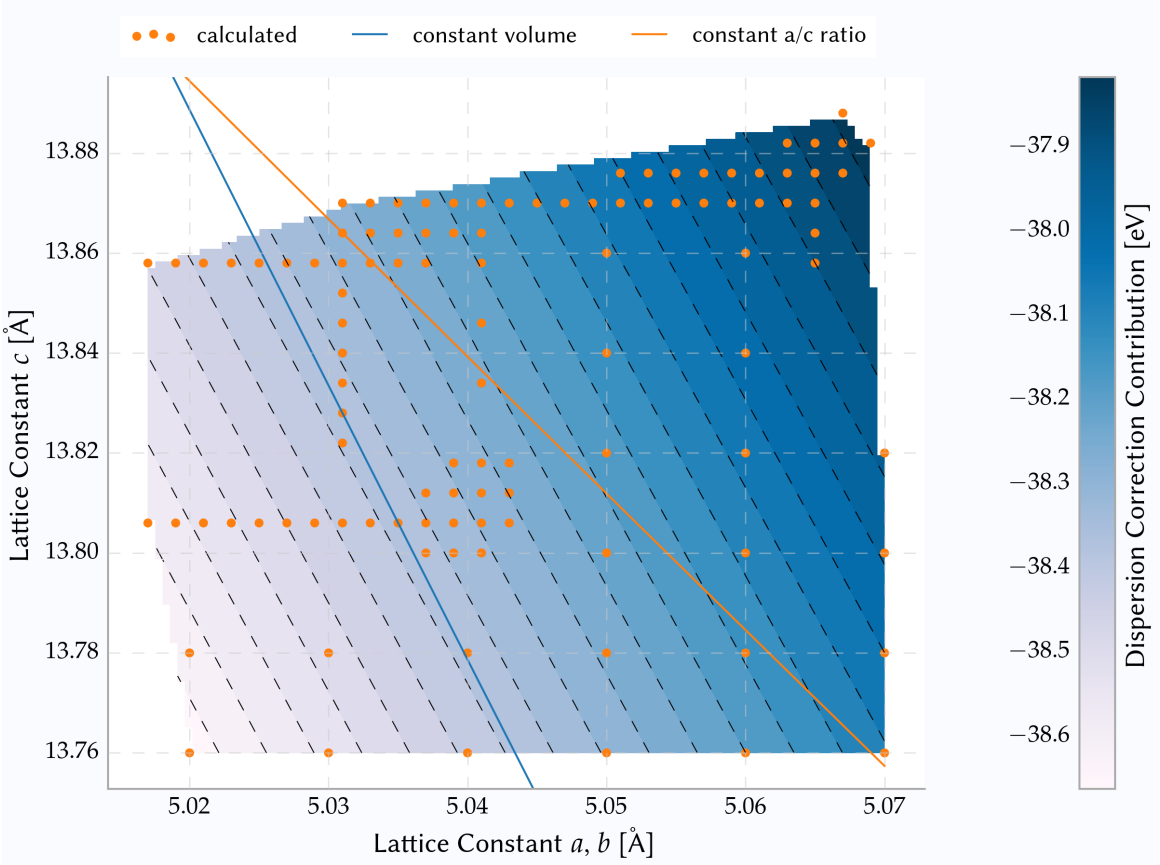
▲ **Figure 2.3** – Band gap of bulk hematite for HSE(12%)-D3 cell geometry depending on the cell geometry. Orange circles are calculated values, the remaining data has been interpolated cubically.

Calculation(s) 893

one of the most important properties for the hematite water interface.

The band gap is expected to depend on the ω parameter of HSE06,[124] which in turn determines the switching between the HF exchange region and the PBE electron exchange domain. When Grimme dispersion correction comes into play, this introduces at least one additional cutoff. In the case of BJ damping, dispersion correction even introduces two additional spatial decay functions.

To estimate the influence of the dispersion correction method on the overall PES, the ideal way would be to perform two scans of the dependency of the total energy on the lattice parameters: one with D3 correction and one without. However, this is computationally expensive. Given the fact that the relative contribution of the dispersion correction E_{D3} to the total energy E is in the order of $E_{\mathrm{D3}}E^{-1} \simeq 8 \cdot 10^{-5}$, it is a reasonable approximation to assume that the PES of hematite looks highly similar with or without dispersion correction. This allows to have a look at the isolated energy contribution of the D3 dispersion



▲ **Figure 2.4** – Dispersion correction contribution to the total energy for HSE(12%)-D3 cell geometry depending on the cell geometry. Orange circles are calculated values, the remaining data has been interpolated cubically.

Calculation(s) 893

correction as an approximation to the real difference between the two potential energy surfaces. Figure 2.4 shows the absolute energy contribution of the dispersion correction. It is directly apparent, that the contribution is mostly constant for a specific volume, disregarding the values of the lattice parameters. This is surprising, as the material itself is highly anisotropic. It can be explained, however, together with the regular spacing of the contour lines. Regular spacing means nearly linear dependence of the dispersion energy on the volume, and linear dependence is only likely to occur for a van der Waals potential far away from the minimum of the dispersion potential. Therefore, the most relevant part of the potential is the tail region which can be well approximated with a linear function. Physically speaking, in most cases the relevant contribution is for large distances. This renders the definition of the actual correction coefficients C_n in Eqn. 1.29 less important. In fact, the influence of the dispersion correction on the total geometry of hematite is negligible, but it

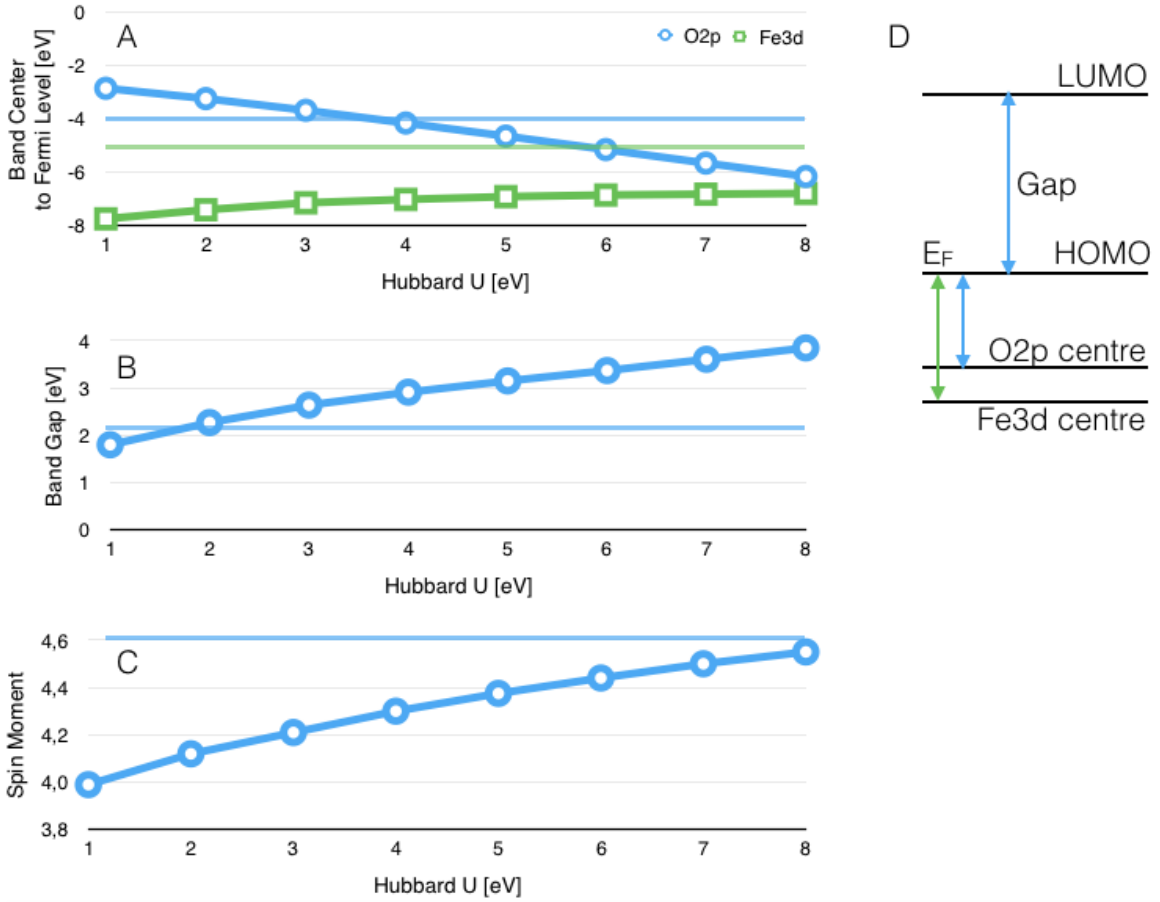
may be of importance for modelling the properties of a two-dimensional slab.[158]

For all electronic properties there is a method suggested mainly to overcome the band gap problems resulting from the SIE of the LSDA approach,[159] namely the DFT+U method (see page 23). The major problem in this context lies in the choice of the U_{eff} value that defines the height of the potential on the atomic spin population and, in theory, shifts the valence and conduction band centres by $U_{\text{eff}}/2$ in a way that the band gap is increased.

For bulk hematite, various values for U_{eff} have been reported[80,81,115,123]—mostly ranging from 4 to 6 V. But to my best knowledge, these values are optimised in order to reproduce the band gap only. Other electronic properties like the band ordering have been considered very rarely[123] and not in a systematic way. On the way towards understanding the limitations of a DFT+U approach in the interface, it is vitally important to check the influence of the U_{eff} value on the relevant electronic properties in the bulk system first. In order to achieve this, a series of calculations has been performed with PBE including the correction in question for a $1 \times 1 \times 1$ super cell. After a geometry optimisation for these cell vectors, the PDOS of the final geometry was calculated.

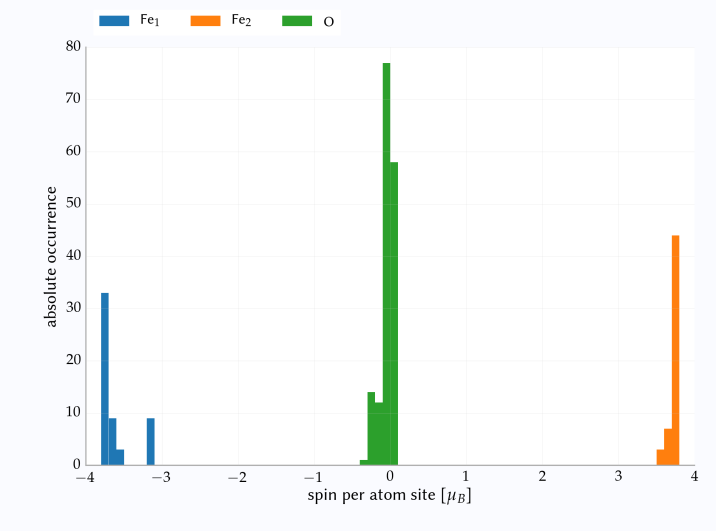
Figure 2.5 shows the expected behaviour for the band gap value, i. e. the band gap is increasing for larger values of U_{eff}, although slower than expected for fully occupied or fully depleted orbitals. With saturating spin population, the slope is getting constant, which matches expectations again. It is surprising however that the band gap matches the experimental value for very small values of U_{eff} (optimum is around 1.5 V), which does not match the published information.

The spin population on the iron atoms increases for larger U_{eff} as well but it approaches the experimental value asymptotically from below, which is reasonable in the physical picture, as there is only a limited number of electrons that can be forced to be located on any given site. The influence of the U_{eff} on the oxygen $2p$ centre is very strong (much stronger than on the iron $2d$ centre) and most surprising. By design, the DFT+U approach only works on a specific orbital, the iron $3d$ orbital in this case. So in principle, there is no direct influence on the oxygen $2p$ to be expected unless there is a mixing of the states. Therefore, the behaviour shown in Figure 2.5 has to be a side effect of forcing electrons onto the iron sites. As the oxygen sites remain close to zero in spin moment (as they should), this means that spin up electrons are forced onto the spin up iron sites and, to the same



▲ **Figure 2.5** – Electronic properties for bulk hematite treated with PBE+U for different potential height U_{eff}. Band centres are defined as the weighted average over the occupied DOS for the full O2p (no O2s) and Fe3d band. Spin population in units of μ_B, all other quantities in eV. The reference for the band centre is the Fermi energy of the bulk crystal, i.e. the energy of the HOMO. Experimental reference values[149,160,161] are shown as horizontal lines in the corresponding panels. Panel D shows the relative position of the levels.

Calculation(s) 7a7, 5eb

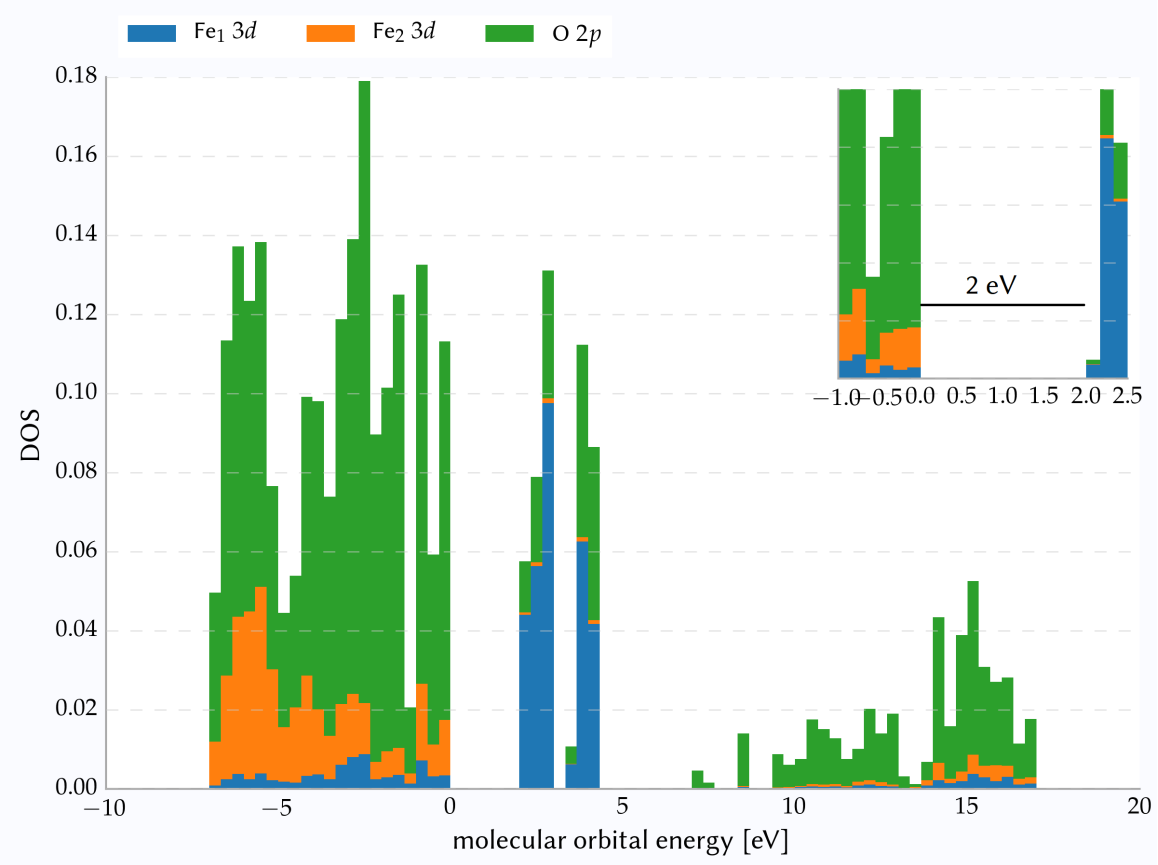


◀ **Figure 2.6** – Spins per atom site type after a geometry optimization. Calculated from spin density integration over the Wigner-Seitz cell of the given atom kind. Fe_1 is initialised with five electrons in the beta channel, while Fe_2 has five electrons in the alpha channel. As shown in Table 2.3, oxygen has paired spins for the hematite setup. Simulation cell with 270 atoms was constrained to GGA+U dimensions,[81] as done in reference literature[149] to facilitate comparison.

Calculation(s) 42e

▲ **Figure 2.7** – Projected α density of states for the three atom kinds for the PBE functional. Beta spin iron is denoted as Fe_1, alpha spin iron as Fe_2. Only 3d (iron) and 2p orbitals (oxygen) contribute to this histogram. Orbital energies are shifted such that the Fermi energy is at zero molecular orbital energy. The inset shows the region around the band gap at the Γ point. While both histograms are normalised, the bins are different for the main graph and the inset. Calculations done with DZVP-MOLOPT-SR-GTH basis set. Top: 270 atoms (3×3×1 unit cells). Bottom: 120 atoms (2×2×1 unit cells).

Calculation(s) 42e, af9



▲ **Figure 2.8** – Projected α density of states for the three atom kinds for the HSE06(12%) functional. Beta spin iron is denoted as Fe_1, alpha spin iron as Fe_2. Only 3d (iron) and 2p orbitals (oxygen) contribute to this histogram. Orbital energies are shifted such that the Fermi energy is at zero molecular orbital energy. The inset shows the region around the band gap at the Γ point. While both histograms are normalised, the bins are different for the main graph and the inset. Calculations done with DZVP-MOLOPT-SR-GTH basis set. 270 atoms (3×3×1 unit cells).

Calculation(s) d9b

extent, the spin down electrons are moved from the oxygen sites towards the iron sites. All this keeps the antiferromagnetic spin pattern unchanged. Depleting the oxygen 2p then leads to an increase of the band centre energy in absolute values*. The rapid change is explained with a comparably high DOS for oxygen 2p right next to the HOMO. As the DOS for iron at the LUMO is much more spread out, the decrease in that band centre energy is much more moderate. In any case, the relative order of the two band centres is maintained, although the band centres approach each other and, eventually, may cross each other, which would then alter the classification of hematite. So there is a limit for U_{eff} in terms of band ordering. The optimal value for U_{eff} judging by the RMSE of the band

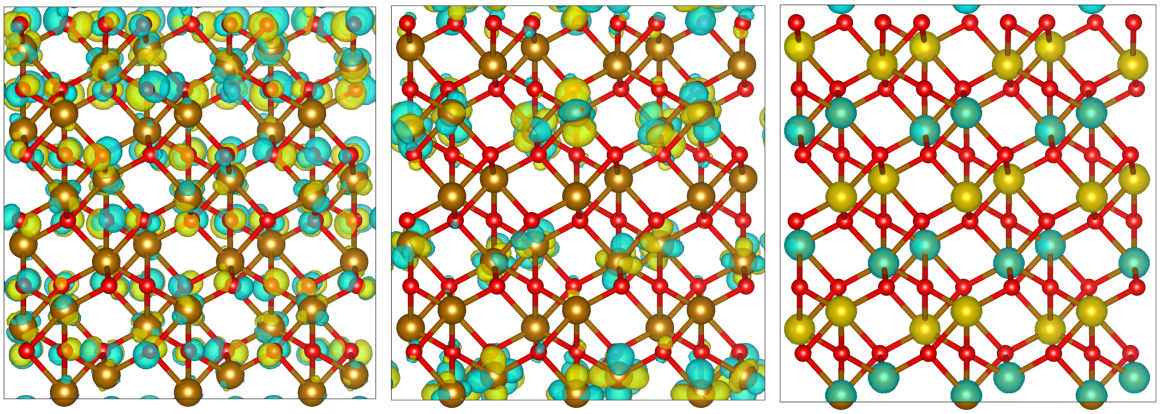
* In this discussion, the band centre energies are treated as positive values for easier comparison to experiments. Physically speaking, these energies are binding energies, and, therefore, negative and relative to the Fermi energy.

positions is around 4 V. Considering the difference in the two energies, 7 V is much better. Because of this ambiguity, the Hubbard U approach has not been considered further in this work for the calculations of the interface structure.

In literature, different values for U for the same material have been given,[80,81,115,123] some of them significantly below the others,[123] which has been attributed to implementation details.[115] While the finite size of the reference system may play a role as well, the qualitative picture is expected to stay the same given the extent to which the hybrid functional calculations have been subject to finite size effects.

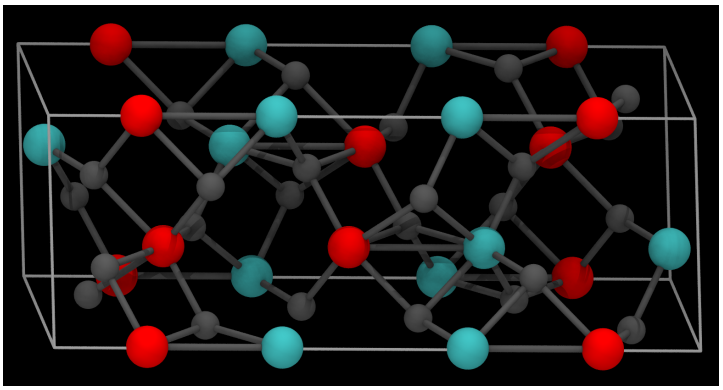
A $3 \times 3 \times 1$ setup of the 30 atom hematite supercell (calculation 689) was analysed using GTH pseudopotentials and the PBE functional. Figure 2.6 shows the distribution of the spin population per atom kind. As illustrated in Table 2.3, the expected spin configuration has two kinds of iron that carry $\pm 5\,\mu_B$ while the oxygen atoms carry zero net spin. Therefore, the simulation has been initialised with this configuration. During the SCF calculations and the geometry optimization, the spin occupation changed. If one applies Mulliken population analysis to the spin density, the absolute value for the spin for iron is $3.83\,\mu_B$. As the reference literature[149] uses Wigner-Seitz fitting, the data in Table 2.1 does so, as well. With this method, the simulation gives a value of $(3.67 \pm 0.05)\,\mu_B$, which is 2 % higher than the reference value. The numerical error from the Wigner-Seitz fitting method is in the order of $10^{-6}\mu_B$ per atom site and, therefore, negligible.

Figure 2.6 shows the spin population after a geometry optimization. While the spin population is expected to be $\pm 5\,\mu_B$ on each iron, the energy minimization shows that for PBE a partial matching of the electron spins is favourable. While the balance between positive and negative spin population is kept, that is the number of iron atoms with a net negative spin population and the number of iron atoms with a net positive spin moment is identical, the actual value of the spin moment is subject to different influential factors. One of which is the population analysis method that is employed. This effect typically is small and in the order of $0.05\,\mu_B$. However, there is one outlier at $-3.1\mu_B$ in Figure 2.6, that cannot explained by this factor alone, as using Mulliken population analysis does not change the picture this dramatically. Most likely, this is due to the unfavourable geometry which results from constraining the unit cell dimensions to match the GGA+U optimised lattice constants. As the purpose of these calculations is to show that using the setup of



▲ **Figure 2.9** – HOMO (left), LUMO (centre), and spin density (right) for the $3 \times 3 \times 1$ super cell for the HSE06(12%) functional. The HOMO is of p nature on the oxygen atoms and of d nature on the iron atoms, while the LUMO is mostly of iron d nature. The spin density shows the antiferromagnetic spin pattern for the same configuration.

Calculation(s) e9a



◀ **Figure 2.10** – Antiferromagnetic ground state spin pattern for a hematite unit cell. Only iron atoms are coloured. Atoms of the same sign spin moment have the same colour. Oxygen atoms carry no spin moment.

atom-centred orbitals reproduces the results obtained using plane wave codes, this point has not been investigated any further.

For a $3 \times 3 \times 1$ super cell, the bulk HOMO and LUMO are visualised in Figure 2.9. In accordance with the projected density of states in Figure 2.8, the HOMO is of oxygen p nature, but has some contributions from the iron 3d orbitals. The LUMO is nearly completely of iron 3d character. Both orbitals are delocalised over all layers due to the symmetry of the crystal.

After some initial doubts,[162] it now has been widely accepted that hematite is antiferromagnetic above the Morin temperature.[43,44,71,89,162] Figure 2.10 visualises the ground state spin pattern. Therefore, the spins on the iron atoms have to be alternating on the sublattice formed by the iron atoms. For hematite, there are different spin ordering patterns. This is due to the geometry of hematite which is not of octahedral symmetry, as it does not follow an ideal corundum structure. Essentially, in the ten atom unit cell, four iron atoms

Atom	Configuration	
Fe	[Ar] $3d^6\,4s^2$	[Ar] $\frac{\uparrow\downarrow}{-2}\,\frac{\uparrow}{-1}\,\frac{\uparrow}{0}\,\frac{\uparrow}{1}\,\frac{\uparrow}{2}\,\frac{\uparrow\downarrow}{0}$
Fe^{3+}	[Ar] $3d^5$	[Ar] $\frac{\uparrow}{-2}\,\frac{\uparrow}{-1}\,\frac{\uparrow}{0}\,\frac{\uparrow}{1}\,\frac{\uparrow}{2}\,\frac{}{0}$
O	[He] $2s^2\,2p^4$	[He] $\frac{\uparrow\downarrow}{-1}\,\frac{\uparrow}{0}\,\frac{\uparrow}{1}$
O^{2-}	[He] $2s^2\,2p^6$	[He] $\frac{\uparrow\downarrow}{-1}\,\frac{\uparrow\downarrow}{0}\,\frac{\uparrow\downarrow}{1}$

▲ **Table 2.3** – Electron configurations relevant to hematite. While isolated atoms are neutral, the high electronegativity of oxygen polarises the bonds.

are forming a line with segments of two alternating lengths. It has been shown[89] that the magnetic ground state features alternating spins on iron atoms separated by the shorter of the two distances. This is why in the simulations of this work, this magnetic state has been used as initial guess for the density matrix.

The data in Figure 2.8 offers another opportunity to compare the band structure of the PBE calculations of this work using atom centred orbitals to the plane-wave calculations by inspecting the PDOS. The density of states for the Fe 3d states the O 2p states look very similar to already published data.[89,90,115,163] In all cases, there clearly is a band gap, which is delimited by both oxygen and iron states. However, the electron orbitals the energy levels in the vicinity of the band gap belong to define the insulator type. Hematite is known to be a charge-transfer insulator.[45,164,165]

The expected picture is reproduced for HSE(12%). While the overall image is the same as for PBE in Figure 2.6, the actual value of the spin population is slightly different, as shown in simulation d9b. Table 2.1 compares various properties. As one can see for the $3 \times 3 \times 1$ hexagonal cell with 270 atoms, geometrical properties like bond distances are reproduced quite well. Only the bond lengths between iron and oxygen are higher than in the reference data set. While this difference of only 0.8% is quite small, it may be the effect of a more hidden effect like a different degree of electron repulsion or reaction to the ligand field. As the deviation is going away from experimental values, this has to be guarded carefully. In order to evaluate the effect of the unit cell, another simulation, b42, has been performed—now with a $2 \times 2 \times 2$ set of unit cells, giving 120 atoms in total.

Figure 2.8 shows that HSE(12%) gives a higher band gap than using PBE, as presented in Figure 2.7. Apart from that the major features, in particular, the band ordering is essentially the same. This means that the lowest unoccupied orbitals are mostly iron 3d orbitals, while the highest occupied orbitals are the oxygen 2p orbitals. Even the relative intensi-

ties and the overall feature positions are very similar. While this is expected, as HSE is a mixture of PBE and only a small percentage of HF—typically 25%, here 12%— and, hence, the features should be similar, it puts confidence in the setup, as the band gap is reacting quickly to the addition of HF exact exchange. Moreover, the experimental band gap of 2.16 (see Table 2.2) matches quite well the majority of experimental band gap values, as shown in Table 1.2.

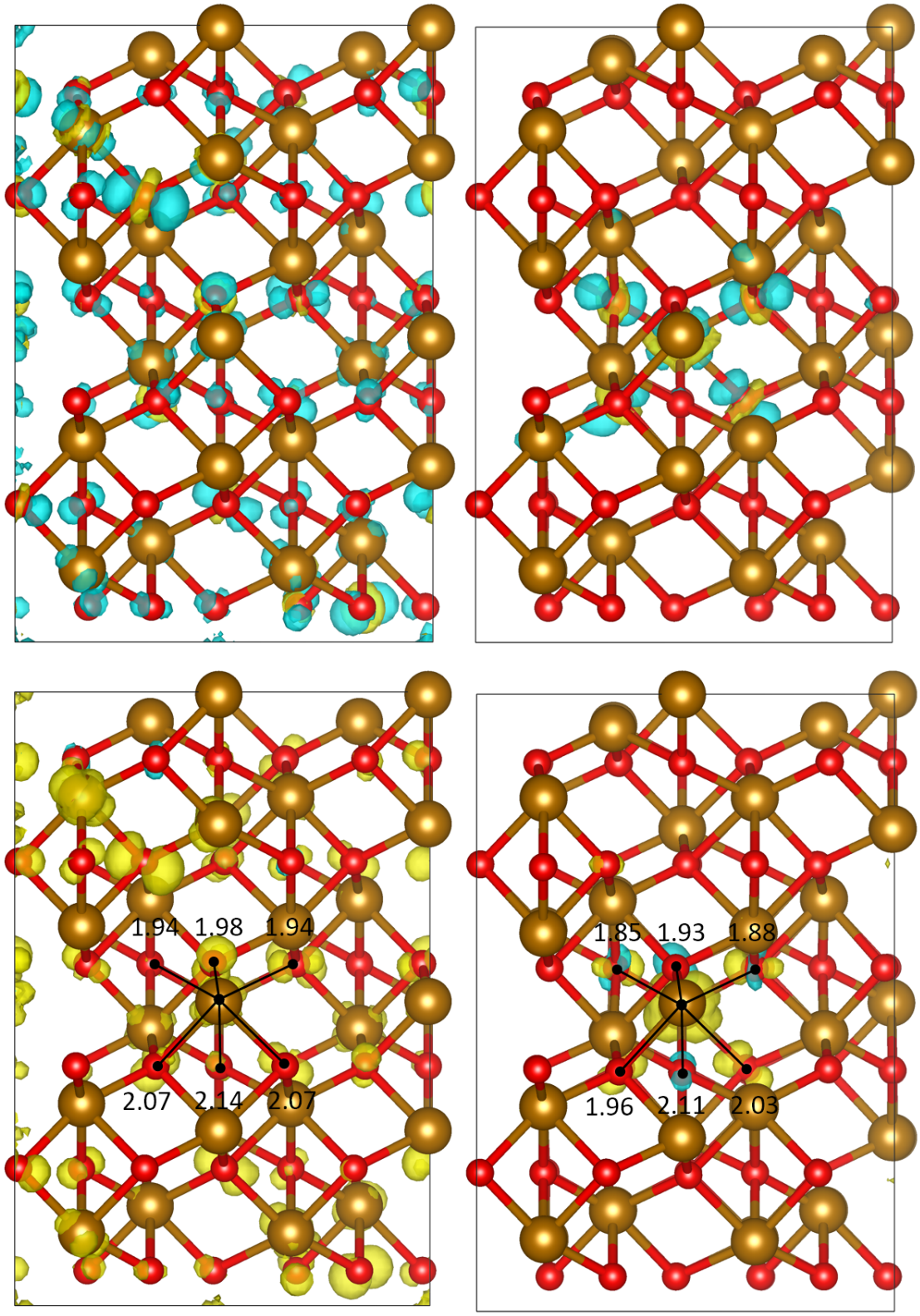
Although the majority of experimental results seem to give similar or statistically identical results for the band gap, as shown in Table 1.2, it is apparent from the gathered data that the measurements are neither accurate nor precise. For the large relative uncertainties of the measurements, this is due to the method that is employed, the so-called Tauc plot, while the outliers around 2.7 eV can be attributed to a mix-up of direct and indirect band gap. This conclusion can be drawn as some publications[47,48,60] disclose their raw measurement data and fitting curves. Those curves also show that the Tauc plot method is not employed correctly in all cases,[47] which almost certainly will introduce further errors. Another contribution that appears to be the most convincing explanation for the discrepancies is the sample preparation. Electronic properties like the band gap are highly sensitive to impurities or doping,[166–170] surface treatments,[171] or defects.[172,173] Moreover, it depends on nanostructure and surface properties,[16,51,58] on the size of the crystals[65,174,175] and crystal layer thickness.[67] Additionally, there are competing effects for absorption depending on the sample preparation,[54] which may also introduce uncertainties. This is why reviews[16] typically specify a range for the band gap.

As for the spin moment of the irons in the simulation, Table 2.2 clearly shows that both the spin population and the band gap is less than in the reference publications. As the lattice constants still are the ones from GGA+U level, it is likely that the simulation is not fully relaxed as the cell vectors could be optimized further. Therefore, another reason for this mismatch could arise from the finite resolution of the cube file data. In CP2K, this would require a higher cutoff for the mgrid. Evaluating a higher mgrid cutoff in simulation fa9 also gives slight deviations of some fitted spin moments from the average value that is observed in this simulation. As both the average value and the standard deviation of the absolute spin moment of the iron atoms remains the same, the slight variations in spin moment are most likely due to artefacts from the finite resolution of the cube files.

2.1.3 Hole

The question of the electronic structure of the bulk hole in hematite has attracted research for a long time now.[176–178] To allow comparison of the results originating from the calculation with HSE(12%) with other functionals, the electronic structure of the hole has been investigated (simulations 692, e9a, and b7b) for two unit cell sizes: $2 \times 2 \times 1$ and $3 \times 3 \times 1$. Since CP2K offers the functionality of taking out the electron from any orbital, two independent calculations have been prepared: one where the electron was taken from an iron site and one where it was taken from an oxygen site. Both runs performed a geometry optimisation for the doublet state. For the $2 \times 2 \times 1$ setup, 73.8 % (52.7 %) of the hole was located on all iron atoms for the relaxed (vertical) geometry, while the $3 \times 3 \times 1$ setup placed 76.6 % (55.0 %) on all iron atoms. In the first case, the relaxation energy was 0.90 eV while the latter one showed a larger relaxation energy of 2.49 eV.

Figure 2.11 shows the hole localisation upon nuclear relaxation. In the neutral geometry, i. e. after electronic, but no nuclear relaxation, the hole is delocalised over all oxygen sites and is mostly of p character, which matches the expectation from the HOMO in Figure 2.9, but is different from the vertical state before electronic relaxation, that is expected to follow the HOMO, as discussed for Figure 2.9. During the relaxation, the particular iron site shifts more towards the middle between the two coordinating oxygen layers. In the initial vertical configuration, the bond lengths to the coordinating oxygen atoms are 1.94, 1.94, and 1.98 Å to the closer oxygen layer and 2.07, 2.07, and 2.14 Å to the more distant layer. This asymmetry comes from the bulk geometry of hematite where each iron layer is split into two sublayers. For the relaxed configuration, one iron site becomes particular and holds the majority of the hole in d character (see below and Figure 2.11) and the bond lengths change to 1.85, 1.88, and 1.93 Å for the closer oxygen layer and 1.96, 2.03, and 2.11 Å for the more distant layer. The geometry obtained during the geometry optimisation supports the small polaron model.[177,179] For the two system sizes $2 \times 2 \times 1$ and $3 \times 3 \times 1$, one specific iron site emerge with a hole contribution of 0.71 or 0.73, respectively. Since the relaxation energy of the hole is subject to strong finite size effects,[180] the final value for this functional could only be obtained from extrapolation to infinite unit cell size. This, however, would require multiple data points for even larger systems, which is currently not feasible at this level of theory owing to the high memory requirements. For all these



▲ **Figure 2.11** – Difference in charge density (top row) and spin density (bottom row) for the vertical configuration (left column) and the relaxed configuration (right column). Isosurfaces are coloured yellow for positive and blue for negative values. Data for a $2 \times 2 \times 1$ super cell of bulk hematite. Upon nuclear relaxation, the hole localises on an iron site. The bond lengths in Angstrom between this site and the coordinating oxygen atoms (as indicated) are given in the bottom row. Differences are given by neutral volumetric data subtracted from charged volumetric data.

Calculation(s) e9a

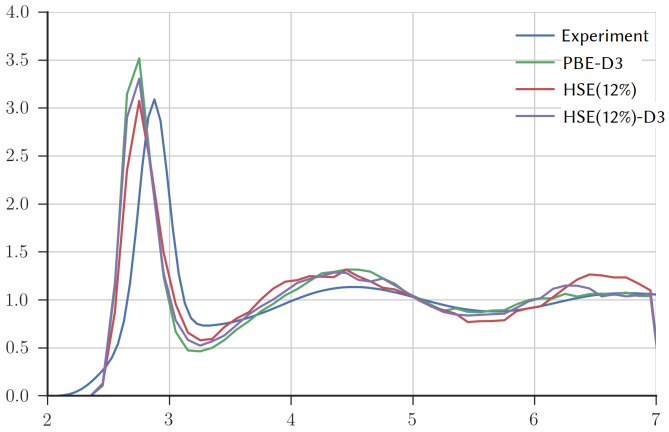
numbers, the Hirshfeld partition scheme has been applied. Previous studies[181] which did not localise the hematite hole may suffer from finite size effects because their system may not be large enough to cover the full lattice reorganisation, thereby giving rise to an intermediate geometry.

The difference in spin density in Figure 2.11 between the neutral and charged case on the neutral configuration shows that the spin moment on the oxygen sites increases, since the removal of a beta electron leaves one unpaired electron in oxygen 2p. At the same time, there are some contributions on the iron sites that have a beta majority, while the iron sites with alpha minority contribute very little, since they only have a very limited population of beta spin electrons. After relaxation, the contribution of the oxygen atoms is greatly reduced and now spatially restricted to the oxygen sites coordinating the beta iron site where the majority of the hole is located. The density shapes are d-like for the iron atom and p-like for the coordinating oxygen atoms.

When comparing these results to experimental data,[161] it is crucial to note that the vertical hole here is for the ideal zero temperature configuration where all sites in the lattice are exactly identical. Therefore, all sites in the identical configuration (which is favoured by the original electronic structure without any hole) are equivalent. Only (random) thermal motions can break this symmetry and give rise to centres which are more favourable than others to localise the hole. Therefore, experimental findings where the vertical hole is found to be localised may be influenced by their finite temperature and the resulting lattice vibrations. For future work, it may be interesting to observe whether configurations similar to the relaxed hole structure discussed above can be found in the thermal fluctuations of neutral bulk hematite.

2.2 Water

While DFT liquid water has been subject to extensive studies in literature, we evaluated the performance of the HSE06(12%) functional, which we found has best performance for hematite because it had not been tested for liquid water in literature before. All simulations for investigating the properties of bulk water with a given DFT method have been carried out with a cubic box of 32 water molecules at 300 K. After an equilibration period of 3.5 ps, the analysed trajectory had a duration of 5 ps to 10 ps. This is a typical setup for evaluating



◀ **Figure 2.12** – Radial distribution function for oxygen-oxygen for bulk water with different functionals compared to experiment[184] (labeled exp).

Calculation(s) 283

the capabilities of the chosen method for bulk water.[182]

Getting dispersion properties right is particularly important for water, as the dispersion energies determine both geometric properties and dissociation energies.[183] It has been shown,[183] that dissociation energies for water clusters typically are ordered in a reverse way for common GGA and hybrid functionals when compared to MP2 simulation results, which gives rise to the need for dispersion correction.

2.2.1 Geometry

Initially, the RDF of water was calculated (see Figure 2.12) using DFT-D3[146] as dispersion correction. As this method employs two free parameters which depend on the functional used, namely the scaling factor for higher order contributions s_8 and the damping $s_{r,6}$, and there are no values available for HSE(12%), the PBE0 settings have been employed, as this functional is most similar to HSE(12%). The resulting radial distribution function features overestimated peaks. This effect is much larger than only the well-documented over-structuring of the water,[185–187] and affects all possible pairs of atoms in bulk water (O···O, O···H, and H···H). The peak positions are the same, though, which gives rise to the belief that the stronger shaped peaks are a result of too strong dispersion correction, which may come from the usage of PBE0 values for a different functional.

The authors of the DFT-D3 method have improved[137] upon their work by changing the potential damping to Becke-Johnson,[143] calling the resulting method DFT-D3(BJ). While they have shown that the form of damping itself does not have a high influence on the overall quality of the results, this form of potential has been introduced in order to

reduce ambiguity in definition. This is an important factor that should enable transferring the parameters across functional families. As the parameters come from benchmarking against highly expensive calculation results at a high level of theory, having this feature of transferability established is desirable.

DFT-D3(BJ) has three parameters, that is one more than DFT-D3 alone. In this case, s_8 is left unchanged, but the physical meaning of $s_{r,6}$ now is transferred to a_1, while a_2 determines the BJ damping. In literature, there are values[146,188] for the functional dependent triple (s_8, a_1, a_2) for PBE0 (1.2177, 0.4145, 4.8593) and for HF (0.9171, 0.3385, 2.8830) which can be used to interpolate these values for HSE(12%) weighting the HF values with 12 % and the PBE0 components by 88 %. This gives (1.1816, 0.4054, 4.6221) for the functional used for the hematite calculations. Although there have been simulations performed with HSE06-D3, the parameters have not been published*.

Even with dispersion correction, HSE06 has an tendency for over-binding, which can be partially mitigated by adding three-body terms to D3.[189]

All variants of dispersion correction studied as comparison give reasonable agreement with the experimental pair distribution function for the oxygen atoms in bulk water.[184] However, the main differences arise for the first peak of the radial distribution function, which typically is the most prominent one in this distribution over a wide range of pressure and temperature.[190] Experimentally, the peak height decreases with increasing temperature. This effect is larger in the temperature range near room temperature than it is near the boiling point at atmospheric pressure. Therefore, slight differences in peak height are typically acceptable, in particular, as the experimental measurement of the radial distribution function requires extensive analysis of indirect observations. Experimental heights for the first peak height for room temperature and atmospheric pressure are highly different.[190,191] While the RDF does not converge completely within 10 ps it is expected that the first peak converges after 2 ps.[182]

* Most likely, this happened by accident. On the website of the group of S. Grimme, there is a tool available for download that is supposed to generate the parameters from reference data. This program is shipped with HSE06 parameters. But as there is no official statement of the group regarding these parameters, they have not been used in the simulation setup.

2.3 Conclusion

It has been shown that HSE06(12%) is able to reproduce both geometry and electronic properties as found in experimental studies. With the caveat relating to the electronic properties, the traditional approach of GGA+U has shown to be applicable for geometry sampling.

With having shown that the functional of choice is able to describe the relevant quantities of both bulk hematite and bulk water, it is now possible to combine these two subsystems to form an interface. The next chapter describes how this has been done and verified.

Chapter 3

Interfaces

Understanding the charge transfer processes at the hematite-water interface is of particular importance for optimizing photoelectrochemical water splitting,[4,16] for understanding of crystallisation behaviour,[6] and for surface reactivity[24–28] including hole properties[192–195] and acidity.[22,23]

With the confirmation from the previous chapter that the chosen functional can describe the electronic properties of interest for both subsystems bulk hematite and bulk water, there is the expectation that the functional can describe the interface reasonably well. In the following, the gained physical insight is compared to experimental results wherever possible to validate the overall approach. At first, the computational approach in particular in relation to the classical pre-equilibration is documented.

3.1 Setup

Investigating the interface of solid hematite and liquid water requires a careful setup procedure that generates system snapshots that are well-equilibrated but not too expensive to obtain. In particular, pressure equilibration is not feasible at AIMD level. Moreover, all physical properties of the underlying system have to be included as far as possible. For hematite, this means selecting the proper surface cut,[36] to reproduce the correct antiferromagnetic ground state,[89] to recover the surface protonation state at the point of zero charge (PZC),[29] and to estimate the minimum slab thickness from experimental information.[196] Moreover, the total system should carry no net spin moment or net electron charge and observe the symmetry of the underlying crystal. This section explains the setup

procedure that has been employed in all interface calculations.

3.1.1 Procedure

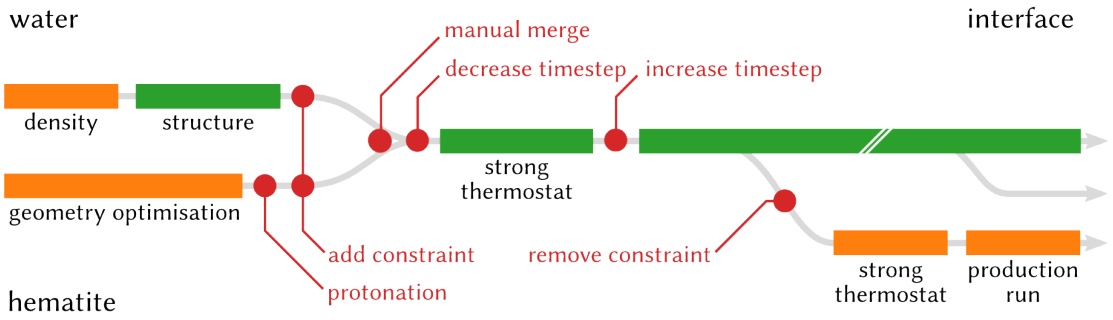
The procedure to follow in order to create a reasonably equilibrated snapshot from an idea how the interface should look like is quite involved. The main problem resides in the timescales that are accessible with hybrid functional DFT. The rotational autocorrelation time of water is around 10 ps which means that the calculation of a RDF as one of the major representatives of the quality of a calculation requires some 10 ps trajectory. This, however, only applies to the first peak of the O···O RDF, as it is the most prominent one and the same statistical error applies to all points on the RDF. Therefore, the later peaks are more likely to be obscured by sampling noise.

Using an implicit solvent model[197] has been considered, but set aside since the conformational information about the water molecules near the surface is required explicitly.

Moreover, any kind of pressure equilibration will take much more than the 10 ps range that is readily accessible.[198] As pressure coupling is known to give heavy fluctuations in instantaneous pressure, it is even more challenging to measure the progress of equilibration, as the target value is only reasonably well defined in the ensemble average. All this also applies to thermal equilibration, but it is considered to be less intrusive to perform strong temperature coupling than to do strong pressure coupling.

This is why a multi-stage system preparation scheme has been developed that combines both DFT calculations and classical MD to come up with a reasonable initial conformation. The overall idea is to prepare the hematite slab first, then constrain it, add rigid water and perform a classical MD. As an example, the following sections will describe the setup for the (001) surface, a $2 \times 2 \times 1$ super cell with roughly 30 Å water and the OH termination, although the same procedure has been used for the iron termination in calculations 590, 129, f8d, acc, eb1, as well.

Figure 3.1 shows the overall idea of the equilibration procedure including the drawing of initial structures from the classical MD calculation.



▲ **Figure 3.1** – Equilibration procedure and branched MD for building the interface. Orange bars represent ab initio calculations, while green bars denote the computationally less demanding classical MD simulations. The branch without any coloured block is the same as the other branch leading to a production run but starts from a different initial conformation of the classical MD.

3.1.1.1 Hematite Slab

The hematite bulk structure known from experiment is optimised with the HSE(12%)-ADMM-D3(BJ) functional (see calculation 893) which gives both the minimum energy cell size and lattice vectors, but also the minimum energy conformation given these lattice vectors. This is necessary, as there are slight differences in the geometry of hematite as it is found in nature and when it is described using our specific functional. Now taking directly the geometry from experiment would lead to two main problems. The first of which is that—due to the strong interaction of the atoms in bulk hematite—the position of the atoms would correspond to a rather high energy in the calculations, so one would observe the relaxation process in the MD. Unfortunately, the time scale of which is not known and this relaxation would require rather strong temperature coupling in order to remove the energy from the system to reach the target temperature. The second problem is that this hematite bulk structure has to be combined with the water interface later on. This requires pre-equilibrated water, as well, and this pre-equilibration requires exact knowledge* of the periodic simulation box. Even small perturbations would distort the hydrogen-bond network of water and, therefore, require a new equilibration of the water part. Both problems can be overcome by performing an extensive search for the minimum energy conformation of hematite given the specific functional.

* More precisely, it is required to know the lattice vectors in the water layer plane, as the actual interface is subject to (unpredictable) relaxation in any case.

3.1.1.2 Water Layer

As the protocol requires classical MD later in the process, the equilibration of the water layer has to be done using a classical water model (for the following, see calculations 4c8, ea6, b0f, 65b, c92). This work uses TIP3P[199] for the calculations. Now, as stated before, the pressure equilibration of the final interface system is not accessible with current computational means. Therefore, two equilibrations of the same water box have been performed, once with the DFTMD and once with classical MD. The DFT calculation was in NPTF ensemble which gave the density of water using the target functional. This density value ($1.0029\,\mathrm{g/cm^3}$) then defined the volume for the classical NVT calculation. The structure from this classical calculation then is used for building the interface. Two of the three lattice vectors are given by the hematite bulk optimization. If $\hat{\mathbf{e}}_\mathbf{z}$ also is the normal vector of the surface, then the lattice vectors in the x and y direction of the non-rectangular unit cell are defined by the hematite calculation. Moreover, the direction of the third lattice vector is implicitly defined by the goal to build a stacked water-hematite layer structure. So the only remaining degree of freedom is the length of the third lattice vector. This length is equal to the thickness of the water layer that is desired for the final setup.

3.1.1.3 Interface

While experiments[31,160] hint towards significant surface relaxation of the structure, it is more useful not to include these relaxations in the simulation setup. If one would do so, there would be no possibility to compare the simulation results to the relaxation structure they found, as then there would be no driven relaxation inherent to the system. Therefore, as little information as possible is used for designing the actual surface.

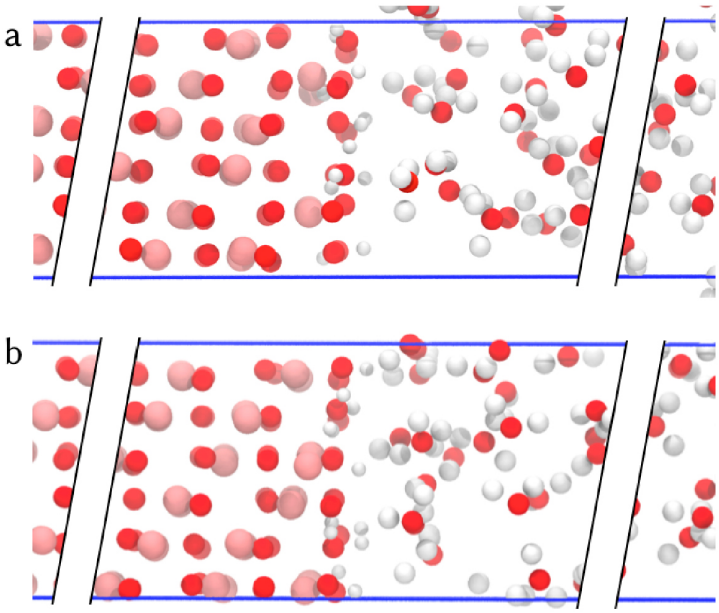
The regular crystal structure is truncated at the surface, followed by protonation of the oxygen atoms at the bare surface. This is required to recover the neutral experimental surface.[31] To this end, the net charge of one non-stoichiometric oxygen layer has to be compensated for by protonation. Since the pristine surface termination does not change upon solvatisation,[31] this already is the protonation scheme that is expected to be the result from the wetting process. The O···H distance is initialised to the value of the water O···H bond length. To reproduce the point of zero (surface) charge (PZC) surface protonation as determined experimentally,[2] all surface oxygen atoms get singly protonated and

all protons are placed directly on top of the oxygen atoms. While this is almost certainly not the ensemble average and generates an artificial dipole, it maximises the volume that may be consumed by the interface. This is reasonable, as in general, there are less deep local minima in conformation space[200] for systems with low local density as there are for systems with high local density, so the optimisation process happening during the MD has a much smaller chance to get stuck in a local minimum of the conformation. It will be shown later on that the equilibration process for the orientation of the OH bonds is fast enough to produce a converged geometry, which in turn produces an equilibrated dipole.

Now the whole periodic water box that has been equilibrated separately from the rest of the system is considered to be one snapshot of the water layer in the interface case (see calculation 7b0). This reflects neither solvation effects[201] nor confinement effects,[202] which is why further equilibration is necessary. This further equilibration is done as follows. When joining the different systems, there is a small gap between the protonated hematite surface and the bulk water set. This gap initially is set to a the O$\cdots$H distance again, but is allowed to change during the equilibration. While a region of low density is expected at the surface,[202] the density distribution is not known beforehand.

For this setup, a force field[203] can be used to describe the interaction between the iron atoms and the water oxygen atoms.[32] For the interaction of the water molecules, the regular TIP3P model is applied, where the LJ term only acts on the oxygen atoms, but the charge is distributed over all three atoms. While force fields typically use a lower partial charge for the atom centres instead of the actual charge, in order to improve interaction energies, which otherwise would be overestimated due to the point charge nature of force fields, this force field uses the regular charge that is found on the atom sites from charge fitting. In this case, this means Fe^{3+} and O^{2-}. As in this setup, the hematite atoms are kept fixed with the exception of the surface protons, only the interaction term between the surface oxygen atoms and the surface hydrogen atom remain to be defined. Any other interaction term would be completely determined by the constrained structure, which would yield a constant energy contribution. As the absolute total energy is meaningless in classical MD simulations, this simplification has no impact on the accuracy of the system.

Now this system is still far from both the minimum energy conformation and the ensemble average. Therefore, a classical MD is performed. Typically, one would apply the



◀ **Figure 3.2** – Snapshots of the c-cut setup. Alternating layers of the iron sites have opposite majority spin moment. The hematite layer is about 14 Å thick, while the water layer spans about 35 Å. Panel a shows the iron termination, while panel b shows the oxygen termination. The blue box represents the fully periodic simulation cell. Image reused from my CC BY publication.[206]

Calculation(s) e9a

Nosé-Hoover[204,205] thermostat in this case, as it gives weak coupling of the system and the heat bath. For this setup, it is necessary to use strong coupling in the beginning, in order to allow for large energy flows in and out of the system that are required at this early stage of equilibration. Using a CSVR thermostat with a very short running average window width of a few 10 fs works best here. Also, one has to reduce the time step of the simulation from 0.5 fs (which would be typical for this setup) to 0.1 fs. This greatly reduces the stress on the thermostat which otherwise would be responsible for ensuring system stability. As this is a temporary measure for a few hundreds of femtoseconds only, the impact of a smaller time step on the performance of the simulation is of little importance.

After this calculation, the thermostat has been switched to Nosé-Hoover chains, and the time step is set to 0.5 fs. This is the equilibration run which also can be used for drawing snapshots from the trajectory to get initial snapshots for the AIMD calculation. In this case, snapshots are drawn after 50 ns of simulation time in steps of 50 ns.

As until this step the hematite is still constrained, a consecutive equilibration using AIMD is still necessary. In this last stage, there are no constraints, so the hematite has to be heated up to target temperature of 300 K. This may be done using comparably strong temperature coupling with CSVR—just as before. After 1 ps, this can be relaxed to Nosé-Hoover coupling again. After another picosecond, the system is sufficiently equilibrated. Apparently, the most sensitive yet easily to determine property is the volume, so once the volume equilibration has been reached, the production runs can be started.

Figure 3.33 and Figure 3.2 show snapshots of the hematite / liquid water systems under investigation in this work. The bulk crystal structure is identical in all the setups, they only differ in the crystal cut exposed to liquid water. All the setups have been prepared using the same procedure outlined above. For each crystal surface, different configurations of the hematite termination or protonation state have been considered. For the c-cut in Figure 3.2, there are two experimentally observed terminations,[32] the so-called *iron termination* and the *oxygen termination*. Coming from the regular hematite bulk structure, the oxygen termination can be regarded as bulk hematite cut right after an oxygen layer. The oxygen termination has the same structure with the exception of one of the two iron sublayers just beneath the surface missing from bulk hematite structure. In both cases, experimental evidence points towards all oxygen atoms at the surface to be singly protonated.[207] This matches the number of protons required to ensure total charge neutrality for the simulation system.

3.1.2 pH

While the system is simulated at the point of zero charge (PZC), modelling the pH in a simulation setup is tricky, but necessary.[208,209] In particular, the limited size of the system has strong implications. Moreover, the way to arrive at an equilibrated structural description of the hematite-water interface is non-trivial. This section describes the procedure developed to obtain a well-defined and reproducible description of the interface system.

The pH value is defined as

$$\mathrm{pH} \equiv -\lg(a(\mathrm{H}^+)) \simeq -\lg(c(\mathrm{H}^+)) \tag{3.1}$$

where c is the concentration and a is the thermodynamic activity of a chemical species which in turn is defined as

$$a(S) \equiv \exp\left(\frac{\mu_S - \mu_S^{\ominus}}{RT}\right) \tag{3.2}$$

where μ_S is the chemical potential of species S and $\mu_S^{\ominus}$ denotes the chemical potential at standard conditions. The activity can be seen a coefficient showing the extent of interactions of particles of one kind. The approximation in eqn. 3.1 holds only for very dilute solutions. Solvated H^+ typically is described as Eigen complex (H_3O^+) or Zundel complex ($H_5O_2^+$). In both cases, the structure fluctuates strongly. The typical autoprotonation

process is

$$2\,H_2O \rightleftharpoons H_3O^+ + OH^- \tag{3.3}$$

which happens at a rate K which is defined by the law of mass action

$$K := \prod_i a_i^{\nu_i} = \frac{a(H_3O^+)a(OH^-)}{a^2(H_2O)} \tag{3.4}$$

where ν_i are the stoichiometric coefficients. As a very dilute solution of water and its autoprotonation products has a chemical potential close to the one of water alone, the difference $\mu_S - \mu_S^{\ominus}$ is close to zero and, therefore, the activity is approximately one. Therefore, one can approximate eqn. 3.4 with

$$K \simeq a(H_3O^+)a(OH^-) \tag{3.5}$$

Unfortunately, the same trick cannot be applied again to deal with the activities of the other products, as their chemical potential in the given chemical environment is not known for the pure solution.

Moreover, it has been shown experimentally[210] that the activity of water undergoes a significant change upon confinement. Above the micrometer scale, the approximation that the water activity is one holds true. As below the micrometer scale the activity drops exponentially, this approximation does not hold in the regime of AIMD calculations, where the water layer has a typical thickness of 30 Å. Using their empirical formula for describing the water activity a_w

$$\log a_w = \frac{-0.0921 \cdot 10^{-8}\,\mathrm{m}}{d} \tag{3.6}$$

where d is the diameter of the confined structure*. For $d = 30$ Å, one gets $a_w \simeq 0.5$.

3.2 Computational Details

When preparing simulations or planning a roadmap for a research project, knowing the scaling behaviour of the setup is important as it limits the system sizes and trajectory lengths that are accessible at reasonable computational costs. More importantly, these limits highlight those points that are the key for tuning the performance of the overall method.

* Strictly speaking, this assumes a capillary structure, but allegedly depends on the confining material anyway. Therefore, this can act as a orientation of the activity behaviour only rather than a quantitatively suitable description.

For example, tuning costly reference calculations that are only to be done in the beginning may less effective than tuning an already cheap calculation that has to be done over and over again.

The scaling behaviour gives information about how the throughput in terms of SPC depends on the number of cores or nodes available for the simulation run. For some components of a MD step in CP2K, a nearly ideal, i. e. a linear scaling is expected.[211] However, this highly depends on the actual simulation method used and on the sparsity of the employed matrices which in turn depends on the phase of the material that is simulated. Moreover, there are some parts of the simulation which are not easily parallelisable at all. These parts gain importance for calculations with larger problems and larger computational resources, because the number of nodes that cannot be employed for any of these calculations increases as well.

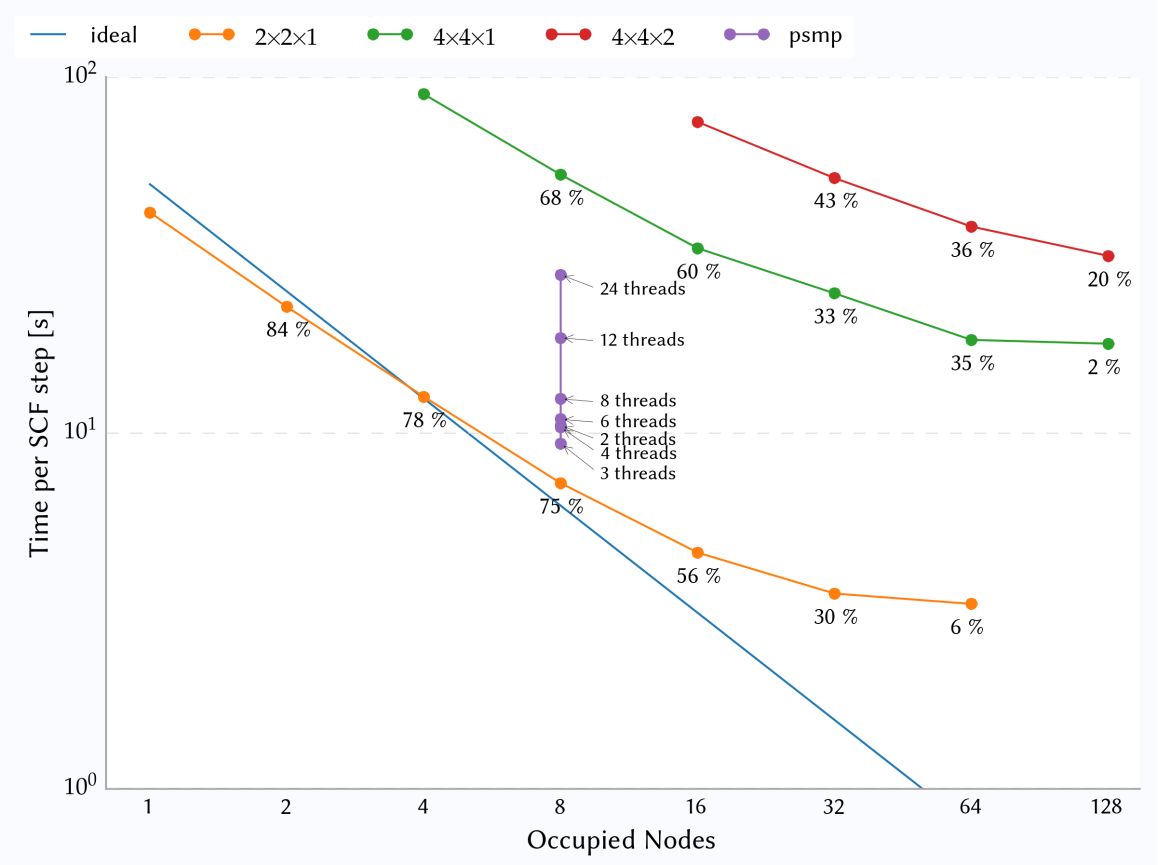
A typical unit of the scaling behaviour is the so-called speed-up s which is based on a normalised throughput $T(n)$ where n is the number of cores or nodes

$$s(n) = \frac{T(n)}{T(1)} \tag{3.7}$$

Typically, $s(n)$ is a linear function of n for small n and has a maximum at some point. The scaling behavior also depends on the architecture of the computer the software is running on and the compile-time options. Therefore, comparing this scaling across clusters or program versions is not generally possible.

The ARCHER supercomputer which has been used for most of the calculations in this work, is of Cray XC30 architecture with two 12-core E5-2697 v2 (Ivy Bridge) processors per node. Figure 3.3 shows the scaling behaviour for representative $4 \times 4 \times 1$ and $4 \times 4 \times 2$ interface models. The setup consists of a $4 \times 4 \times 1$ hexagonal super cell of hematite (120 atoms) with a liquid phase consisting of 384 molecules which has been brought in contact with hematite in the direction of the **c** lattice vector. The water has been pre-equilibrated in a cubic box using the same functional that has been used for the hematite super cell (HSE(12%)*-ADMM-D3(BJ)). After wrapping the water into the hexagonal simulation box of $20.164 \times 20.164 \times 46.482$ Å^3, a geometry optimisation has been performed until the energy change between consecutive optimisation steps fell below 0.01 Hartree. The volume

* This notation denotes the HSE06 functional with the share of HFX reduced from 25% to 12%, motivated by previous studies.[149]



▲ **Figure 3.3** – Scaling behaviour of CP2K over number of nodes (24 cores each). *popt* and *psmp* refer to the CP2K parallelisation schemes (see text). Three systems were evaluated: a $2 \times 2 \times 1$ super cell of hematite with 96 water molecules (green), a $4 \times 4 \times 1$ super cell of hematite with 384 water molecules (red) and a $4 \times 4 \times 2$ super cell with 384 water molecules (purple). *psmp* was benchmarked for the smaller system. Annotations for *psmp* show the number of SMP threads per MPI process. Annotations for the other curves show the marginal efficiency of additional nodes when compared to the previous data point.

Calculation(s) 520, 253, fd5

that has been added to the simulation box in order to accommodate for the water has been calculated from the density of the pre-equilibrated water box.

Throughout this work, three main methods have been employed or evaluated: PBE, PBE+U and HSE with ADMM. The calculations for HSE are so expensive that only the ADMM[134] renders them feasible. A test run of the interface system* showed that on 32 nodes† the initial calculation of the wavefunction does not converge even though the correct spin population was hinted in the input file by using CP2K_INPUT/FORCE_EVAL/SUBSYS-/KIND/BS.

The data from simulations 6f2 and 520 shows that for 32 nodes, PBE+U is about 82 %

* 2040 electrons in the overall hematite-water mixed interface. † The calculations have been done on ARCHER which is a Cray XC30 machine where each node consists of two 2.7 GHz, 12-core E5-2697 v1 (Ivy bridge) series processors with 64 GB memory.

faster than HSE and that PBE is about 7 % faster than PBE+U.

3.2.1 Scaling Behaviour

Figure 3.3 shows the scaling behaviour for a simple interface. As long as the overall geometry and the relative volume of the liquid and solid interface remain the same, the actual structure of the sample interface does not matter. Therefore, the simple interface used in this setup has to real physical meaning, as the surface termination has been ignored. The setup consists of a $2 \times 2 \times 1$ hexagonal super cell of hematite (120 atoms) which is brought into contact with a liquid phase consisting of 96 molecules which has been attached to the hematite in the direction of the **c** lattice vector. The water has been pre-equilibrated in a cubic box using the same functional that has been used for the hematite super cells (HSE(12%)-ADMM-D3(BJ)). After wrapping the water into the hexagonal simulation box of $10.082 \times 10.082 \times 46.482\,\text{Å}^3$, a geometry optimisation has been performed (simulation 448) until the energy change between consecutive optimisation steps fell below 0.01 Hartree. The volume that has been added to the simulation box in order to accommodate for the water has been calculated from the density of the pre-equilibrated water box.

From the scaling information, it is evident that *psmp* is less efficient than *popt*. Even though it is not shown in Figure 3.3, this is also the case for the $4 \times 4 \times 1$ system. For deciding which setup is a reasonable compromise between required resources and duration of the overall simulation run, the marginal efficiency is useful. Given the duration t of a single SCF step, the marginal efficiency $\Delta\eta$ of two setups can be calculated from the number of nodes involved N as follows

$$\Delta\eta(N) := \frac{t(N/2)}{t(N)} - 1 \tag{3.8}$$

A value of 0 means that doubling the computational resources yields no benefit at all, a value of one is ideal scaling. Negative values denote decreased performance and values above 1 are valid for superlinear scaling. Assuming a desired limit of 50 % marginal efficiency, for the smaller system the scaling behaviour is acceptable for up to roughly 20 nodes (480 cores) while the larger system scales up to 25 nodes (600 cores).

For planning simulations, it is helpful to estimate the overall costs in terms of both the time it takes to actually perform the simulations and the computational budget that has been made available at the high performance computing facilities. If the duration of a

Setup	Nodes	kAU	Duration [d]
$2 \times 2 \times 1$	1	124	14.4
$2 \times 2 \times 1$	32	339	1.2
$4 \times 4 \times 1$	4	1074	31
$4 \times 4 \times 1$	16	1586	6.3

▲ **Table 3.1** – Simulation costs on ARCHER in their budget units kAU and the simulation duration for 1 ps of trajectory for a system of $2 \times 2 \times 1$ supercells of hematite with 96 waters in total (1968 electrons) or the quadruple of this system (7872 electrons) and HSE(12 %)-ADMM-GTH-D3(BJ).

Calculation(s) 520, 253, fd5

single SCF step is given as q (in seconds) and the number of steps necessary for completion of a single SCF cycle is given as n while N denotes the number of nodes required then for the simulation time d the required budget units U can be calculated from the time step Δt and the conversion factor c which gives the number of budget units per node hour

$$U = \frac{q}{3600} n N \frac{d}{\Delta t} c \tag{3.9}$$

With the typical values $n = 15, \Delta t = 0.5\,\text{fs}$ and the conversion factor specific to ARCHER $c = 0.36$, this can be simplified to

$$U(d = 1\,\text{ps}) = 3qN \tag{3.10}$$

The results for ARCHER are given in Table 3.1.

3.3 Neutral Interfaces

3.3.1 Hematite c-cut (001)

3.3.1.1 Oxygen Termination

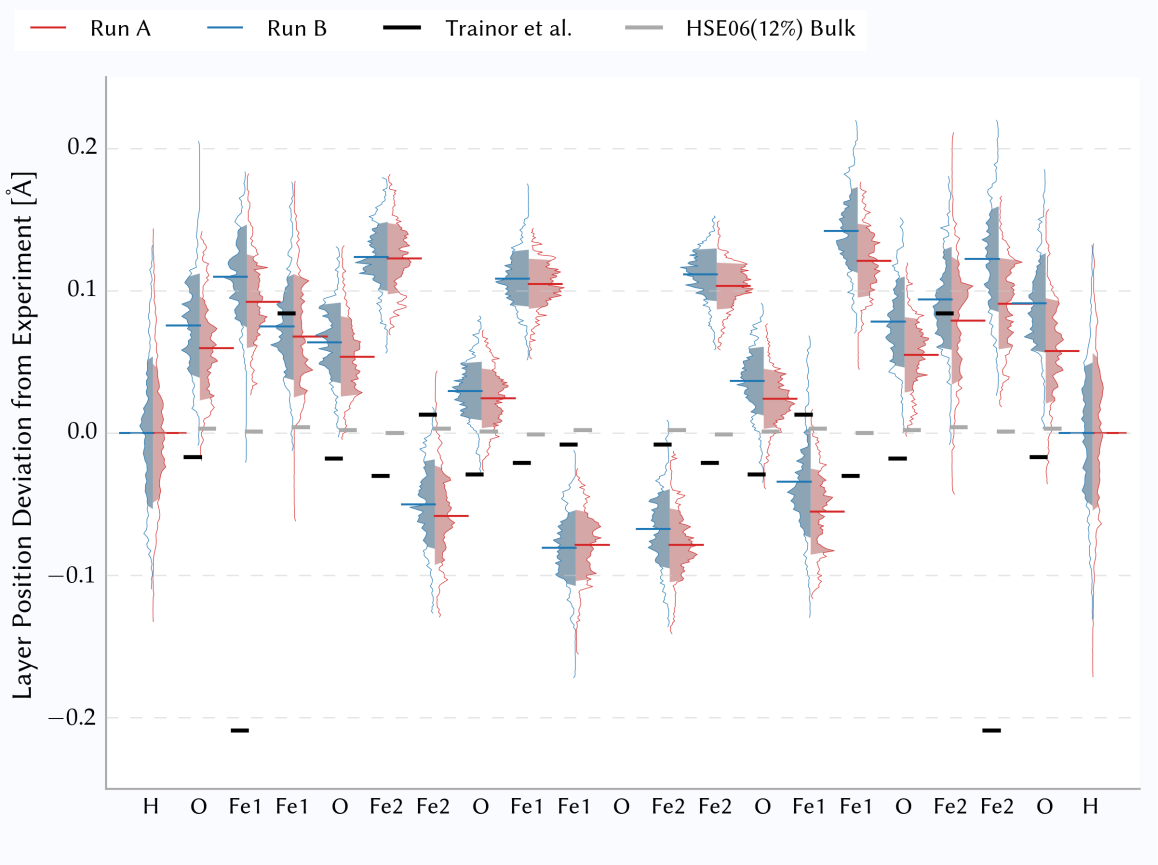
One of the most prominent geometric or structural properties is the interlayer spacing of hematite near the surface. There are numerous experimental measurements[30,31,212] of this property as it is directly accessible by CTR experiments in particular and X-ray methods in general. This methodology, however, does not allow to resolve hydrogen atoms as their scattering cross section is too small. Still, the experimental approaches are similar starting with a clean and ideally step free (001) surface. Then, the hematite slab is brought into contact with water, followed by immediate *in situ* measurements of the system. Some experimentalists see a significant ageing effect[17] which means that the time between sample

preparation and actual measurement may have substantial impact on the measurement results. However, not all papers state the time delay between these two steps, so it is hard to quantify the ageing impact on the comparison of the different sources.

Figure 3.4 compares the results of this work to one of the most detailed experimental studies available so far.[31] In this reference, data about the interlayer spacing for iron and oxygen atoms down to the fourth oxygen layer is available. This is ideal for comparison, as the middle layer of oxygen the hematite slab in the simulations of this work is exactly the fourth one from either side of the slab. The results of this work have been compared to their reference data by building the position of the layers along the third axis of the hematite crystal by summation of the interlayer distances. The position of the innermost oxygen layer (the fourth one from the surface) has been set to zero in both cases. This alignment causes the innermost oxygen layer to have zero variance in our setup and a position ideally matching the experimental data. Essentially, no comparison is possible for this innermost oxygen layer and the outermost hydrogen termination. At least for the outermost hydrogen termination, a histogram of the layer positions is available in Figure 3.4.

As the ensemble is NPTF, the simulation cell changes of the course of the calculation. This requires a slightly more complicated evaluation of the layer positions than just taking the centre of mass along the surface normal vector. The layer positions are calculated by representing the cartesian coordinates of the atoms in question in fractional coordinates of the simulation box. Then, the third component of these positional vectors is averaged over all atoms of a given layer. Finally, this value is transformed back into cartesian coordinates for the z component only.

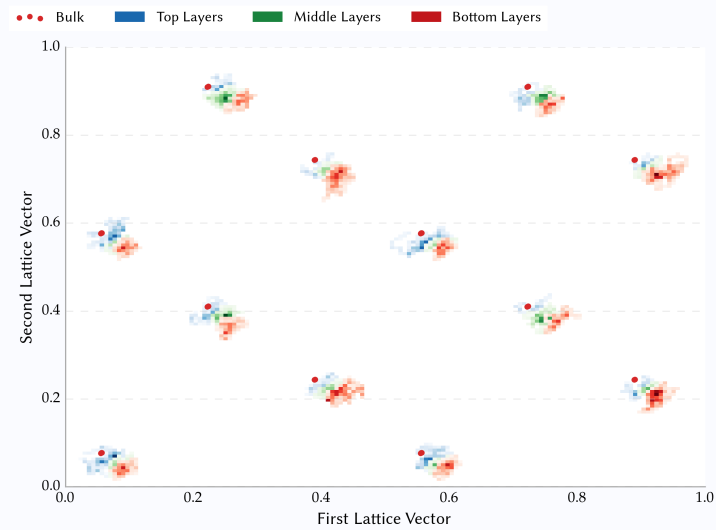
The resulting values then follow a distribution that reflects the vibrations of the atom cores, although smeared out due to the averaging. Figure 3.4 shows these distributions for two runs which have been started from two different initial configurations but are otherwise identical. It is interesting to see that the two runs agree with each other. Moreover, the results are symmetric across the hematite slab which is expected as there is no physical difference between the two surfaces—even the spin orientation of the first iron layer in the antiferromagnetic pattern compared to the surface normal vector is identical. From the coordinate distributions one can readily calculate the average position and the standard deviation. Again, the two runs give identical results in a statistical sense: the average



▲ **Figure 3.4** – Layer positions compared to the layer structure reported in literature from fits to experimental data[31] for the cross section through the hematite layer. Thin lines denote the distribution of positions for the two independent runs. Shaded areas denote the standard deviation, coloured bars the average position. The middle oxygen layer has been used as overlay reference point, which is why it has no distribution of positions in the two runs. The outermost hydrogen termination layers are not observed experimentally, so their ordinate position is arbitrary. Their position distribution is still valid. MD calculation and bulk optimisation (light grey) carried out at the same level of theory. Dark grey bars denote the positions from Trainor et al.[31] The iron indices distinguish between different iron sites in antiferromagnetic spin pattern.

Calculation(s) e49, 22c

values for one run are always within the standard deviation of the other run, often much closer. It is still surprising to see that the histograms of the layer position have quite long tails. They cannot result from the cumulative calculation of the reference position, so they must reflect an intrinsic increased atom mobility near the surface, which would relate to lower bond strengths. It is clearly visible that the second iron layer from the surface already exhibits much smaller tails. The position of the outermost oxygen layer also gives information about the average density of hematite near the surface. When comparing the position of this oxygen layer to the one calculated from bulk layer spacing, one can see that the oxygen layer for the interface is farther from the hematite slab than the corresponding



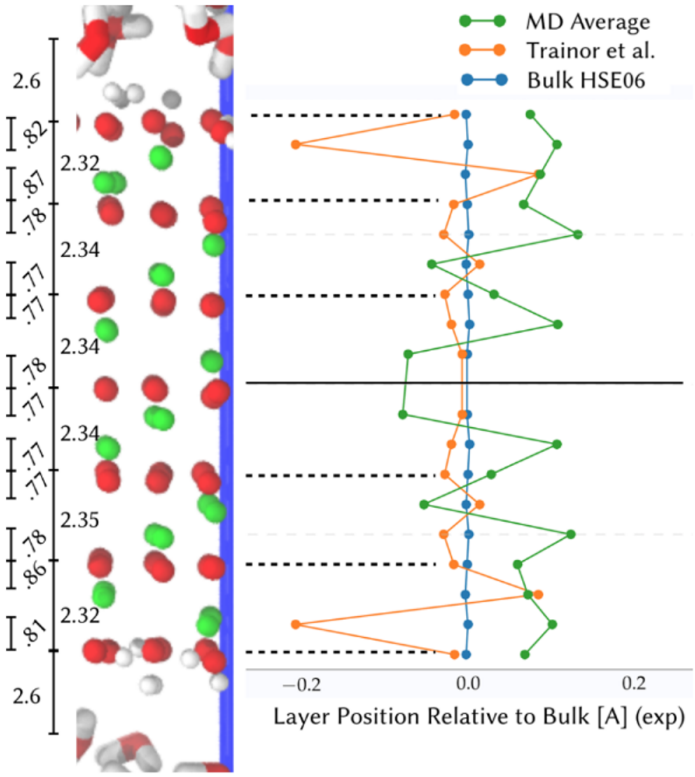
▶ **Figure 3.5** – Position histogram of the iron atom positions in the MD over two independent runs (coloured) compared to the average bulk MD positions of them in hematite. The positions have been projected onto the two lattice vectors of the unit cell **a** and **b** in order to allow square binning. Each histogram colour is for two layers of iron atoms in the hematite slab out of the six layers in each MD.

Calculation(s) e49, 22c

oxygen layer would be for bulk systems. This means that the density of the slab is a bit lower than the bulk system. This effect is small though: 0.2 Å total thickness difference for the whole slab of roughly 13 Å thickness is about 1.5 %. This small yet remarkable effect is unlikely to be an effect from the functional used for system treatment, as the hematite density of the bulk simulation matches the experimental density within the line width of Figure 3.4.

Of course, the analysis here assumes that the equilibration phase was long enough in order to thermalise the frozen hematite structure during the equilibration procedure in Figure 3.1. This has been checked and will be discussed in the context of Figure 3.23.

Layer Structure One of the most promising links between theory and experiment is the layer structure. More precisely, the changes in the interlayer spacing as a reaction to the presence of the surface can be easily derived from the atomistic representation of the system in theoretical calculations but is also accessible to measurements like the CTR setup. In this kind of experiments, the reaction of the diffraction peaks from x-ray radiation on the presence of the surface is analysed. The basic idea is that the presence of a solid-liquid interface reduces periodicity from three dimensions to two[213] which in turn broadens the refraction peaks. After the measurement, a model is constructed and theoretical spectra are calculated and compared to the actual experiment. This step is the most vaguely defined one, as the number of degrees of freedom is prohibitively large to conduct unbiased fitting



◀ **Figure 3.6** – Interlayer spacing for the oxygen termination of bulk hematite compared to bulk calculations at the same level of theory as well as experimental data.[31]

Calculation(s) e49, 22c

in this step.* Therefore, a set of assumptions and symmetry considerations is employed to reduce dimensionality. In one of the most commonly cited articles on the hematite-water interface for the (001) surface, there is[31] one assumption which does not seem to hold true in our simulation: the in-plane positions of the atoms in the first iron layer beneath the surface.

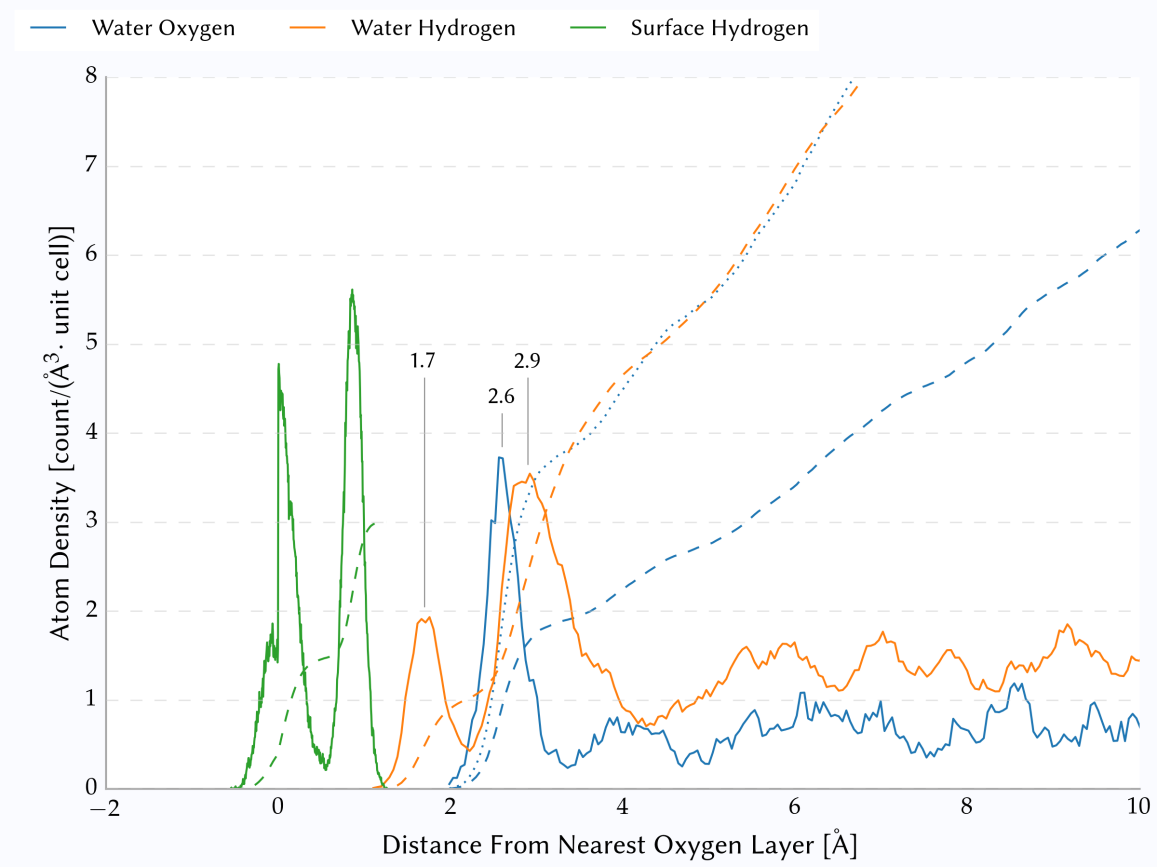
Over the course of the simulation, the relative positions of the iron atoms in the iron layers have been collected and merged into a two-dimensional histogram in Figure 3.5. For better comparison, the bulk positions of the iron atoms in these layers have been added to the diagram. From the data, it is clearly visible that the outermost iron atom positions are not kept at the same place in the lattice when projected onto a plane parallel to the interface. This means that the model of Trainor et al.[31] may be oversimplifying. This is supported by the fact that the average distance between the water oxygen atoms and the topmost oxygen layer of hematite is surprisingly low: 1.9Å. This is below the typical contact distance of oxygen atoms of 3.04 Å[142] and has been recognised as potential problem in the literature.[17]

* Usually, each atom has five fitting parameters: three (fractional) coordinates, one occupation percentage and one Gaussian width for the thermal fluctuations. In materials like hematite which has a strong anisotropy in the internal thermal fluctuations,[214] one Gaussian width per dimension is used, pushing the numbers of parameters to fit to seven per atom.

Figure 3.6 shows the layer spacing as obtained from AIMD in direct comparison to experimental values. The layer spacings have been averaged in a particular way. At first, a plane has been fitted to the middle layers of oxygen in the hematite slab. Because of the symmetry of the slab, these two oxygen layers are equally far away from the surface and represent the same bulk layer. Now the plane defined by them in each and every time step allows to calculate the distance from that plane for every atom in all time steps. Averaging of these distances over both atoms in a layer and time gives average positions of the individual layers. The difference of which forms the interlayer spacing. With the exception of subsurface iron layers, the calculated layer spacings are close to the experimental values.[31] The difference may result from two issues with the experimental analysis: the authors assumed that the in-plane positions of the iron atoms in the subsurface layer is identical to the positions of these atoms in bulk hematite (which is not the case according to the AIMD calculations) and the analysis allowed for unphysically short O-O distances between surface oxygen atoms and solvation layer positions.[17]

Solvation Structure Figure 3.7 shows the density profile of the water solvation. More precisely, the atom number density per unit cell surface is shown. This makes direct comparison to the accessible surface of hematite possible. For the hydrogen atoms, two peaks emerge of the trajectory data. The first of which relates to in-plane hydrogen bonds along the surface or atoms mostly aligned into that direction. This peak seems to have a shoulder for slightly negative values, but the peaks are not clearly separated with the given trajectory length. The integral of this first peak is about 1.5 hydrogen atoms per unit cell surface. As each unit cell features precisely three oxygen atoms, this means that half the termination hydrogen atoms of the fully terminated surface are staying in plane which is a surprisingly high share. The second hydrogen peak is centred around a distance to the outermost oxygen layers of about 1 Å. This is about the bond length between the surface oxygen and the corresponding hydrogen atom. As the second peak is centred around this value, it seems likely that at the other half of the time, the hydrogen atom of the surface is pointing towards the water layer. This is confirmed by the integral of the atom number density as it approaches three after the second peaks which matches the total number of hydrogen atoms per fully protonated unit cell surface.

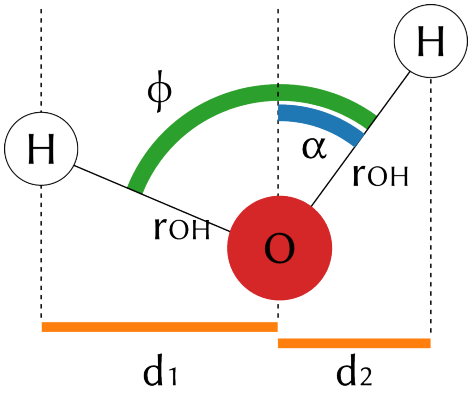
In order to investigate the internal water structure, the atom number density has been



▲ **Figure 3.7** – Atom number density per unit cell along the hematite layer normal vector which is identical to the third lattice vector of hematite. Dashed lines denote the integral of the stroked lines of same colour. The dotted line is twice the dashed blue line (the water hydrogen count) to allow direct comparison to the orange dashed line (the oxygen water count).

Calculation(s) e49, 22c

split into the elements of water in Figure 3.7. For the first hydrogen peak in the water component, the integral is about one meaning that this peak holds less hydrogen atoms than the first oxygen peak holds oxygen atoms. This means that part of the water bonds have to point towards the surface which implicitly leaves water bonds pointing roughly parallel to the surface and part of the water bonds have to point away from the surface such that the dipole vector is aligned perpendicular to the interface. In numbers, this can be readily seen from the integral atomic number density. The first oxygen peak consists of a little less than two atoms. This means that four hydrogen atoms are expected to be attached to these oxygen atoms. One of them forms the first hydrogen peak and the other four form the second peak. The shoulder on the second peak for large distances from the surface comes from those atoms where the dipole vector is aligned perpendicular to the surface. From the equilibrium angle ϕ in the following picture and the bond length r_{OH}



▶ **Figure 3.8** – Schematic drawing to visualise the variables in the tilt angle discussion of surface water molecules.

in the following picture, one can calculate the expected orientation for the water molecule that belongs to the first kind, where one bond is pointing towards the surface. This requires taking d_i from the data in Figure 3.7

With the peak positions from Figure 3.7 and variables as shown in Figure 3.8, the distances are given as

$$d_1 \simeq 0.9\,\text{Å} \qquad d_2 \simeq 0.3\,\text{Å} \tag{3.11}$$

while basic geometry gives

$$d_1 = r_{\text{OH}} \sin(\phi - \alpha) \tag{3.12}$$

$$d_2 = r_{\text{OH}} \sin\alpha \tag{3.13}$$

In order to eliminate one parameter of the equations we can consider only the fraction

$$\frac{d_2}{d_1} = \frac{\sin\alpha}{\sin(\phi-\alpha)} =: d' \tag{3.14}$$

There is no general analytical solution for α but as a Taylor expansion for small α, there is a solution

$$d' = \alpha \csc\phi + \alpha^2 \cot\phi \csc\phi + \mathcal{O}(\alpha^3) \tag{3.15}$$

which can be solved for α

$$\alpha \simeq -\frac{1}{2}\tan\phi \pm \sqrt{\frac{1}{4}\tan^2\phi + d' \sin\phi \tan\phi} \tag{3.16}$$

Now the only parameter is the angle formed by the bonds in water. For model systems, a value of 104.5° is often employed.[199] With $d' \simeq 1/3$, this gives

$$\alpha \simeq 20° \tag{3.17}$$

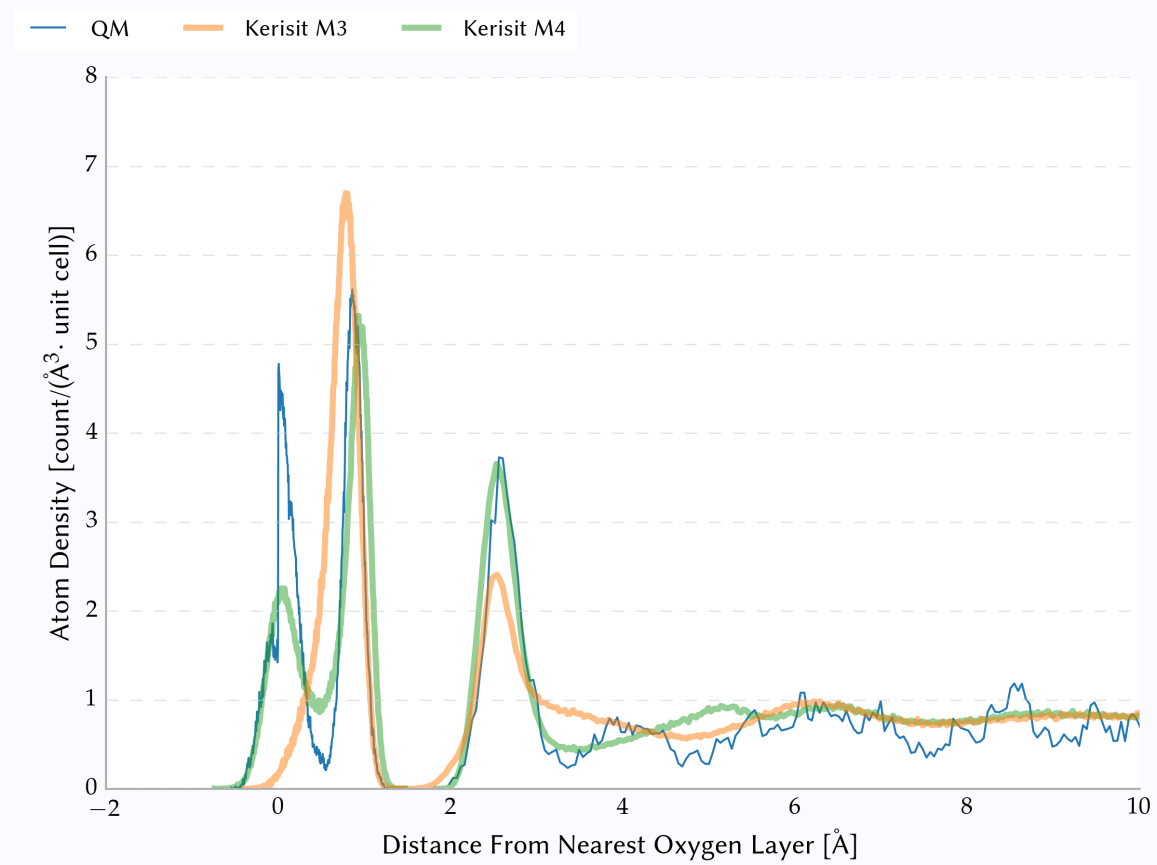
While this estimation is sensitive* to the imprecise value for d_2, it shows that given the density profile, about half of the water molecules in the first solvation shell have to point towards the surface with exactly one bond. This is surprising as now both the surface protons as well as the solvation protons are faced to each other which seems unlikely from an electrostatics point of view. Most likely, there is a balance between hydrogen bonds with hematite as donor and those where water is the donor. This is investigated in more detail in Figure 3.22.

Now the other interesting answer Figure 3.7 can give is about the required thickness of the water layer. When comparing the integral density of the water elements, it is clear that for short distances from the surface, the perturbation in the average orientation of the water molecules is substantial. This is expected, as the water surface disrupts the internal structure of water. After a bit more than 4 Å, the atom count for hydrogen atoms in water is precisely double the atom count for water oxygen atoms and mostly stays this way for increasing distance from the surface. This means that from the density distribution, a water layer of about 10 Å would be enough. In this simulation, we kept a much larger water layer. Therefore, we can assume that the water layer we have is thick enough in order to mimic bulk properties in its middle.

Comparison to Classical Force Fields When compared to classical calculations[32] as done in Figure 3.9, it is clearly visible that the ab-initio picture is different from the classical picture, most likely due to force-field limitations.[215] In particular, the spatially averaged atom number density of the terminating hydrogen atoms follows a different distribution where the distribution of the surface oxygen atoms and their protons is not properly captured.†

Electronic Properties Figure 3.10 shows the position of the conduction and valance band centres for subsystems of the whole interface, which is required in order to assess the stability of the electronic component of the calculation. As expected, the spin-up and spin-down channels give the same band centres for all elements except for iron because the antiferromagnetic nature of the material requires a net spin moment on the iron centres. The statistical error of the band centre position is very small for both the valence and the conduction band part, which means that the electronic structure is rather stable on the time

* If d_2 were set to 2.25 Å, the estimate for α would be 15°. † I particularly thank Dr. Kerisit for providing the raw data of his MD trajectory which allowed me to analyse the atom number density beyond the graphs given in this published work.

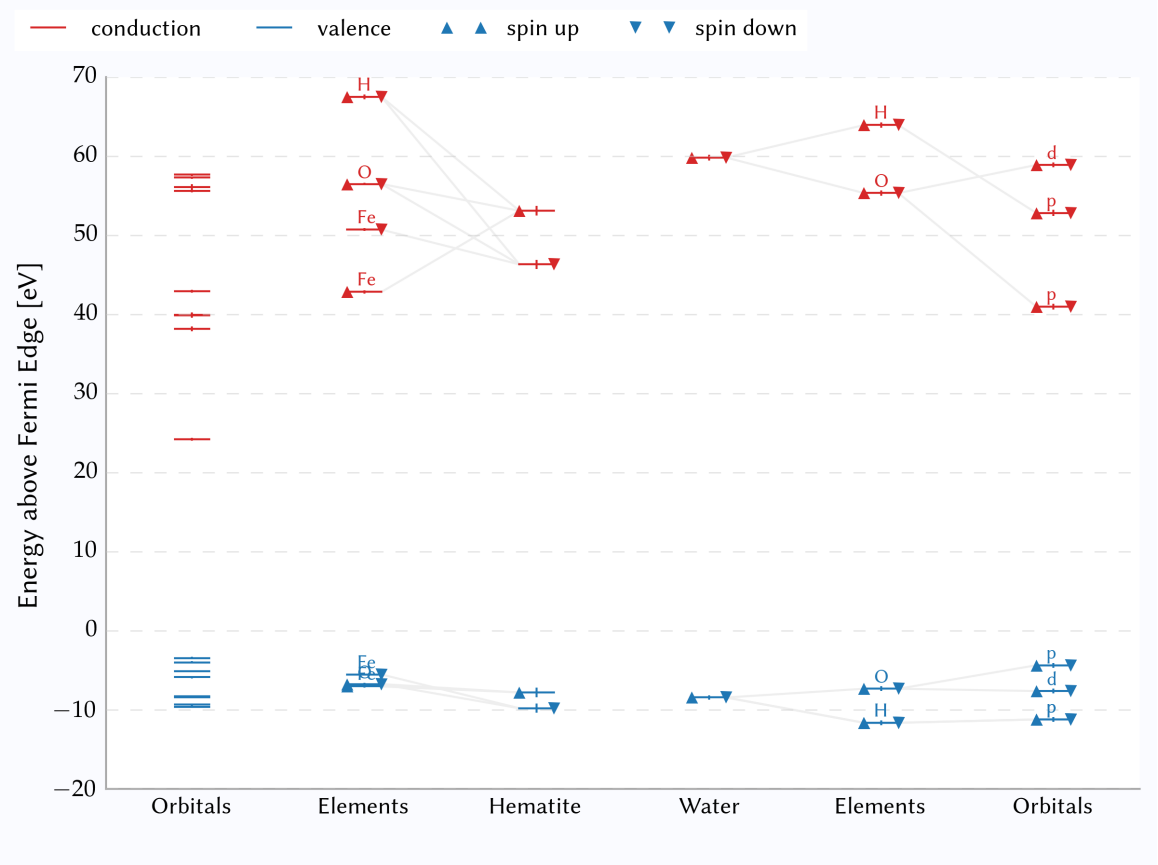


▲ **Figure 3.9** – Atom number density of the oxygen-terminated (001) hematite/liquid water interface compared to classical force fields.[32]

Calculation(s) e49

scale of a few picoseconds that has been assessed here. While the spin splitting of the iron band centres is comparably large, the splitting of the overall hematite part is only a bit smaller, which means that the DOS is dominated by the iron components for the hematite layer. This is in agreement with results for a single bulk hematite structure. This is particularly interesting, as the number of oxygen atoms compared to the number of iron atoms in the interface system is even higher than in the bulk structure, as setting up a symmetric interface at both sides of the slab requires another oxygen layer on top of the last iron layer. A stoichiometric setup would only be feasible by introducing a vacuum layer[216] (which brings its own challenges[217]), thereby reducing the computational efficiency, since only one surface would be sampled in the same trajectory rather than two independent surface at the same time.

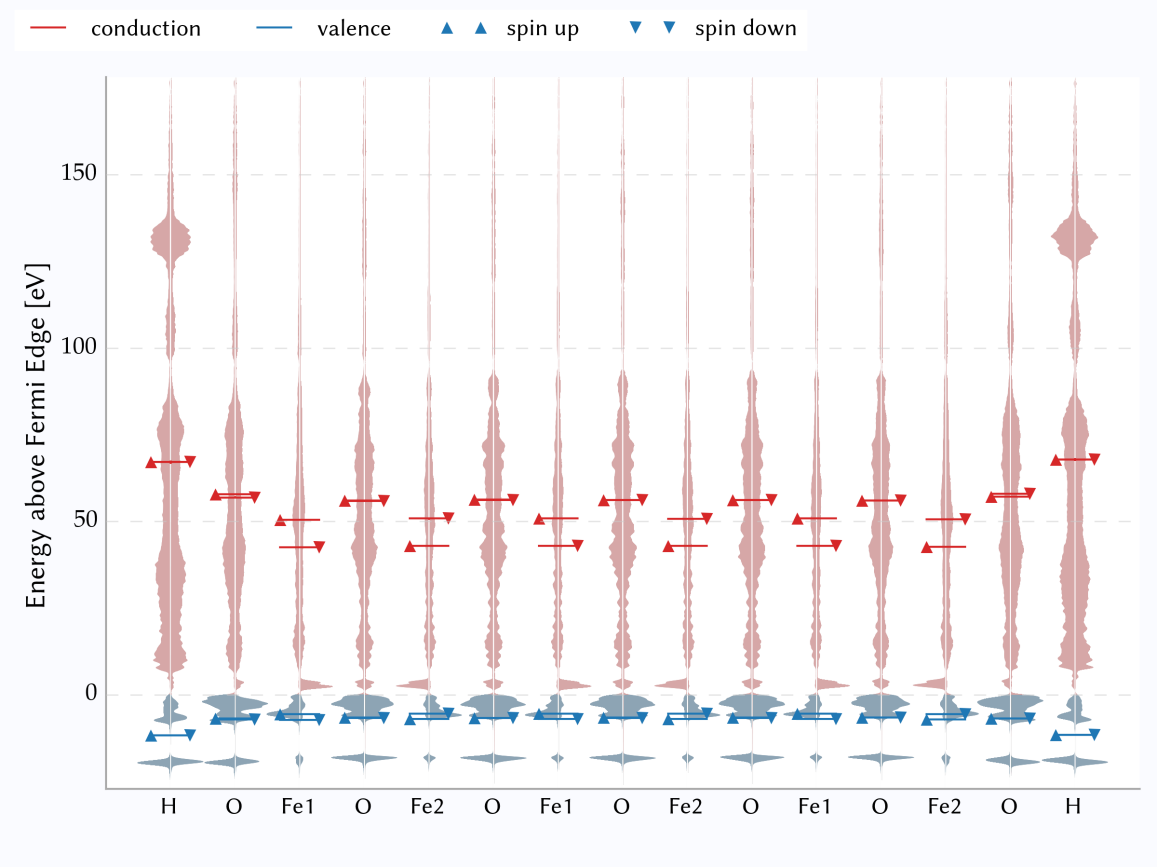
The small variance of the DOS band centres for the iron atoms in hematite is an indicator that the hematite layer in the middle of the simulation cell is thick enough for inves-



▲ **Figure 3.10** – Energy levels for both the conduction and the valence band of the interface. Spin splitting is shown. Error bars denote statistical error of this energy level over the trajectory.

Calculation(s) e49

tigation of bulk properties, which is crucial for correctly mimicking the interface structure in a fully periodic simulation system. Of course, this is no definite proof, as this requires comparing the PDOS in the cross section of the material. This is done in Figure 3.11 which shows the PDOS averaged over the trajectory but without smearing for each layer of atom kinds in hematite. It is to be noted that there are distinct atom kinds for the different spin alignments of iron atoms in order to allow for the antiferromagnetic spin pattern. While this is without effect on the physics and the results of system as the two atom kinds have identical basis sets, one can easily identify the opposite spin orientations. In Figure 3.11, the mirrored PDOS for the two spin channels can be seen right away. Moreover, this visualisation confirms the hypothesis that the electronic properties are well converged within the hematite layer. It is striking to see that the overall shape of PDOS is converged after even only very few layers from the surface. This may be interpreted as a strong shielding of the surface dipole and the water charge distributions. While the static dielec-

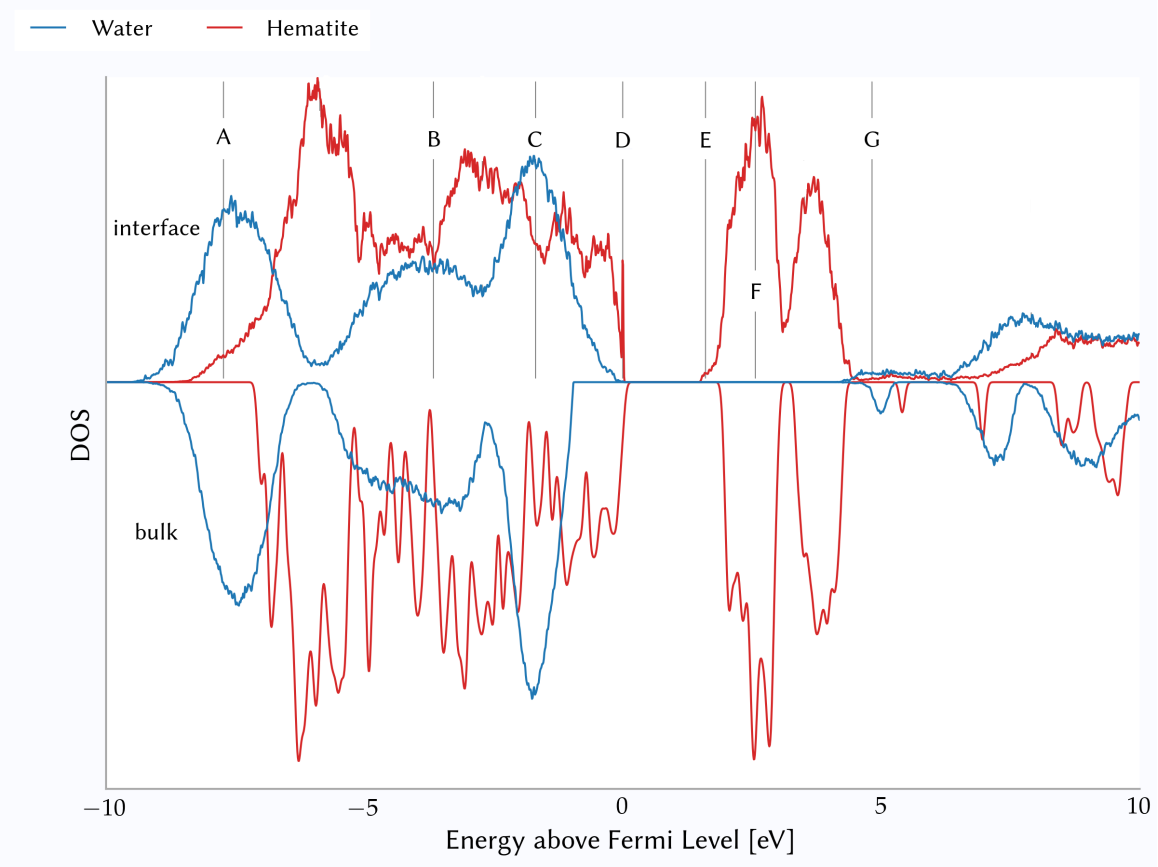


▲ **Figure 3.11** – PDOS for valence and conduction band for a cross section through the hematite layer averaged over a single run. For each layer the spin up channel (left half) and spin down channel (right half) is shown filled. Band centres are shown as bars with arrows indicating the spin channel. The iron indices distinguish between different iron sites in antiferromagnetic spin pattern.

Calculation(s) e49

tric constant of hematite is significantly lower than the one of water (and, additionally, anisotropic),[218,219] it is surprising to see that the screening length is very short compared to the roughly 10 Å that are required in (salty) liquid water.[220] Direct comparison of this graph with the Hartree potential in Figure 3.14 confirms strong screening of the surface in hematite, because the water Hartree potential decreases quite quickly to the hematite average on crossing of the interface. The thickness of the surface layer that is required for relaxation to bulk properties is comparable in both cases. As there is no rigorous definition of the two layers in question, no numbers are presented here.

Although hard to spot, Figure 3.11 presents error bars for the band centres as they are calculated from the variance over the whole trajectory. While this naturally ignores the existing autocorrelation of the values as the nuclear velocities are comparably low in the solid, it gives a good idea of the stability of the electronic structure of hematite in this setup.

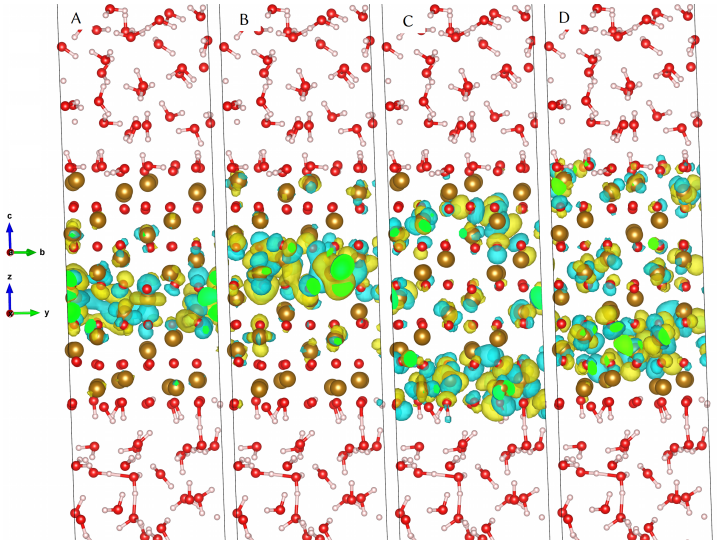


▲ **Figure 3.12** – PDOS for the two components of the interface system (top half) compared to the DOS for the isolated systems (bottom, mirrored). The interface results are averaged over two independent runs. No smearing for the interface part; 0.05 eV smearing for bulk hematite, 0.02 eV smearing for bulk water. Labels are as follows: Water $1b_2$ (A), $3a_1$ (B), $1b_1$ (C), edge of the valence band (D), edge of the conduction band (E), iron 3d (F), and water valence band edge (G).

Calculation(s) e49

This is mostly a confirmation of the equilibration procedure outlined in Figure 3.1.

One property which is of interest for interface is the change of the DOS. The effective two-dimensional constraints give rise to surface states[221,222] which then reduce the effective band gap of the interface as well as the one of the components which can be investigated by comparing the PDOS of the two phases to the DOS of the separate homogeneous and one-phased systems. Figure 3.12 does this comparison. It has to be noted that the results for the bulk systems are not averaged over a MD trajectory of sufficient length and, thus, slightly different smearing widths were necessary. While the overall DOS shape looks similar for both components in their pure calculations and in the interface setup, it looks as if the water component is shifted towards lower energies by roughly 1 eV. At the bottom of the conduction band of hematite, one can see that the interface introduces additional states



▶ **Figure 3.13** – HOMO (panel A+B) and LUMO (panel C+D) for the hematite/liquid water interface as obtained from a single thermally equilibrated snapshot of the oxygen terminated (001) surface. For both HOMO and LUMO, the left panels (A+C) show the alpha spin channel, while the right panels (B+D) show the beta spin channel.

Calculation(s) e49

or lowers existing ones into the former band gap.

Selected states in Figure 3.12 have been analysed by inspecting the corresponding MO visually. Although manual classification of the three-dimensional grid data is not quite a rigorous method of analysis, is has been used here in order to get a first impression of the MO shape. The results therefore are strictly qualitative. For more reliable quantified results, other tools like the electron localisation function[223–225] or similar approaches[226] would be necessary. The point A is near the maximum of the lowest-energy PDOS peak of water but has been chosen to have as little hematite contribution as possible. This peak mostly belongs to the $1b_1$ molecular orbitals of water, which matches the relative MO ordering given in reference data.[227] In any case, the overall PDOS of the water component of the interface is well reproduced and is very similar to the PDOS for the bulk water system, in particular the nature of the orbitals stays the same. There are two main differences between the interface and the bulk case for water here: a shift in energy and a different peak width. The shift in energy is due to the approximate[228] Fermi level alignment[229] of the two separate systems that has been used in this graph, while the narrower peaks in the water bulk DOS arise from the smaller amount (32 molecules instead of 92) of water molecules used in the bulk simulation. Very short simulations with a number of water molecules matching the one from the interface calculations confirm that this difference is the only reason for the different peak widths.

Figure 3.13 shows the HOMO and the LUMO of a single snapshot from the MD calculation. Both frontier orbitals are located on the hematite slab as the PDOS in Figure 3.12

already suggested. From the visualisation in Figure 3.13, the HOMO has some Fe3d character, but is mostly O2p – similar to what has been found for bulk hematite. The LUMO is mostly Fe3d and is delocalised over more hematite layers. In general, the HOMO is localised on the middle layers of hematite, while the LUMO is close to the surfaces. The occupation of different layers in Figure 3.13 comes from the antiferromagnetic spin pattern of hematite which means that the slab has iron layers of alternating alpha and beta spin majority. Other snapshots of the MD calculation give the same picture, although the exact extent of the composition depends on the individual configuration.

Hartree Potential and Electron Density The Hartree potential is defined as the effective electrostatic potential of the overall charge distribution that would be felt by an additional non-interacting test charge at any given point in space under the constraint that no relaxation of the current charge density due to the additional fictitious charge is allowed. This gives an insight into the averaged Coulomb potential due to the charge distribution in the system. Without averaging over the trajectory, interpretation of this property is particularly difficult, as subtle changes in conformation can have a significant influence on the Hartree potential.

In this system, the local Hartree potential is of little interest, as it is highly dependent on the local orientation of the atoms. It is more interesting to slice the system into thin bits parallel to the surface of the hematite. This way, one can check how far the electrostatic influence of the interface reaches into the water subsystem.[230,231] While the high ε_r of water (about 80) suggests a strong shielding of any local charge concentration, it is hard to predict this value without any simulation. For the analysis in Figure 3.14, the two sides of the interface have not been wrapped together into one single graph as this would obstruct the results due to possible overlaid potential decays. As the graph also features the average layer positions for each oxygen atom layer of hematite and the twofold iron layers in between the oxygen layers, it is possible to directly compare the atom positions to the potential. The only limitation here is that the standard deviation of the Hartree potential is quite substantial, which arises from the limited length of the trajectory. Unfortunately, there is another contribution to the statistical error which cannot be resolved by increasing the trajectory length. As the third lattice vector of the simulation box is designed to be parallel or anti-parallel to the surface normal vector, it is helpful to calculate the projection

in the slices coming from the internal plane wave grid from CP2K. This way, there is no need to alter the attribution of the voxel in the three-dimensional grid to new slices, which would be prone to numerical errors. Moreover, there is only one alternative that has the potential to yield improved results in the sense of less statistical noise: fitting a plane through the separate layers of hematite* followed by averaging the normal vectors†. The advantage here would be a more accurate representation of the surface plane orientation and, therefore, a clearer picture for the Hartree potential depending on the distance to the surface. The problem here is that the averaged normal vector of the surface fluctuates over time with a magnitude that is comparable to the variation of direction when using the third box vector of the simulation box. Therefore, there is very little potential gain but much higher complexity and potential new error sources which may compensate the small gain or even overcompensate it. Therefore, in this analysis, only the projection in slices along the third lattice vector are reported.

There is another effect that has to be accounted for in the analysis, namely the slight movement of the hematite relative to the water in the simulation. While the overall trajectory is kept free of COM movement in CP2K, this is not guaranteed for any arbitrary subsystem in general and certainly not the case for the hematite/water subsystems. This means that for each frame the slices of the Hartree potential have to be shifted along the third box vector until the deviation between the hematite COM in the given frame of the trajectory and the hematite COM in the reference frame‡ is minimised. As the slices are discrete, there is no perfect overlap of the two frames in most cases. This error, however, is very small, as the overall simulation box has 525 slices in total, so the expectation value of the error Δ_c in position due to discretisation is

$$\langle \Delta_c \rangle = 0.5 \cdot \frac{\mathbf{c}}{525} \Rightarrow \frac{\Delta_c}{|\mathbf{c}|} \simeq 0.001 \tag{3.18}$$

With a relative error of about one thousandth of the length of the third lattice vector, that typically has some 42 Å, the error is easily acceptable.

Figure 3.14 shows that the Hartree potential exhibits sharp peaks at the average oxygen layer position and less pronounced peaks for each iron layer. This is expected as there are two iron layers between any two neighbouring oxygen layers. While in the rest of this work

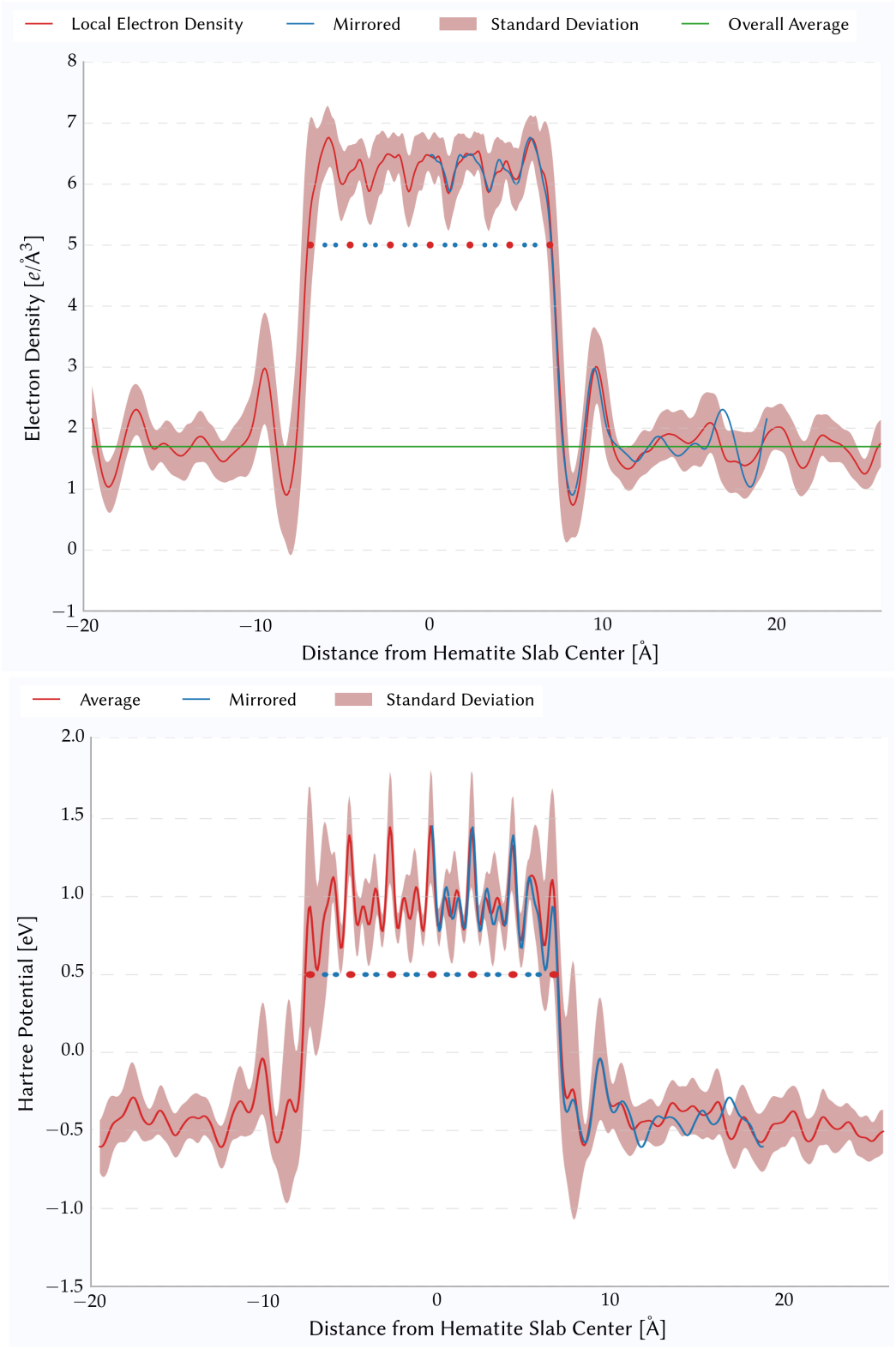
* This can be done using a SVD of the atom coordinates after shifting a layer such that the COM of which is at the origin. † This may require flipping the normal vectors such that they are almost parallel, as any unbiased plane fitting method gives results that are symmetric on inversion of the estimate normal vector. ‡ This reference frame is arbitrary. Here, the first frame of the trajectory is chosen.

these two layers are usually referred to as one single layer, here the distinction is crucial as these two sublayers are at slightly different coordinates in the hematite structure. As the (theoretical) time-averaged charge of the oxygen layer is identical to the time-averaged charge of the two iron sublayers combined, the charge in the iron layers is expected to be distributed over the two sublayers, resulting in the smaller peaks in the Hartree potential.

As the inner oxygen layer is the symmetry plane* for the hematite subsystems (as long as the spin orientation is ignored), one can map the Hartree potential at this point. This gives insight into the stability of the property given the amount of sampling that was affordable for these calculations. Figure 3.14 therefore shows both the complete cross section as well as the left part of the averaged Hartree potential mapped onto the right part of the graph. This way, one can immediately see that the peak positions, heights and widths for the oxygen layers are identical on both sides. This means that the sampling for this component of hematite is well-represented in the simulation. For the iron layers, there is no difference for the peak position and peak shape but the peak heights are significantly different for the two curves that have been superimposed. This may come from slight differences in geometry already present in the crystal structure of hematite. What is reassuring here is that the decay towards the water subsystem is nearly identical for both sides of hematite. This gives rise to the idea of superimposing the trajectory data for the two surfaces in order to increase sampling of dynamic properties. The Hartree potential is very stable for the first 4-5 Å from the outermost oxygen layer which means that this part has to be regularly structured. Otherwise, the random fluctuations of the water molecules would obstruct forming a clear Hartree potential profile as they do for bulk water which is farther from the hematite surface. One can clearly see that the Hartree potential is irregular with constant standard deviation for large distances from the hematite surface—just like it is expected for bulk water. The Hartree potential difference of 1.4 eV is related to the surface dipole (with contributions both from solvent molecules due to their average orientation in the first solvation shell and OH groups at the surface) and the interface capacity[82] but for a meaningful evaluation of both properties, counter-ions in the solvent would be required.[82]

While Figure 3.14 does not include the core electrons that are treated by GTH pseu-

* This is a good approximation. In fact, the symmetry is slightly broken due to distortions present in the crystal structure.



▲ **Figure 3.14** – Top: Electron density along the hematite layer normal vector which is identical to the third lattice vector of hematite. Red and blue dots denote the minimum energy geometry position of oxygen atoms and iron atoms respectively. Bottom: Hartree potential of a single positive non-interacting test charge along the hematite layer normal vector which is identical to the third lattice vector of hematite. Red and blue dots denote the minimum energy geometry position of oxygen atoms and iron atoms respectively.

Calculation(s) e49

dopotentials but only those that are explicitly treated in the DFT approach here, there is a clear picture of charge localisation. Clearly, the two surfaces of the neutral hematite layer are very similar if not identical given the statistical uncertainty resulting from the finite sampling duration. Right between the hematite and the water, the electron density is the lowest for the whole cross section. This is not surprising as the regular lattice structure of hematite comes to an end there and is terminated by hydrogen atoms which may point to the water at some time over the trajectory but even if they do, these hydrogen atoms carry little electron density as the charge is expected to be mostly localised on the oxygen atoms the hydrogen atoms are bonded to because these oxygen atoms have a much higher electronegativity. The same goes for the intramolecular bonds in the water layer. Due to the strong repulsion of the oxygen atoms of the hematite surface and the oxygen atoms of the water formed by coulomb repulsion and van der Waals interactions, there are no atoms to be found in this narrow gap that carry a significant electron density. This gap is mostly populated by the surface hydrogen atoms, as shown in Figure 3.7.

About 2.5 Å away from the surface, there is a clear peak in electron density which coincides with the first layer of oxygen atoms, as shown in Figure 3.7. The fact that there is no visible influence of the corresponding hydrogen atoms is expected for the same reasons as discussed above. Reversing this logical step, this means that the remaining part of the electron density across the water layer can be attributed to the water oxygen atoms only. As their atom number density is about constant, this matches the expectations. For comparison, the average electron density on the left side of the hematite slab in the way the cross section is visualised in Figure 3.14 is mirrored along the central plane of hematite, the middle oxygen layer. It is clearly visible that the decay of the electron density depending on the distance from the top hematite oxygen layer is identical within the standard deviation which is a hint towards reasonable convergence.*

Comparing the superimposed average atom positions in Figure 3.14 to the electron density in the hematite layer, one can clearly spot areas of higher electron density around the oxygen layers. While the core electrons are treated with pseudopotentials leaving six and sixteen explicit electrons for oxygen atoms and iron atoms respectively, their number and spatial organisation is different in hematite. Each oxygen layer in this cross section has

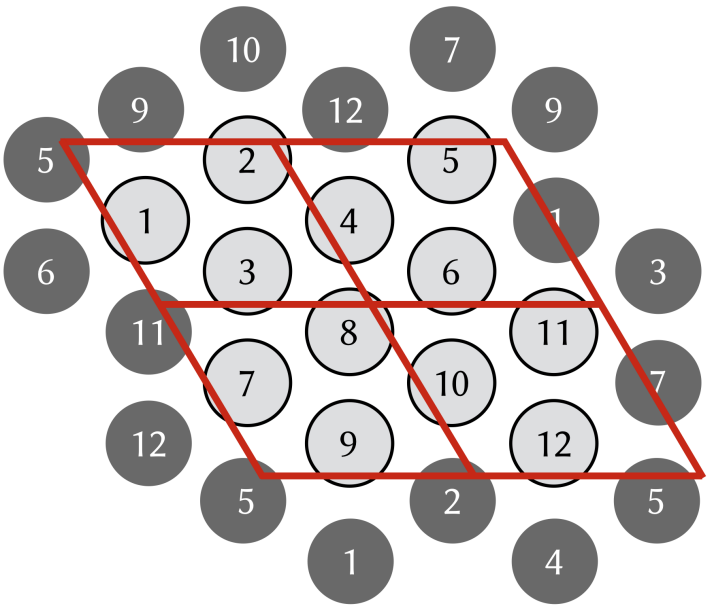
* Unfortunately, this is no measure of global convergence, as the electron density rearranges much faster than the atomic nuclei. Therefore, it is only valid to say that for the atomic configurations sampled in this calculation, the electron densities give a converged and consistent picture of the average instantaneous charge distribution.

twelve oxygen atoms and eight iron atoms where the oxygen atoms share their average coordinate along the cross section and the iron atoms do not. The iron atoms come in a two-layer substructure which means that their electron count (per sublayer) is 72 and 64 for the oxygen and iron atoms respectively. This population relation is well-reflected in the electron density shown in Figure 3.14.

Surface Dynamics The major advantage of a MD calculation of the system in question rather than taking the information from SPC stretched out over the trajectory of a classical calculation is the ability to get dynamic information regarding the system as long as the underlying process is fast enough to be captured by the available trajectory.

As shown in Figure 3.7, there are two distinct peaks in the atom number density of surface termination hydrogen atoms near the interface. In the following, hydrogen atoms that remain in the surface on roughly the same vertical position as the oxygen atoms they are bonded to, are referred to as in-plane sites whereas those protons that are pointing towards the water are called out-of-plane sites. Over the course of the trajectory, the nature of a site may change, i. e. a proton may move away from the surface or move closer to the surface plane. Therefore, the character of the sites has to be determined for every frame and every site independently. Given the histogram shown in Figure 3.7, the criterion was chosen as follows: for each frame and each surface (top or bottom surface of the slab), a plane was fitted through the oxygen atoms of the top layer using SVD. Then, the shortest signed distance between this surface and each proton was calculated such that positive distances relate to a proton position above the surface pointing towards water and negative distances belong to protons below the surface oriented towards the bulk hematite slab. If this signed distance is larger than 0.5 Å (roughly half the average O$\cdots$H bond length in water) then this proton is considered to belong to the out-of-plane group, otherwise, it is treated as in-plane proton.

Once this criterion has been applied to all frames and all termination sites, it is possible to identify surface patterns. Given the crystal structure of bulk hematite, there are only two symmetry operators for the surface: the **a** lattice vector and the **b** lattice vector. Although it looks like it at first glance, rotation is no symmetry operator and neither is inversion. To illustrate the symmetry operations due to PBC, Figure 3.15 shows a top view of a single surface. In total, there are four unit cells exposed to the surface, each of which has three



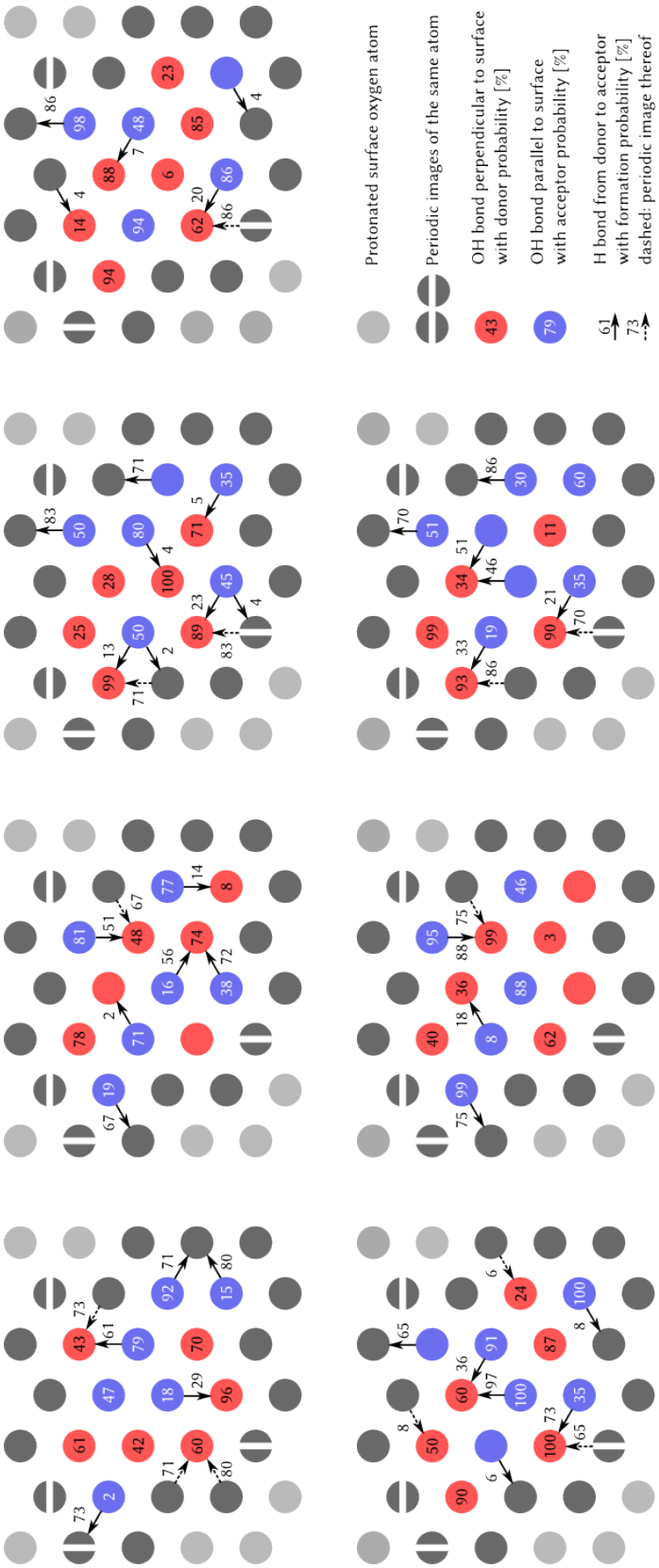
◀ **Figure 3.15** – Surface termination sites of the hematite slab (top view from the water layer). Light grey sites are uniquely numbered sites in the simulation setup, dark grey sites are the periodic images thereof. Red bars denote the unit cell size. Each site is a protonated oxygen atom.

oxygen sites. If the four unit cells are labelled (a, b, c, d) in their original order, then there are three more orders that are identical to this one: (b, a, d, c), (c, d, a, b), and (d, c, b, a), as shown in Figure 3.15.

In the following, a binary vector (0 for in-plane sites, 1 for out-of-plane sites) is called a surface pattern. As outlined above, some of these patterns are identical under the symmetry of the system. For all frames and the two surfaces of the hematite slab, the surface patterns have been determined and folded back on each other to obey the symmetry rules of the system. This results in seven highly prevalent patterns, as shown in Figure 3.16.

Over the course of the simulations, distinct patterns have been observed (see Figure 3.16) that are stable on the time scale of 0.7 ps to 1.8 ps. Switches between these stable surface termination patterns are fast and happen spontaneously within 50 fs.

Interestingly, no chaining of hydrogen bonds across the surface is observed where a series of in-plane oxygen terminations form hydrogen bonds. In fact, for these stable surface patterns, there is not a single hydrogen bond formed between two in-plane oxygen terminations at all, despite the fact that the oxygen-oxygen distance is advantageous for forming a hydrogen bond.[232] The hydrogen bonds that are formed come in two distinct patterns: a three-site one where two in-plane sites form a hydrogen bond to an out-of-plane site that in turn forms a hydrogen bond to a water molecule, while there is a two-site pattern that consists of an in-plane site accepting a hydrogen bond from a water molecule, forming one itself to an out-of-plane site that donates to a water molecule again. Together with



▲ **Figure 3.16** – Top view onto the surface: The seven stable patterns of surface termination orientation found in simulations with their hydrogen bond structure. Probabilities are conditional probabilities, that is, they state how likely it is to observe a particular hydrogen bond given this particular surface pattern. Any probabilities below 2 % are omitted for clarity.

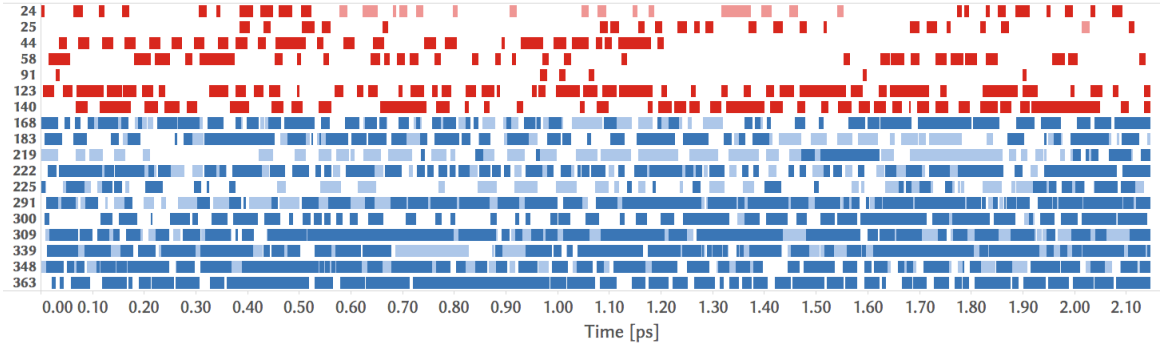
Calculation(s) e49, 22c, 6cf, e47, ab3, 10d, 0da, 6a0, 87b, 523

the absence of any hydrogen bond chains between in-plane sites, this hints towards a very localised interaction driving the surface termination pattern.

Two features are observed that seem to support that the surface structure is driven by the coordinating water on the surface: some hydrogen bonds are stable for nearly the whole lifetime of the surface patterns and some sites (whether in-plane or out-of-plane type) are completely avoided. The stable hydrogen bonds of comparably high lifetime[233] offer a link between the surface pattern and water adsorption on the surface which suggests a lower water mobility at the surface. Together the fast switches between distinct and stable surface patterns, this points towards changes in the water solvation structure that provide the energy to alter the stable surface termination patterns.

While the presence of self-similar surface termination patterns in Figure 3.16 is a sign that the system setup is large enough to capture the features spontaneously emerging in this system, there are other rare events that interesting for interface dynamics that would highly benefit from prolonged trajectories. Those are switches where an in-plane site breaks a hydrogen bond with one out-of-plane site just to form a new one with a different out-of-plane site next to it or those where an in-plane site that is hydrogen bonded to another surface oxygen becomes an out-of-plane site with a hydrogen bond to bulk water. The latter process apparently requires a reorganisation of the water molecules close to the surface. Features like these are the case in point why besides the current approximations that render our calculations possible, further method development is required to fully capture the details of surface dynamics properties.

Generally, the detection of the presence of hydrogen bonds is based on geometric criteria that are evaluated instantaneously. Geometric properties are comparably easy to evaluate and are suitable for both classical and ab-initio results. Using an instantaneous criterion is helpful for long sparse classical trajectories where the cost of storing every MD frame is avoided. However, a purely geometric criterion is somewhat artificial, as the actual geometry of a hydrogen bond heavily depends on the nature of the molecules in question.[232] This is commonly compensated by adjusting the criterion to match expectations or experimental estimates based on the involved compounds. While this approach works well for the occupancy analysis of hydrogen bonds, it significantly overestimates the breaking frequency and, therefore, underestimates the life time of a hydrogen bond.



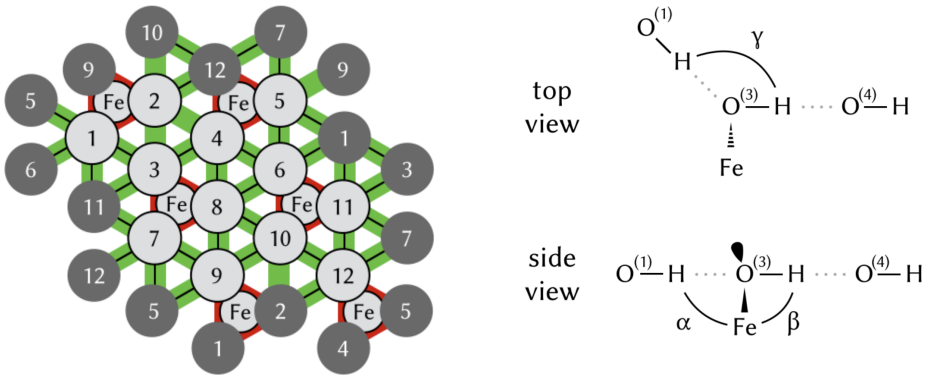
▲ **Figure 3.17** – Illustration of underestimation of hydrogen bond lifetimes for geometry based criteria. Timeline of the presence of hydrogen bonds for selected surface oxygen indices (y-axis). The coloured bars show the persistence of hydrogen bonds based on a purely geometry-based criterion. Red bars represent hydrogen bonds with a hematite oxygen as donor, blue bars have a water oxygen as hydrogen bond donor. Saturated colours are used for hydrogen bonds within the same subsystem, i. e. hematite-hematite or water-water while shaded bars are used for inter-subsystem hydrogen bonds. The hydrogen bond criterion in this case requires an O-H-O angle of at least 160° and allows a maximum O-O distance of 3.5 Å.

Calculation(s) e47, ab3, 10d, 0da, 6a0, 87b, 523

Figure 3.17 illustrates this problem for a selection of oxygen donors from a MD trajectory of the hematite-water interface.

As shown in Figure 3.17, it is commonly observed that the purely geometric criterion is violated for very short times. In order to improve the quality of the results in this case, a new parameter to hydrogen bond detection has been added: τ gives the maximum duration a violation of the geometric criterion is allowed. Introducing this temporal parameter slightly complicates the evaluation of the criterion, as a history of past hydrogen bonds has to be kept for τ. In this work, τ has been calibrated by calculating the average hydrogen bond lifetime for the bulk water phase of the setup and match it to literature estimations.[233] In the end, the criterion for detection of hydrogen bonds is threefold: the O-H-O angle has to be at least 160°, the O-O distance has to be at most 3.5 Å and a temporary violation of these criteria is allowed for at most 30 fs.

As mentioned before, only particular features are observed in the patterns spontaneously forming over the course of the trajectory. In particular, there are two site features, where an in-plane surface termination site accepts a hydrogen bond and simultaneously forms another one along the surface to an out-of-plane site which in turn is HB donor to the water layer. When viewed from the side of the interface, this two-site pattern has a U shape. The other feature is a three-site feature that can been seen as an extension of the two-site model since in this case the out-of-plane OH group is accepting HB from two



▲ **Figure 3.18** – Left: Circles with numbers are protonated surface oxygen sites. Labels enumerate all oxygen sites and illustrate periodic images (grey). Red bars: HB formation suppressed by iron atom close to surface. Green bars: Remaining possibilities for HB formation. Black lines: HB actually observed. Right: Illustration of the geometry of two hypothetical HB chains 1-3-4 and 9-8-6, see text for details.

in-plane sites instead of one. From the side, this looks like a W shape.

The lack of any in-plane chains where at least one in-plane site forms a HB to another in-plane site can be conclusively explained by a combination of geometric and electronic arguments. First, as shown in Figure 3.18, there is a largely regular grid of protonated surface sites. To some extent, this resembles a (slightly distorted) hexagonal pattern. Below the surface oxygen layer is the first iron layer which in turn consists of two sublayers one of which is closer to the surface. Due to the chemical environment, the iron atoms are positively charged. Having a look at the observed HB in any surface termination patterns in Figure 3.18 shows an interesting pattern: not a single HB is formed between surface oxygen atoms if a subsurface iron atom is close, but nearly all other geometrically possible HB are found at least once over the course of the trajectories. This can be seen as a strong repulsion between the positively charged subsurface iron atom and the proton.

This set of forbidden HB only leaves a certain number of three-site chains that could potentially be part of an in-plane HB chain. This can be further narrowed down by investigating the coordination of the central oxygen atom in two chained HB. In the right panel of Figure 3.18 this is illustrated for an example set of surface oxygen atoms. Since the distances between the oxygen atoms in questions are merely the same, the angles between them get more important.[232,234] Since the ideal coordination of an oxygen atom is tetrahedral with angles of 109.5° each, this can be used to evaluate the likelihood of HB chains for certain oxygen atoms on the surface. Table 3.2 lists all the potential unique three-site chains

Oxygen indices	α [°]	β [°]	γ [°]	RMSD [°]
1-3-2	135	144	64	63
1-3-4	135	94	124	33
4-6-8	135	83	60	62
2-4-6	94	83	176	73
4-8-9	83	94	176	73
4-8-10	83	144	124	46
4-8-6	83	135	60	62
9-8-6	94	135	124	33

▲ **Table 3.2** – Angles according to Figure 3.18 for the unique chained hydrogen bonds possible across the surface of hematite. RMSD is given relative to the optimal angle value 109.5°.

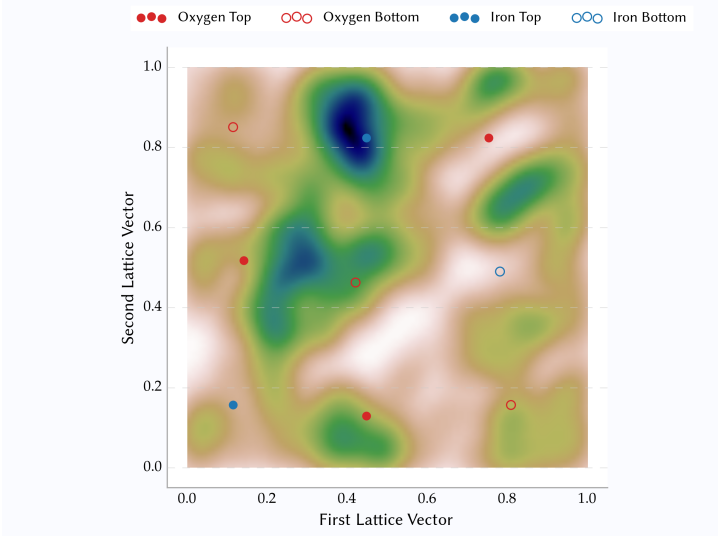
Calculation(s) 634

and their RMSD compared to the ideal geometry. It is evident that only two options have a low RMSD. Moreover, these two have a RMSD from the ideal geometry comparable to the one of the two-site features typically seen in the trajectories. To explain why the latter is observed and the former are not, an electronic structure argument is necessary.

As shown in Figure 3.14, the Hartree potential features a steep increase at the surface defined by the average position of the top layer oxygen atoms. Figure 3.18 shows the location of the lone electron pair. For the two remaining in-plane chains, this lone pair would be above the surface pointing towards the water layer, while for the two-site feature, this electron pair is below the hematite surface. This means that for any in-plane chain along the surface, a high energy penalty due to the Hartree potential for this electron charge prevents any formation in the trajectory while still allowing the clearly prominent two-site feature.

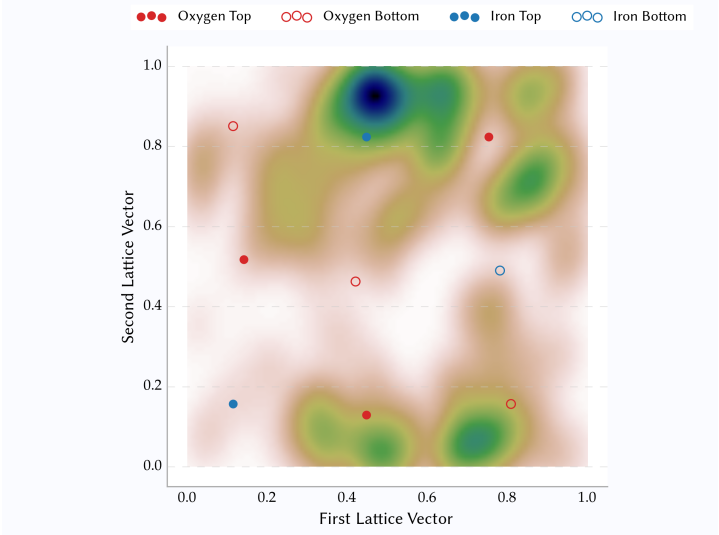
While the atom number densities in Figure 3.7 along the normal vector of the interface converge quite quickly within a few hundreds of femtoseconds, the distribution of atoms across the surface is much slower. Figure 3.19, Figure 3.20, and Figure 3.21 illustrate this for the neutral interface state. The distributions in these plots have ben calculated using Kernel Density Estimation (KDE). This method takes each sample point from a set of data points and calculates the most likely underlying distribution from which the data points have been drawn. In the 2D case, this method is essentially summing Gaussian functions around each data point.

Figure 3.19 shows the distribution of water oxygen atoms over the c-cut surface. The



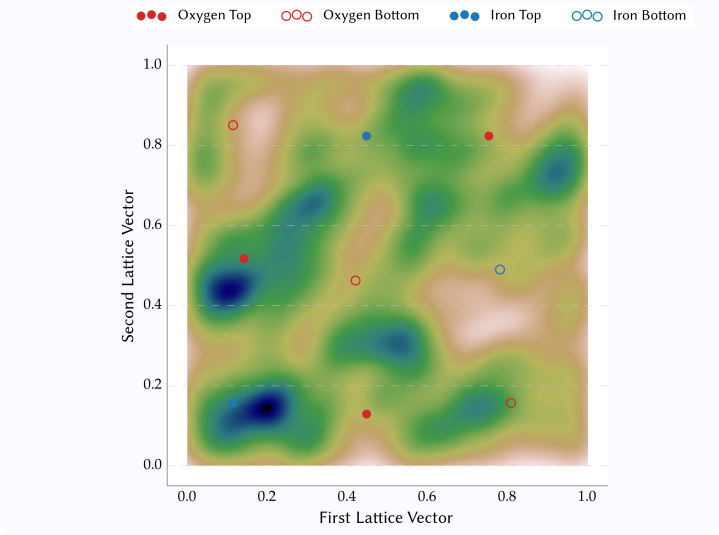
◀ **Figure 3.19** – Local density of the water oxygen atoms of the first solvation shell over a single unit cell of hematite. Coordinates are projected onto lattice vectors to allow square binning. The raw density has been fitted using a two-dimensional maximum likelihood estimation (Gaussian KDE). The nine points (see legend) denote the bulk position of the exposed hematite atoms of the surface.

Calculation(s) e49, 22c



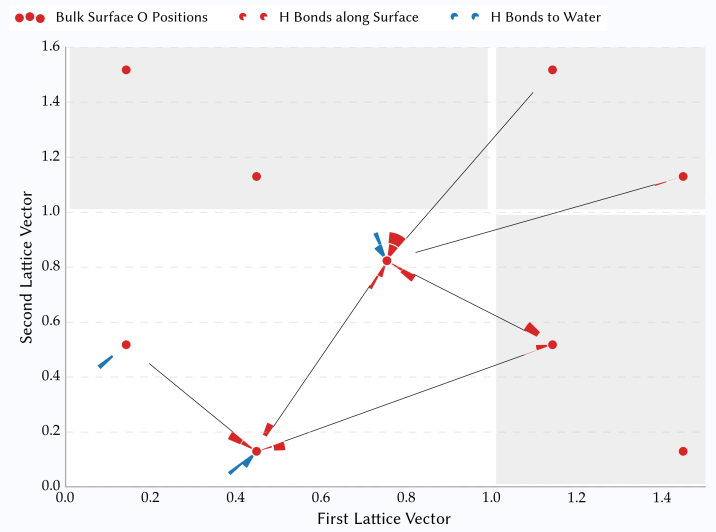
◀ **Figure 3.20** – Local density of the water hydrogen atoms of the first solvation shell over a single unit cell of hematite. Coordinates are projected onto lattice vectors to allow square binning. The raw density has been fitted using a two-dimensional maximum likelihood estimation (Gaussian KDE). The nine points (see legend) denote the bulk position of the exposed hematite atoms of the surface.

Calculation(s) e49, 22c



◀ **Figure 3.21** – Local density of the water hydrogen atoms of the second solvation shell over a single unit cell of hematite. Coordinates are projected onto lattice vectors to allow square binning. The raw density has been fitted using a two-dimensional maximum likelihood estimation (Gaussian KDE). The nine points (see legend) denote the bulk position of the exposed hematite atoms of the surface.

Calculation(s) e49, 22c

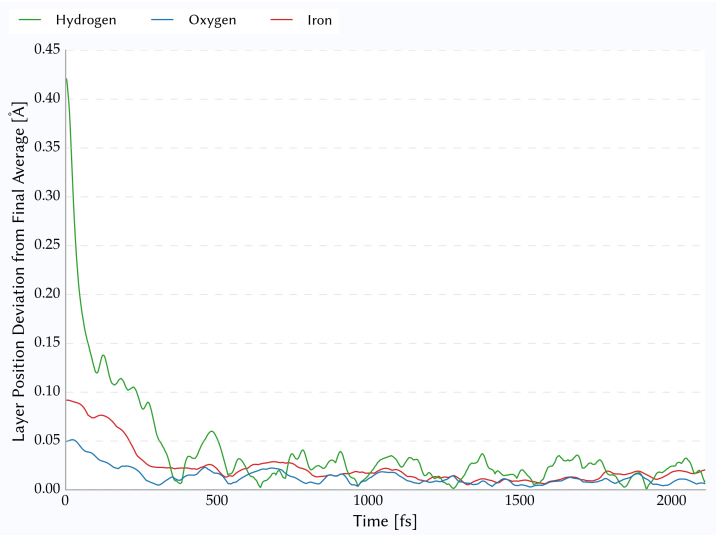


▶ **Figure 3.22** – Hydrogen bonds due to termination hydrogens at the interface. The bulk positions of the oxygen atoms are shown as dots. To each oxygen atom there is a hydrogen atom which may point towards different neighbours while forming a hydrogen bond. The percentage of the time where there is a hydrogen bond from one oxygen to another is shown as pie diagram around each surface oxygen. Blue pie elements denote hydrogen bonds towards water oxygen atoms.

Calculation(s) e49, 22c, 782

strongest peak is located right above the subsurface iron sites as exposed by the surface. Due to the positive charge of that iron site, negatively charged oxygen atoms are attracted. Note that in all figures showing the 2D distribution over the surface, the coordinates are given as fractional coordinates of the two lattice vectors that span an angle of 60 degrees rather than the 90 degrees shown in the diagram. The surface unit cell shown is fully periodic. It is of little surprise that the water hydrogen atoms as shown in Figure 3.20 are close to the corresponding oxygen atoms in Figure 3.19. However, the distribution of the light hydrogen atoms is much less diffuse than the one of the heavier oxygen atoms. To some extent, the hydrogen bond network is visible in the distribution of the hydrogen atoms in the first solvation shell of water, since the protons are mostly distributed around those oxygen atoms that are closest to the water.

For the second (very weak) solvation shell, the water hydrogen atoms in Figure 3.21 are much more equally distributed. This hints towards the water being much more mobile in this shell than it is the case in the first solvation shell. Since it has been shown earlier in this work that the hematite surface exhibits stable surface patterns, it is expected that the first solvation shell is taking part in this surface structure and, hence, is somewhat less mobile than bulk water. For protons at the surface however, a higher volatility of the the configuration has been shown, which is particularly pronounced for the second solvation shell. Usually, finite sampling of the system produces distributions which feature more pronounced peaks than the real physical distribution. This means that the broad distribution in Figure 3.21 is unlikely to be a result of finite sampling.



◀ **Figure 3.23** – Average deviation of the atomic layers in the hematite slab from their average position over the last 100 fs of the equilibration run. The graph covers the whole AIMD part of the equilibration procedure. Curves smoothed using an EWMA with a halflife of 40 fs. Constant finite fluctuations are expected due to thermoinduced vibrations.

Calculation(s) 140

Figure 3.22 shows the hydrogen bond network internal to the hematite surface only. Every line connecting two oxygen sites denotes that there has been a hydrogen bond present at least once. The donor site shows a pie chart of the relative probability of finding a hydrogen bond in this direction. It is interesting to see that the data is very similar for the two independent runs that have been performed for the c-cut surface. Even though the initial solvent configuration of the trajectories was completely different and uncorrelated, the relative probabilities are very similar. This hints towards sufficient sampling at least for this type of interaction.

When preparing the simulation at this level of theory, the main question remains whether the calculation can be assumed to be equilibrated. This is particularly important for the comparably short trajectories at hand. However, determining whether a time series is converged without continuing the time series is hard. Therefore, Figure 3.23 is only a hint towards the equilibration region of the MD calculations. The figure shows the deviation of the average layer position from the final layer position. For all elements H, O, and Fe, the final average position is obtained after about 500 fs. The final variance resulting from the thermal fluctuations of the system is obtained shortly after at about 600 fs. This means that the equilibration time period of 2 ps should be sufficient to ensure structural equilibration.

Surface Pattern Estimator Since the ab-initio MD is computationally expensive, it is desirable to find a way to estimate the surface pattern distribution from the observations found in the current trajectory. This could be used for showing convergence of surface patterns to estimate how long a trajectory has to be to cover all physically relevant surface

patterns. Moreover, this distribution could help extrapolating to other surface structures that have not been observed in the trajectory so far. Finally, a model of the surface pattern distribution could be linked to the corresponding energies over the probabilities which then could help quantifying contributions to the surface termination energy that would be hard to converge otherwise.

As mentioned before, the surface termination pattern is a set of binary states for each site consisting of a OH group. Under a number of assumptions, this can be treated with a 2D isotropic and hexagonal Ising model.[235,236] The required assumptions are:

Hexagonal Pattern The oxygen sites are roughly hexagonal in the bulk hematite minimum energy conformation already. The area covered by the oxygen atoms in the MD due to thermal fluctuations covers the spots the atoms should occupy in the regular ideal hexagonal lattice.

Isotropy Of course, HB are directional. For the Ising model, it has to be assumed that the shape of a OH group can be well approximated by a single sphere. Due to the small effective radius of the proton, this approximation may be acceptable despite it is a stronger approximation than the first one in this list.

Interactions Second-nearest neighbour interactions have to be small, as the model itself only incorporates nearest neighbour interactions. While it is hard to prove the validity before testing the model, it is much easier to do it afterwards: if second-nearest neighbours (or any higher order) is of significant relevance, the model cannot be self-consistent*.

State-depending Coupling While there are many different variants of the Ising model, some including different parameters for the coupling of neighbours depending on their binary value (meaning different couplings for states $(0,0)$, $(0,1)$, $(1,1)$), this is to be avoided in this case to avoid overfitting. Increasing the number of free parameters almost always improved the quality of fit but may lead to less predictive power of the overall model.

In the general Ising model, the total energy function is given as

* Self-consistent estimation means that the model should predict patterns to be likely that have been included in the training set as being particularly likely.

$$H_{\alpha,\beta}(\boldsymbol{\sigma}^{\mathbf{k}}) = \alpha \sum_{ij} \sigma_i^k \sigma_j^k + \beta \sum_j \sigma_j^k \tag{3.19}$$

$$:= \alpha A(\boldsymbol{\sigma}^{\mathbf{k}}) + \beta B(\boldsymbol{\sigma}^{\mathbf{k}}) \tag{3.20}$$

where $\boldsymbol{\sigma}^{\mathbf{k}} \in \{-1, 1\}^N$ is the binary vector of the surface termination states for the N sites for pattern k. Subscript indices thereof iterate over the vector entries. The actual order is of no relevance to this model. $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are the model parameters. α can been seen as the energy required to create a domain wall between the binary states, β is the equivalent of an external field acting on the sites creating a imbalance between the probability of sites being in either binary state. Using the partition sum, this energy relation can be transformed into a probability $p_{\alpha,\beta}$ normalized by $z_{\alpha,\beta}$:

$$p_{\alpha,\beta}(\boldsymbol{\sigma}^{\mathbf{k}}) := \frac{\exp(-H_{\alpha,\beta}(\boldsymbol{\sigma}^{\mathbf{k}}))}{z} \tag{3.21}$$

$$z_{\alpha,\beta} := \sum_{k'} \exp(-H_{\alpha,\beta}(\boldsymbol{\sigma}^{\mathbf{k}'})) \tag{3.22}$$

At first glance, this formulation of the model is convenient, because the probabilities directly relate to the frequencies of observations of certain patterns in the trajectory. However, having this number alone is not sufficient for estimating the parameters α and β from the frequencies. The solution employed in this case is the so-called maximum log-likelihood estimator (MLLE). The basic idea is to rewrite the problem such that the parameters of Eqn. 3.21 become variables to an optimisation problem with all the observations and, thereby, their frequencies as parameters. This switch of perspective works by calculating the set of α and β that are most likely to give the observations actually made in the trajectory.

The general theory of the MLLE for this problem is outlined as follows: let there be a target probability function f with parameter vector $\theta = (\alpha, \beta)$:

$$f(\cdot|\theta) := p_{\alpha,\beta} \tag{3.23}$$

This target function f depends on the actual observation $\boldsymbol{\sigma}^{\mathbf{k}}$. With the set of all n observations $\boldsymbol{\Sigma} \in \{-1, 1\}^{N \times n}$ the total (not necessarily normalised) probability is the product of

these target functions

$$f(\mathbf{\Sigma}|\theta) = f(\sigma^1) \cdot f(\sigma^2) \cdot \ \cdots \ \cdot f(\sigma^n) \tag{3.24}$$

$$= \prod_i f(\sigma^i) =: \mathcal{L} \tag{3.25}$$

Now $\mathcal{L}$ is called the likelihood function that gets maximised over the further course of the calculation. This requires an analytic model of the probabilities which is given in Eqn. 3.21 and can be inserted into the definition of $\mathcal{L}$. Consequently, the partial derivatives of $\mathcal{L}$ are set to zero to find extremal values. It is mathematically equivalent to maximise the log-likelihood function $\ln \mathcal{L}$, which is a commonly used trick to simplify the expressions defining the conditions of the maximum for the parameter space.

For the first parameter, α, the extremal condition is

$$\frac{\partial \ln \mathcal{L}}{\partial \alpha} = \frac{\partial}{\partial \alpha} \ln \left[\prod_i f(\sigma^i|\theta) \right] \tag{3.26}$$

$$= \frac{\partial}{\partial \alpha} \sum_i \ln \left[\exp(-\alpha A) \cdot \exp(-\beta B) \cdot z^{-1} \right] \tag{3.27}$$

$$= \sum_i \left[-A(\sigma^i) - \frac{\partial}{\partial \alpha} \ln z \right] \tag{3.28}$$

$$= -N \frac{\partial}{\partial \alpha} [\ln z] - \sum_i A(\sigma^i) \tag{3.29}$$

with N as the count of patterns. Equivalently, N can be seen as the sum of weights of distinct patterns. In the actual implementation, N is treated as the sum of the pattern weights for convenience. It is helpful to solve part of this equation separately

$$\frac{\partial}{\partial \alpha} [\ln z] = \frac{1}{z} \frac{\partial}{\partial \alpha} z \tag{3.30}$$

$$= \frac{1}{z} \sum_j \frac{\partial}{\partial \alpha} \exp(-\alpha A^j) \cdot \exp(-\beta B^j) \tag{3.31}$$

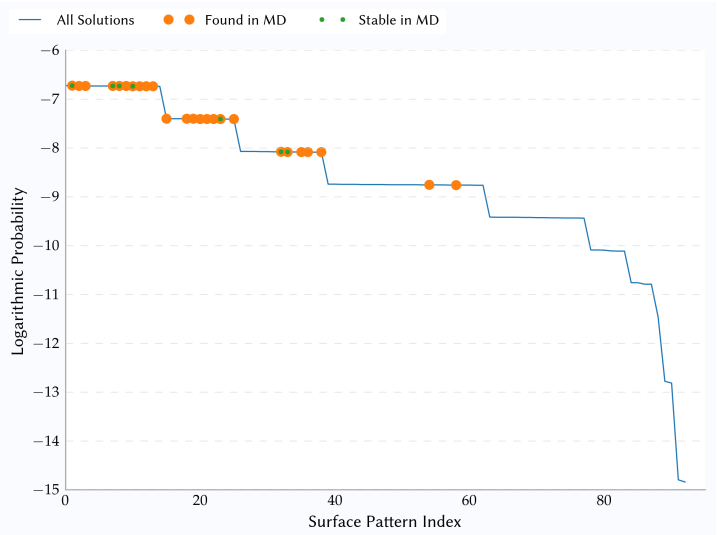
$$= \frac{1}{z} \sum_j (-A^j) \exp(-\alpha A^j) \cdot \exp(-\beta B^j) \tag{3.32}$$

Inserting this result in eqn. 3.29 gives the final condition

$$\frac{\partial \ln \mathcal{L}}{\partial \alpha} = N \frac{1}{z} \sum_j A^j \exp(-\alpha A^j) \cdot \exp(-\beta B^j)$$

$$- \sum_i A(\sigma^i) = 0 \tag{3.33}$$

The calculation for the second parameter, β, works the same way. Since the values $A(\sigma^i) := A^i$ and $B(\sigma^i) := B^i$ are only dependent on the pattern, they can be precalculated making



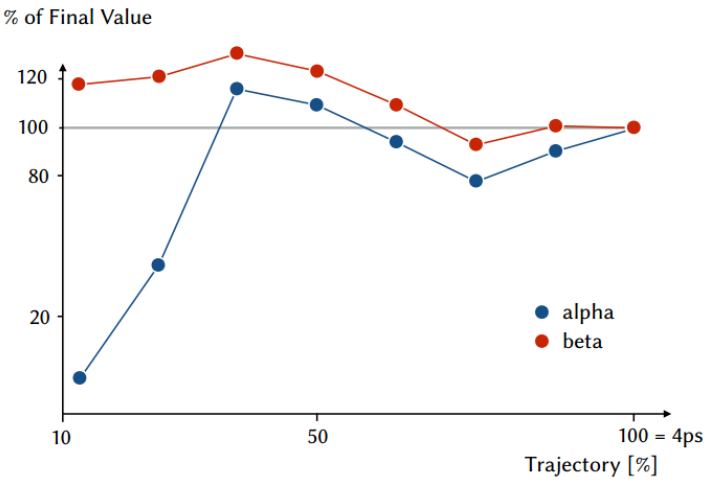
◀ **Figure 3.24** – Results of the MLLE estimation for the Ising model based on an isotropic hexagonal pattern. All possible distinct surface patterns are on the x axis, sorted by their probability (y axis). Highlighted patterns are observed in the MD (orange) and a subset thereof is stable for over 0.7 ps (green).

Calculation(s) 4d9

the actual optimisation computationally cheap. With the help of the analytic gradients, a fast steepest-descent algorithm has been implemented.

With these relations, the parameter vector θ can be calculated for a given set of observations and their corresponding frequencies. There is a significant caveat here: the observations are only for a specific realisation of a surface pattern. However, there are many equivalent representations for this pattern—a consequence of the periodic boundary conditions of the physical system and the high internal symmetry of the hexagonal structure[237] assumed for the Ising model. It is crucial to the efficiency of this approach to distribute the observed probability equally over all equivalent representations before evaluating the expressions for α and β.

Figure 3.24 shows the result of this approach. First of all, the high level of symmetry of the underlying system in conjunction with the periodic boundary conditions of the hematite setup only allows for 93 distinct surface patterns that cannot be mapped on any other of the patterns. The effect of the symmetry considerations is strong here, since there are 4096 possible surface termination states in total per surface of the hematite slab. It is clearly visible that the model is self-consistent in a way that all surface states that have been observed in the MD and, therefore, been used in the fitting procedure to build the underlying model, are predicted to be likely from the model. For example, neglecting the symmetry operations mentioned above give a completely different and, more precise, not self-consistent picture for the same property. There are a few surface patterns that are considered to be likely as from the model, as there are gaps between the surface patterns found



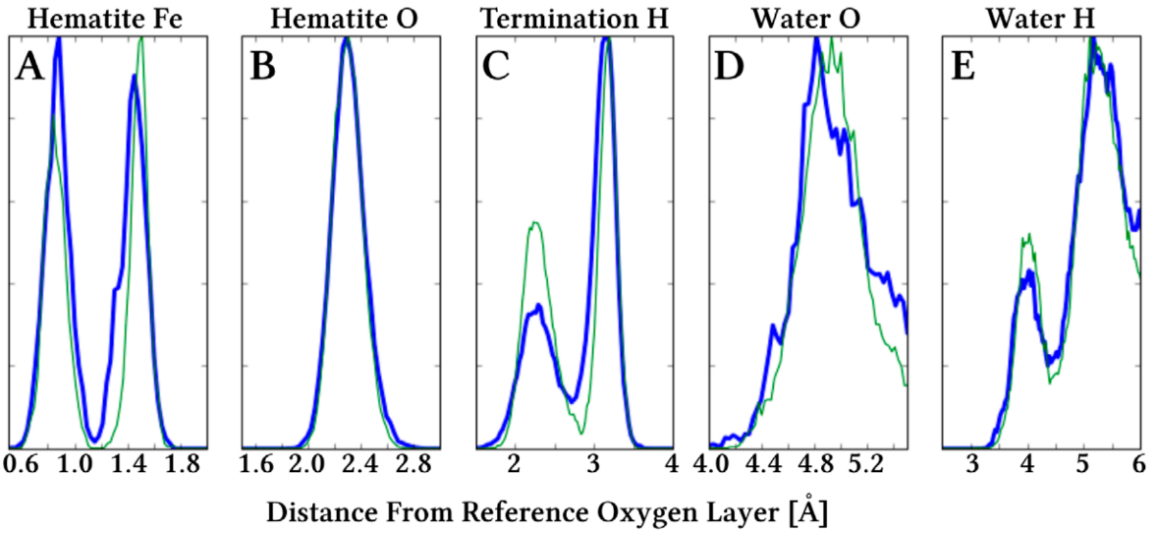
▶ **Figure 3.25** – Convergence of the two parameters of the Ising model over the course of the full MD trajectory. Data shown in percent of the final value which is estimated for the whole trajectory.

Calculation(s) 4d9

in the MD in Figure 3.24.

Whether predictions with shorter MD trajectories would yield similar results is shown in Figure 3.25. It is clearly visible that even for much smaller segments of the final MD calculation, the estimated model parameters are close to the final ones. Most interestingly, the amount of trajectory required to stabilise the model parameters is of same magnitude of the the amount of trajectory required to describe the flip between stable surface protonation patterns.[238] Since the α parameter of the model is related to the probability of having two identical motives (OH in-plane or OH out-of-plane) on adjacent sites, it is plausible that the dynamic picture of the switching events is only captured on the appropriate time scale. For the β parameter, which is related to the share of in-plane sites, the value stabilises quickly, which again corresponds to the observation from the MD trajectory that the interconverting patterns have roughly the same share of in-plane sites, but vastly different surface configurations.[238] Under the caveat of partial coverage of the second nearest-neighbour interactions, this is a promising result likely to be useful for finding other similar surface termination patterns with much shorter trajectories of the all-atom representation.

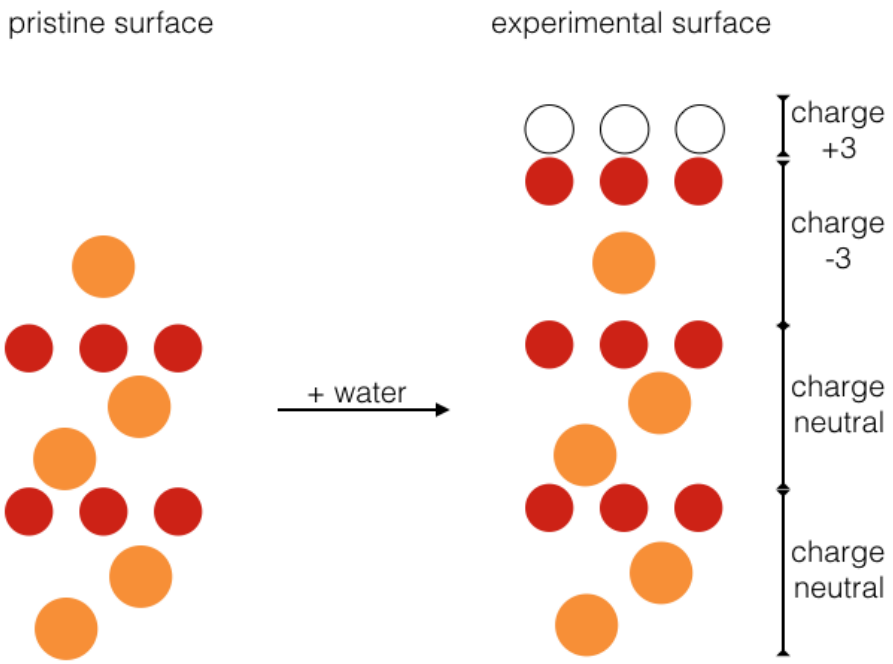
Finite Size Effects and Functional Dependence The setup typically used in this work, the $2 \times 2 \times 1$ super cell of hematite is at the limit of what was feasible with the given budget. To estimate the finite size effects, a system with larger surface area at the expense of hematite layer thickness has been prepared and simulated following the same protocol as for the regular system. In this setup, two iron layers and two oxygen layers have been removed from the hematite slab. Because of the antiferromagnetic spin pattern of hematite, it is not possible to remove just one pair of layers. Moreover, the oxygen sites



▲ **Figure 3.26** – Finite size effects on the atom number density. Thin green curve shows data from the $2 \times 2 \times 1$ super cell and the thick blue curve shows the data for the $3 \times 3 \times 1^*$ super cell with more than double the surface area for the oxygen terminated (001) surface. Reference oxygen layer is the first subsurface oxygen layer of bulk hematite. Image reused from my CC BY publication.[206]

Calculation(s) 696, dc9, d0e, 55e

of bulk hematite are only nearly symmetric under inversion due to the slight distortion of the iron atom coordination. This makes it preferable to remove two layer pairs to obliterate the need to invert the coordinates to map the two surfaces into one single analysis. This flattened but enlarged system is referred to as $3 \times 3 \times 1^*$ system. It is slightly more expensive than the regular system. Figure 3.26 shows the difference in the atom number density for the two setups. In total, two differences are visible: the distribution of protons from the surface termination is different and the mobility of the iron atoms from hematite is increased, giving rise to broadened atom number density peaks for the $3 \times 3 \times 1^*$ system. The peak positions are not significantly changed. The two differences are connected: since the position of the subsurface iron atoms is crucial to the hydrogen bonding network as shown in this work, their higher mobility makes them more likely to push protons into the out-of-plane configuration, which is exactly the change that is observed. The higher mobility of the subsurface iron atoms is attributed to the thinner hematite bulk layer rather than to the larger surface area.

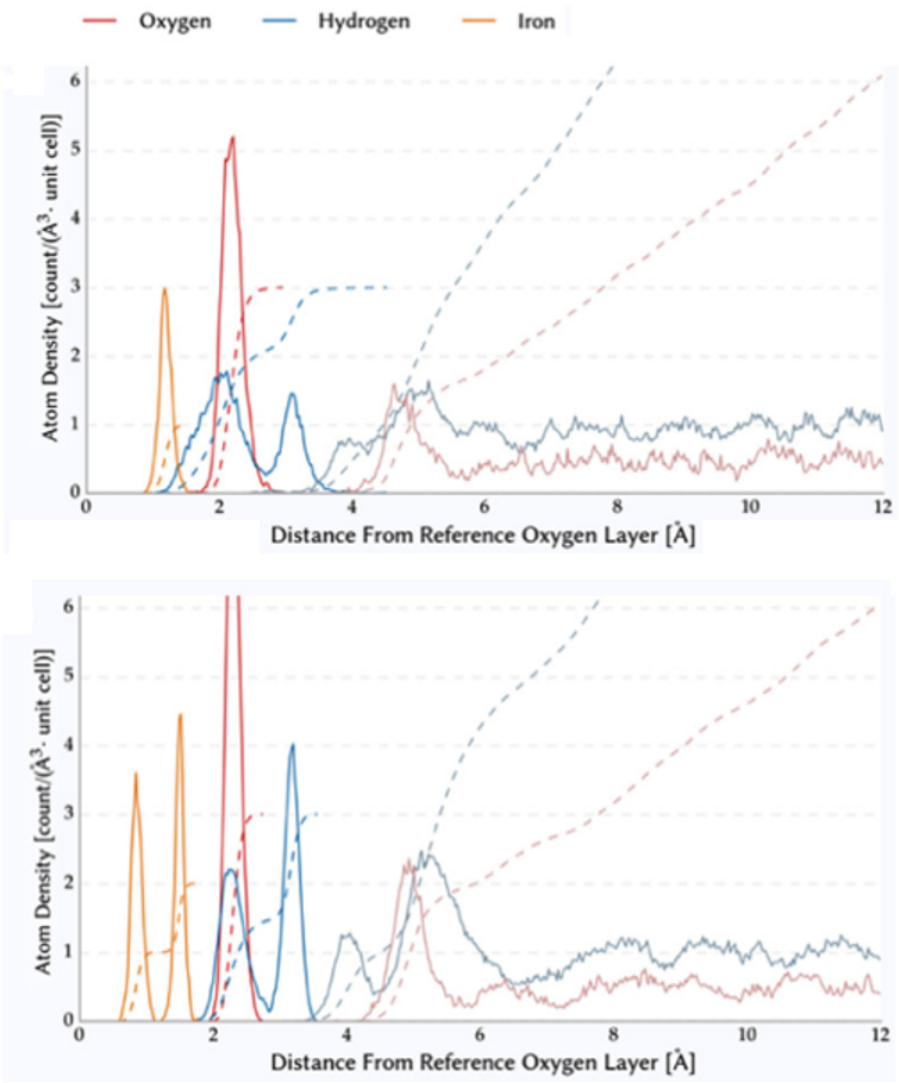


▲ **Figure 3.27** – Protonation of the experimental iron termination. The pristine iron termination (left) is experimentally found to be oxygen covered after solvatisation. The primitive cells of hematite are neutral but the outermost block is missing one iron to be neutral. The formal charge of 3 has to be recovered by singly protonating each surface oxygen.

3.3.1.2 Iron Termination

First, it is important to highlight that the iron termination mentioned here is the solvated iron termination, i.e. the pristine surface[?] after it has been brought into contact with water. Experimental studies[31] have found that the pristine iron termination is covered with oxygen atoms during solvation. Since the aim of this work is to describe the equilibrated structure of this interface, the experimentally confirmed structure has been chosen as starting point for the setup (called two-domain structure in the aforementioned reference).

Figure 3.27 shows the impact of solvatisation on the pristine iron-terminated surface. Upon wetting of the surface, experimental results[31] show that this originally iron-terminated surface is covered with oxygen atoms instead. Since X-ray experiments cannot resolve protons due to their low scattering cross-section, the surface has to be protonated such that the conditions from experiment are recovered. Since the primitive cells of hematite have a zero net charge, the topmost unit that is exposed to water which missed one iron compared to the bulk configuration has a formal charge of -3. To compensate for that, three protons have to be distributed per surface unit area. The natural way of doing so is to protonate every chemical identical oxygen once, which is what has been done in this work.

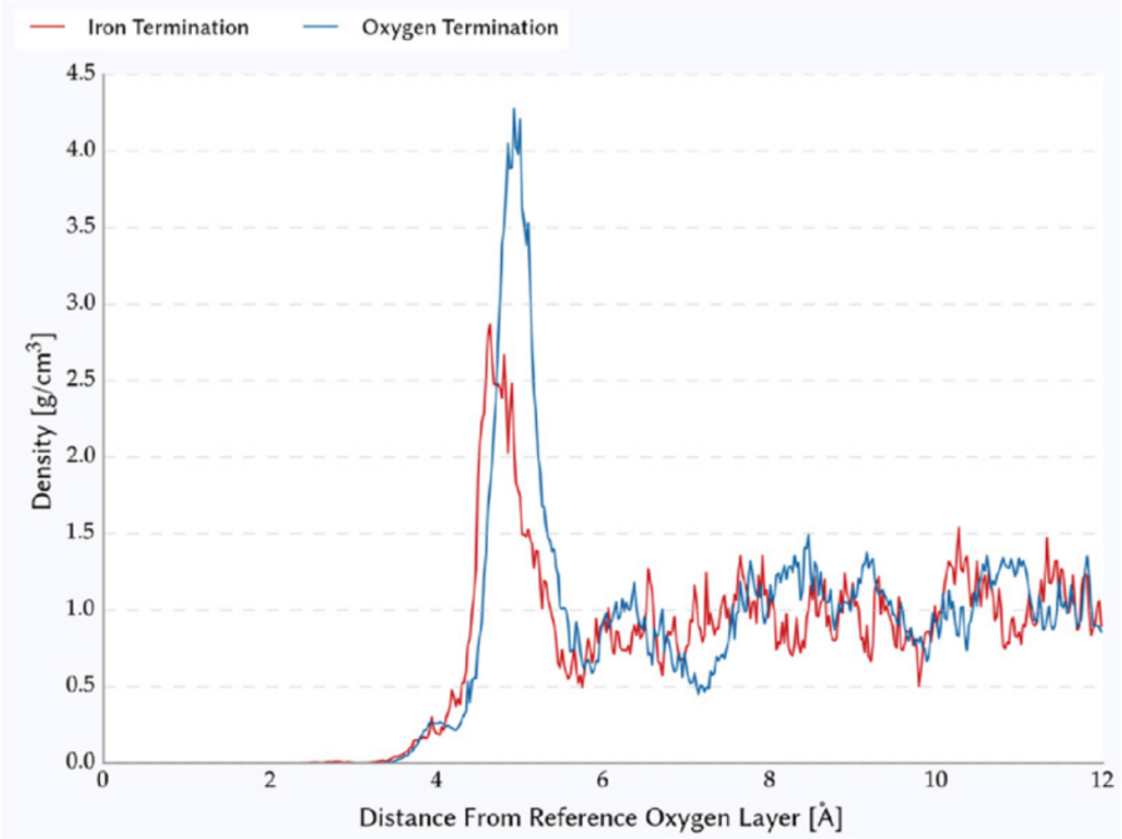


▲ **Figure 3.28** – Atom number density profile for the iron termination (left) and the oxygen termination (right). Dashed lines show the integral of the atom number density of same colour. Shaded lines represent the atom number density for bulk water. Image reused from my CC BY publication.[206]

Calculation(s) e49, 22c, 6cf

Figure 3.28 allows for direct comparison of the atom number density as obtained for the two terminations that have been investigated. The most notable difference is the distribution of the protons of the surface OH groups. For the iron termination, the peak for the in-plane configuration of the surface OH bonds is much wider than for the oxygen termination. This is a consequence of more space being available due to the reduced number of iron atoms in the subsurface iron layer. Moreover, the ratio between the population of the in-plane and out-of-plane configuration is essentially different: for the iron termination, the ratio is 2:1 while the oxygen termination exhibits a 1:1 split. This means that for the iron termination, significantly less protons are available for interaction with the solvent. The solvation shell is of similar thickness for both terminations. Figure 3.29 illustrates this in direct comparison for the two terminations. While the oxygen termination has a stronger pronounced first solvation shell, the iron termination generally features less structure in water, which is attributed to the difference in the hydrogen bond network at the surface.[206] The shift between the two peaks of the first solvation shell results from the different interlayer spacing between the iron layer and the surface oxygen layer due to the different number of iron atoms in said iron layer.

An interface effectively interrupts the hydrogen bond network on bulk water at the surface. The exact way how this happens strongly depends on the internal structure of the solvation layer. One indicator for the stability of hydrogen bonds at the interface is their life time. To reliably determine the life time of volatile hydrogen bonds is a particular challenge[232,239] mostly due to the somewhat arbitrary geometric criterion detecting a hydrogen bond.[240,241] In the following, a hydrogen bond is considered to be present, if the O-O distance is shorter than 3.5 Åand the O-H-O angle is at least 160 degrees. In order to allow fo the temporary violations of this criterion for at most the time τ as motivated by experimental results related to libration motion of water hydrogen bond networks,[233,242,243] τ was chosen such that hydrogen bond lifetime distribution[233] for bulk water is recovered for the bulk-like part of the simulation. In this work, two approaches have been followed: calculate the average lifetime of all uninterrupted hydrogen bonds of a unique combination donor-proton-acceptor or calculate the autocorrelation of the number of uninterrupted hydrogen bonds as time series, a method proposed by Stillinger.[239] The results are shown in Table 3.3. While there are small differences between the relative lifetimes as computed by

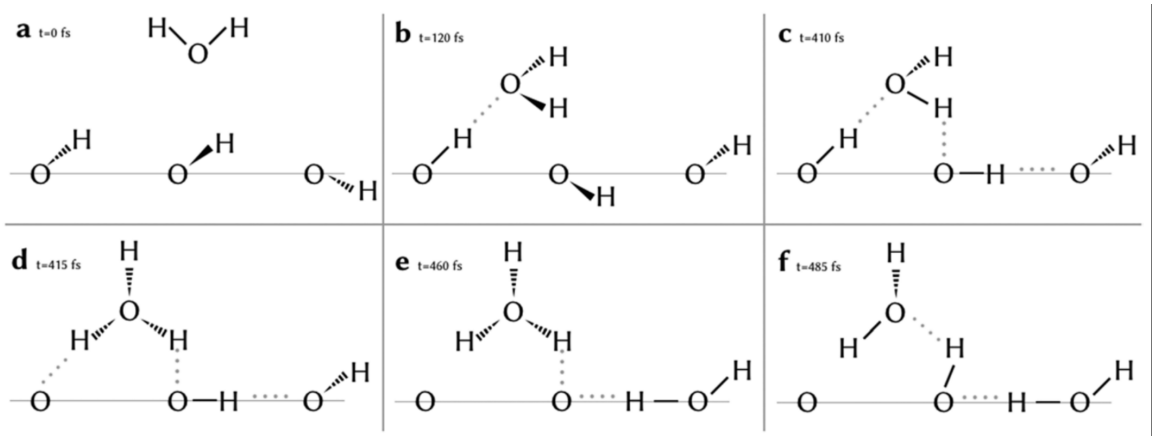


▲ **Figure 3.29** – Atom number density profiles for the oxygen atoms in bulk water for the two terminations investigated. Plane of reference is the plane fitted through the average positions of the second subsurface oxygen layer of hematite. Image reused from my CC BY publication.[206]

Calculation(s) e49, 22c, 6cf

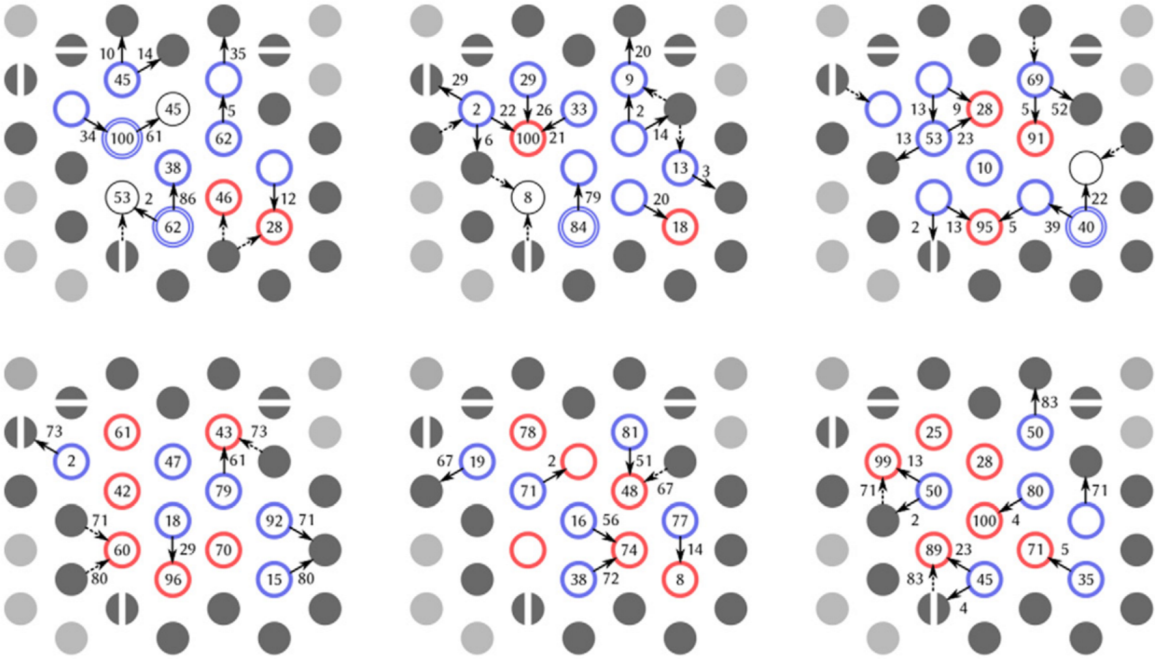
these two methods (this can be seen as a measure of accuracy of the methods), the overall picture is similar: hydrogen bonds from water to hematite have a substantially shorter lifetime fo the iron termination than for the oxygen termination. This difference is considered to be a direct effect from the the higher likelihood of finding a proton parallel to the surface for that termination and the lower surface coverage of the iron termination. In this case, the PES for adsorbed water molecules is flattened which increases the mobility of these molecules and, hence, lowers the lifetime of hydrogen bonds. Low coverage of a surface is typically a cause for increased disorder at the surface[244] for systems where the hydrogen bond within water is not much stronger than the surface-water hydrogen bond,[245] which is the case for hematite, as shown in Table 3.3.

During the AIMD simulation, the protonation of individual sites in general is free to change. For the oxygen termination however, this is never observed: every singly proto-



▲ **Figure 3.30** – Stages of the autoionisation process observed for the iron termination surface. The grey line denotes the surface of the hematite slab. Image reused from my CC BY publication.[206]

Calculation(s) 6cf, e47, ab3, 10d, 0da, 6a0, 87b, 523



▲ **Figure 3.31** – The most stable surface hydrogen bond patterns for the iron termination (top row) and the oxygen termination (bottom row). Symbols identical to Figure 3.16. Sites with thin black circles are not protonated. Sites with two circles are doubly protonated. Probabilities are conditional probabilities, that is, they state how likely it is to observe a particular hydrogen bond given this particular surface pattern. Any probabilities below 2 % are omitted for clarity. Image reused from my CC BY publication.[206]

Calculation(s) 6cf, e47, ab3, 10d, 0da, 6a0, 87b, 523

Donor	Acceptor	τ Fe [%] Stillinger	τ Fe [%] Average	τ O [%] Stillinger	τ O [%] Average
Water	Water	100	100	100	100
Water	Hematite	31	34	83	82
Hemati	Water	116	89	114	75
Hemati	Hematite	20	31	19	48

▲ **Table 3.3** – Lifetime τ of hydrogen bonds relative to the lifetime of a hydrogen bond within water for the two terminations (*iron* and *oxygen*). Data is shown for both methods (*Average* and *Stillinger*).

nated surface oxygen remains singly protonated for the duration of the whole simulation. This picture is very much different for the iron termination of the c-cut. While all surface oxygen sites have been singly protonated as well, the protons quickly rearrange and form doubly protonated oxygen sites (-OH_2^+), thereby creating oxyanion (-O^-) sites where the protons originated from. This configuration then is stable until the end of the trajectories and has been observed at least once on every surface (note that every single trajectory has two surfaces) simulated in independent setups. In one case, this conversion took place twice on a single surface. The resulting stable configurations are shown in Figure 3.31. In none of these configurations, the net protonation of the surface changes.

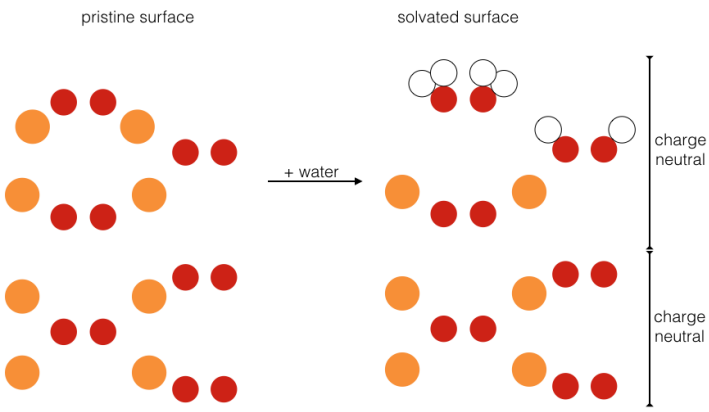
For this transfer, two pathways are observed. Either there is a direct migration of a proton from one surface oxygen site to another neighbouring surface oxygen site, or there is a solvent assisted process. The direct migration requires a hydrogen bond to be formed as precursor of the migration. Since the distance of the surface oxygen atoms is favourable for the formation of a hydrogen bond, there are many opportunities for this transfer. This is likely the reason why this mechanism dominates: all but one of the aforementioned formations of a doubly protonated -OH_2^+ site at the surface employ this pathway.

The latter case, the solvent assisted migration of a proton, involves six steps and an intermediate hydronium ion (see Figure 3.30). At first, a solvent water molecule approaches the surface (step a) and is oriented to be a hydrogen bond acceptor from a singly protonated oxygen site at the surface (step b). In the next step, a cooperative chain of hydrogen bonds is formed between the solvent molecule and two other singly protonated oxygen sites at the surface (step c). This cooperative chain of hydrogen bonds from the first surface oxygen site via the solvent molecule to the second surface oxygen site is stabilised due to polarisation effects[232,246] and is the precursor of the hydronium ion (step d) which is held at the surface by two hydrogen bonds. Although hydronium hydrogen bonds are stronger

than regular water hydrogen bonds,[247] the newly formed hydrogen bond in the inverse direction of the hydrogen bond in the precursor geometry breaks shortly after (step e). After 50 fs–the typical for the lifetime of a hydronium ion in water[248,249]–the hydronium ion looses one proton to the (now doubly protonated) second surface oxygen site. This way, charge has been transported, but no mass, since the proton of the (now bare) first oxygen site now is part of a solvent water molecule. There are accounts of similar mechanisms on a TiO_2 surface.[250,251] During the presence of hydronium, there is an additional solvent water molecule above the hydronium ion which is close to the Eigen cation which has been shown to enhance proton hopping.[252]

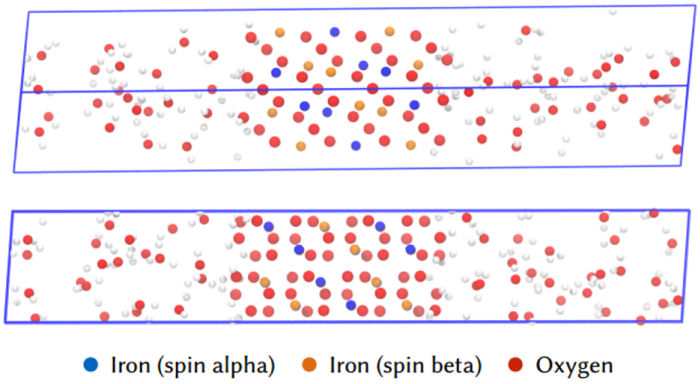
The mechanism described above is similar to the Grotthuss mechanism and shows that the presence of hydronium ions can influence the surface protonation.[253] While the overall mechanism was observed only once, the total duration of 0.5 ps shows that this process is fast and involves an hydronium ion that is of transient nature only. Without the inclusion of quantum nuclear effects, the frequency of autoionisation is expected to be underestimated.[254] Experimental evidence lists short hydrogen bond lifetimes,[255] cooperative hydrogen bonds,[256] low water surface coverage,[79] and reduced OH group coverage at the surface[257] as supporting factors for a high proton mobility, all of which are present in this system. The direct transfer has been observed for quartz,[258] while the resulting stable[259] Fe=O sites have been found experimentally,[74,75] as well.

Functional Dependency For the iron termination, the functional impact has been evaluated by performing calculations at the PBE+U level with eight independent trajectories (6cf, e47, ab3, 10d, 0da, 6a0, 87b, 523) covering 9.5 ps surface dynamics. Seven of the trajectories show spontaneous deprotonation of neighbouring surface oxygen sites, dissociation of them and formation of solvated O_2, which exposes bare iron sites to the solvent. Since this surface structure is energetically favoured for the bare surface without solvent molecules for PBE+U and PBE level only,[73,80] but experiments see oxygen atoms exposed to the solvent,[17,31,212] it can be concluded that PBE+U does not reliably reproduce the experimental surface termination.



◀ **Figure 3.32** – Protonation scheme upon solvation of the pristine r-cut as obtained from calculations and experiment.[36,260]

Calculation(s) 318



◀ **Figure 3.33** – Snapshot of the r-cut setup. For better orientation, the iron sites with opposite majority spin moment are highlighted in different colours (blue, orange). The hematite layer is about 10 Å thick, while the water layer spans about 35 Å. The blue box represents the fully periodic simulation cell.

Calculation(s) 318

3.3.2 Hematite r-cut (1-102)

The hematite r-cut has been investigated extensively in previous studies,[36] in particular in regards to the stability of surface structures. Based on their work, the pristine crystal cut is expected[36] to have a missing surface iron layer, as illustrated in Figure 3.32. This has been confirmed experimentally.[260]

Similar to the calculations of the c-cut, the more reactive[261–263] r-cut has been investigated. For a description of the system and the setup procedure, see page 57. For the purpose of this section, the most important difference between the r-cut and the c-cut discussed before are the surface oxygen atoms. While the c-cut in both terminations (oxygen termination and iron termination) only features chemically identical oxygen sites, the r-cut surface has three different oxygen sites, denoted $^{\mathrm{I}}$O, $^{\mathrm{II}}$O, and $^{\mathrm{III}}$O.

For the r-cut in Figure 3.33, the protonation scheme is not entirely clear,[29,207] although it is commonly speculated[36] that of the three chemically distinct oxygen sites that are (mostly) exposed to water, the one being exposed the most is doubly protonated, while the one closest to bulk hematite is not protonated at all. The middle site (as seen perpendicular to the hematite surface) is expected to be singly protonated. Later sections of this work go

Layers	d Experiment[260] [Å]	d Experiment[29] [Å]	d Simulation [Å]
^{I}O–^{II}O	1.28 ± 0.07	1.25 ± 0.06	1.260
^{II}O–^{2}Fe	0.64 ± 0.04	0.69 ± 0.04	0.934
^{2}Fe–^{III}O	0.48 ± 0.04	0.38 ± 0.03	0.436
^{III}O–^{4}O	1.39 ± 0.07	1.41 ± 0.04	1.410
^{4}O–^{3}Fe	0.38 ± 0.04	0.37 ± 0.03	0.406
^{3}Fe–^{5}O	0.84 ± 0.04	0.77 ± 0.03	0.935
^{5}O–^{4}Fe	0.69 ± 0.03	0.77 ± 0.02	0.928
^{4}Fe–^{6}O	0.354	0.354	0.375

▲ **Table 3.4** – Spacings d between layers of hematite in the r-cut for the 2-1-0 protonation along the surface normal axis. Table taken from my published work with permission from ACS.[260]

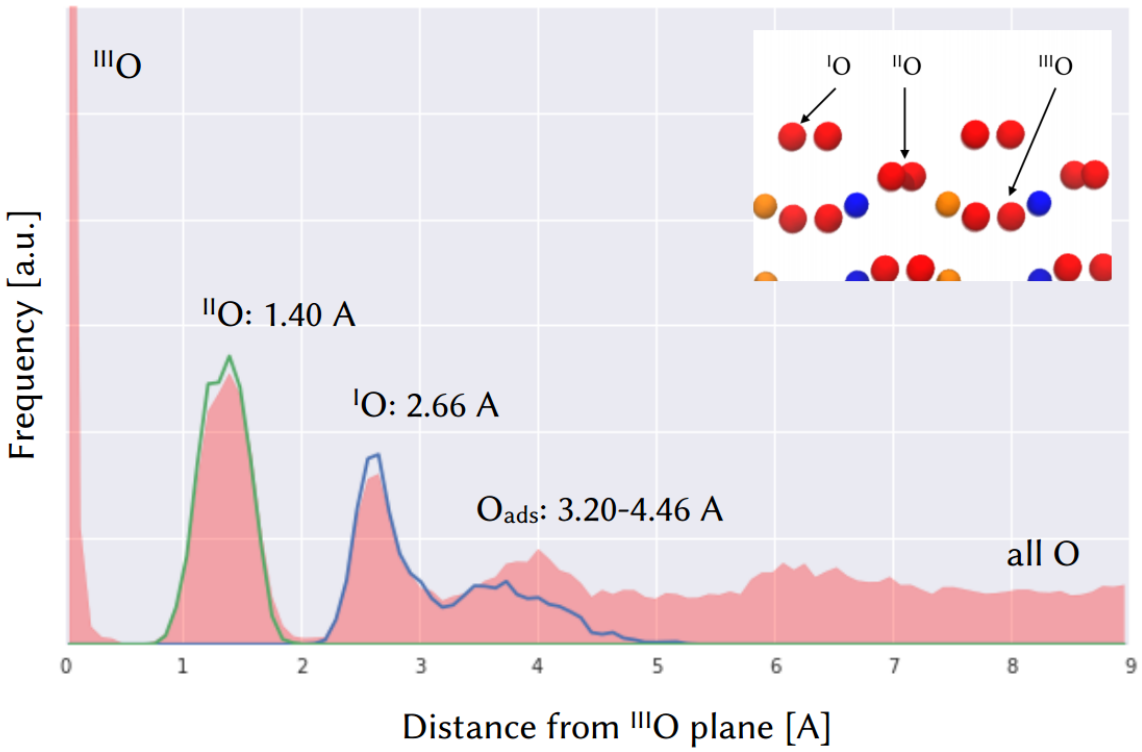
Calculation(s) 318, ec8, 29a, 2b2, 633, 0db, 568, 6fd

into further detail and compare this protonation scheme with the one obtained from AIMD calculations.

All r-cut calculations in this work have been carried out in the unit cell shown in Figure 3.33 which is built directly from the schematic in Figure 3.32.

Layer Structure The best comparison to experiment is available for bulk structures, since the CTR method commonly employed is most sensitive to this part of the system. Table 3.4 compares the data from experiment to the ones obtained from simulation. It is clearly visible that for nearly all layer spacings the results agree very well, the only exception being one of the O-Fe sublayer geometries. However, previous work has shown that these iron sites are the ones most affected by finite size effects.[206]

Because the protonation scheme of the surface is not experimentally known and a matter of debate in literature, two protonation schemes have been evaluated: 2-1-0 and 1-1-1, where the numbers designate the protonation state of the three distinct oxygen sites at the hematite surface going from ^{I}O, the one closest to water to ^{III}O,the one closest to bulk hematite (see inset Figure 3.34). The first protonation has been simulated for 19.5 ps and the other one for 7.8 ps (82e) after appropriate equilibration (ffe, 4c4, a63, 2cc). Convergence has been confirmed by comparing a subset of the atom number density in Figure 3.34 from the first half of the data against the equivalent subset of the second half of data from independent trajectories.



▲ **Figure 3.34** – Atom number density of oxygen atoms (shaded red) for the 2-1-0 protonation (see text). Oxygen atoms initialised as ^{II}O, shown as green curve. Oxygen atoms initialised as ^{III}O, shown as blue curve. Oxygen atoms from bulk water are included in this diagram. Both sides of the hematite slab are averaged here.

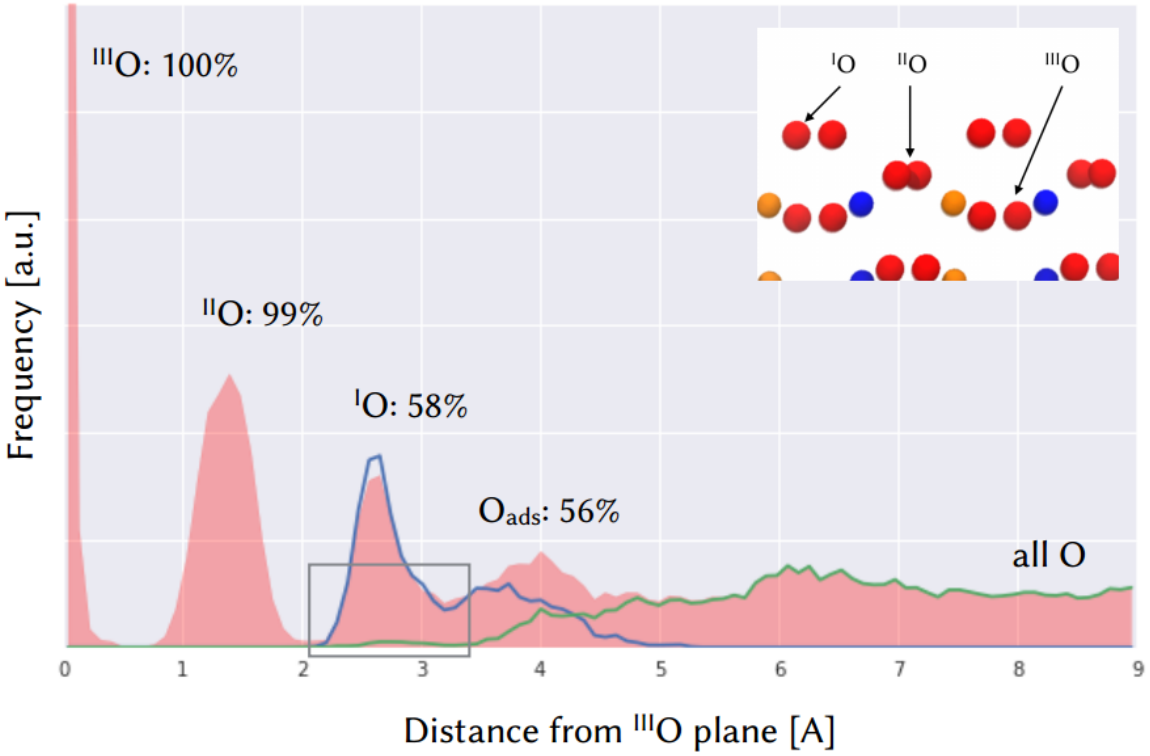
Calculation(s) 318, ec8, 29a, 2b2, 633, 0db, 568, 6fd

Solvation Structure Figure 3.34 shows the distribution of the oxygen atoms at the surface. This is of particular interest for comparison to experiment since the scattering cross section of oxygen atoms is large enough to be picked up in CTR experiments. However, the underlying assumption in the analysis of CTR experiments is that each atom has a stationary average position and that the probability of any single atom to be found around its average position can be modelled by a Gaussian distribution with (optionally) different widths for different axes. Based on the atom number density information obtained from the MD simulation, it is clear that this may be justified for the ^{II}O peak, but not for ^{I}O. In particular, there is a significant share of ^{I}O oxygen atoms that become adsorbed water on the picosecond time scale. Therefore, in direct comparison with experiment, one has to take into account this limited validity for the surface area. However, with the AIMD data this can be partially mitigated: if experiment and simulation agree on the bulk structure and the ^{II}O peak position, this gives some credibility to the AIMD picture of the ^{I}O information and the interchange between aquo groups in the ^{I}O peak and adsorbed or even bulk water.

For the peak positions as outlined in Figure 3.34, there are both experimental[29,30,264] data as well as classical[32,33] MD calculations with ClayFF. While the experiments place the ^{I}O peak at 1.90-2.33 Å from the ^{III}O position, this work places the peak slightly farther out at 2.66 Å. Similarly for the ^{II}O: experiment sees the peak at 1.07-1.12 Å and AIMD places it at 1.40 Å. Classical simulations place ^{II}O at 1.34 Å and ^{I}O at 2.34 Å. This means that there is good agreement for the ^{II}O peak, but significantly different placements for the broader peak ^{I}O. It is shown later in this work that the doubly protonated oxygen atoms in this peak are highly mobile and exchange with adsorbed water molecules. This violates the underlying assumptions of the experimental CTR measurements, which may explain the difference resulting positions.

For water molecules adsorbed at the surface, there is no ClayFF information available, since the corresponding peak is not observed.[32,33] However, experimental work places the adsorbed water peak at 3.2 Å. This value is the lower bound of the distribution of adsorbed water molecules in the MD data in Figure 3.34 which extends to 4.46 Å. Again, adsorbed water is not exactly compatible with the basic assumption of the CTR experiments of having an equilibrium position for every atom similar to a bulk crystal. The very short distance of 0.9 Åbetween the outermost ^{I}O peak and the adsorbed water peak (!) hints towards inconsistent results for surface atoms.

One of the free variables in the analysis of experimental CTR data is the occupancy of given site, i.e. the fraction of equivalent sites being occupied for an infinitely large system sample. The analogue for AIMD data is the average atom number density integrated over the peak positions. Figure 3.35 shows the results for the 2-1-0 protonation. While both the ^{II}O and ^{III}O peak are fully populated, it is interesting to see that after equilibration the ^{I}O peak is only populated to 58%. Since it has been initialised with 100% occupancy when the simulation has been prepared, this means that a significant part of the oxygen atoms (and, hence, the aquo groups) have become adsorbed water. Tracing those oxygen atoms that have been part of this 100% occupied peak upon system initialisation (the thin blue line in 3.35) confirms this. Vice versa, the same happens to water molecules that adsorb at the surface but have been part of bulk water before (the green line in 3.35). Both effects contribute to a population of 56% of the adsorbed water peak. In all cases, the percentages relate to the number of occupied sites per surface unit cell.



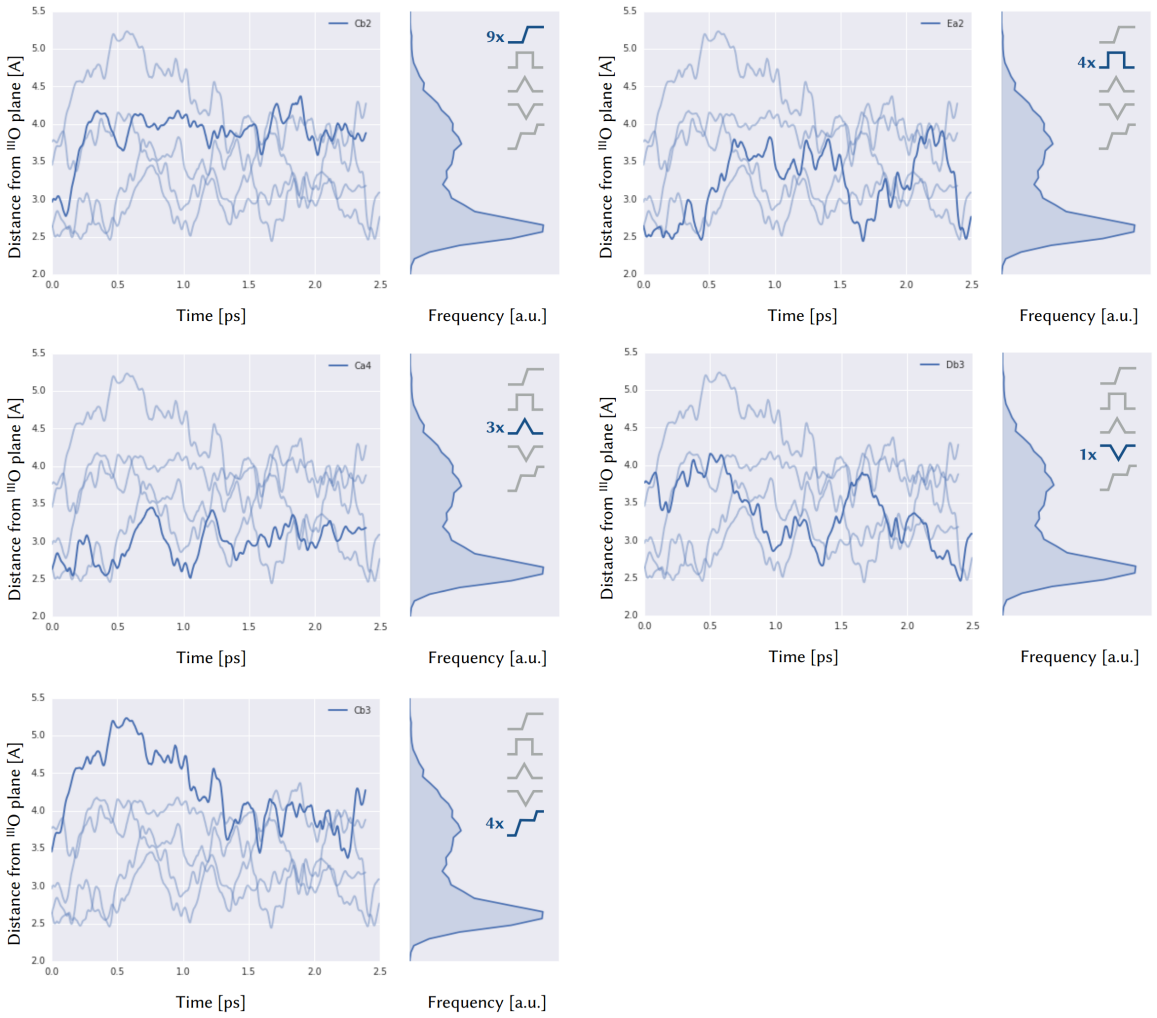
▲ **Figure 3.35** – Occupancy per peak over atom number density of oxygen atoms (shaded red) for the 2-1-0 protonation (see text). Oxygen atoms initialised as water oxygen are shown as green curve. Oxygen atoms initialised as hematite protonation are shown as blue curve. Both sides of the hematite slab are averaged here.

Calculation(s) 318, ec8, 29a, 2b2, 633, 0db, 568, 6fd

The lability of the doubly protonated oxygen atoms is also illustrated by a small peak of atom number density of the oxygen atoms that have been initialised as oxygen atom in Figure 3.35 at the location of the ^{I}O peak. Since placing an atom at the ^{I}O site requires breaking and subsequent formation of a new (weak) oxygen-iron bond, observing this process on the picosecond time scale hints not only towards the mobility of all molecules and atoms involved, but also on a particular weak oxygen-iron bond. The weak bonding assumption is confirmed by fairly extended bond lengths. This is discussed in greater detail below.

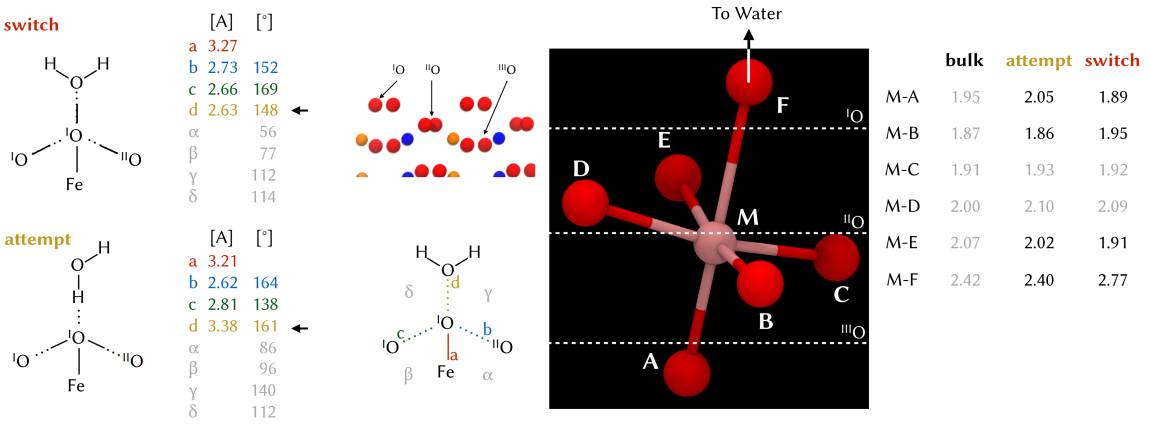
Besides the initial migration of surface aquo groups and bulk water to form the adsorbed water layer (as discussed above), exchange between the ^{I}O sites and adsorbed water is observed.

Figure 3.36 shows representative time series for the motifs observed in the simulation. For the first motif, a ^{I}O becoming adsorbed water, there are nine observations on the picosecond time scale. This comparably high number is in part due to the overall trend of



▲ **Figure 3.36** – Representative time series for individual oxygen atom positions for different motifs. Each panel consists of two diagrams with the raw time series and the atom number density for the ^{I}O and adsorbed water peaks. Insets in the atom number density plots show the number of events observed for this motif. Top left: ^{I}O becoming adsorbed water. Top right: ^{I}O becoming adsorbed water and ^{I}O again. Mid left: ^{I}O attempting to convert to adsorbed water. Mid right: adsorbed water attempting to convert to ^{I}O. Bottom left: ^{I}O becoming adsorbed water and subsequently becoming bulk water.

Calculation(s) 318, ec8, 29a, 2b2, 633, 0db, 568, 6fd



▲ **Figure 3.37** – Comparison of coordinating geometries for sites involved in a conversion of a ^{I}O site into adsorbed water. Bond lengths are shown for all relevant bonds. For hydrogen bonds, the angle formed between donor, proton and acceptor is given. For the metal coordination, bond lengths similar to the bulk value are shaded.

Calculation(s) 29a, 6fd

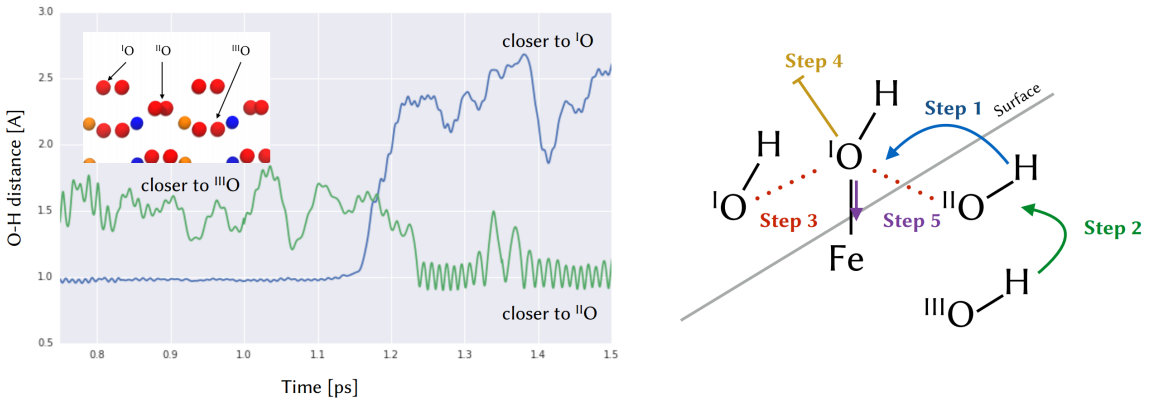
converting ^{I}O to adsorbed water, while the high mobility of the sites in general, as demonstrated by the other motifs in Figure 3.36, contributes as well. Other motifs include temporary conversion of the ^{I}O to adsorbed water, which requires breaking of a weak Fe-O bond.

The low absolute number of events makes it hard to calculate any reaction rates, but at least allows an estimate of the time scales on which the processes outline above take place. The interconversion on the picosecond time scale shown here is an important information when building a CTR model of the interface, since the high mobility invalidates the picture of a static interface. Moreover it is interesting to look into the driving forces for the motions: as shown in Figure 3.36, there are unsuccessful attempts undertaken by doubly protonated oxygen atoms to leave the ^{I}O peak and become adsorbed water and vice-versa. By comparing the successful interconversions to the pure attempts, it is at least possible to devise an environment which is in favour of the interconversion.

Interestingly, all nine successful switches share one topology and all three attempts share another topology. Figure 3.37 compares the two configurations that differ both in their (hydrogen) bond orientation as well as in the distance of the water molecule involved to the surface. In both cases, a solvent water molecule is required. For a successful switch, it acts as hydrogen bond acceptor from the ^{I}O site that becomes adsorbed water later in the process. The other proton of the doubly protonated ^{I}O site forms no strong and stable hydrogen bond for the duration of the switch. Now the ^{I}O oxygen has one strong hydrogen

bond pulling from the water molecule above, one strong hydrogen bond sideways along the surface to a neighbouring ^{I}O, one hydrogen bond to a ^{II}O below as well as a very weak Fe-O bond holding it in place. In this configuration, the hydrogen bond from above exerts the strongest force on the ^{I}O oxygen. For the attempt, however, the ^{I}O is a hydrogen bond acceptor from the solvent water molecule, the hydrogen bond to the neighbouring ^{I}O site is very weak and the one to the subsurface ^{II}O is very strong. While the Fe-O bond is similarly extended and, thereby, weakened,[232,254] the hydrogen bond to the solvent water molecule is much weaker than in the switch case since the the O-O distance is 3.4 Å for the attempt as compared to 2.6 Å in the successful switch. This points towards the solvent interaction being the driving force do convert a ^{I}O site to adsorbed water, or, more precisely, the relative strength of the hydrogen bond between solvent molecule and ^{I}O site as compared to the hydrogen bond between ^{I}O and ^{II}O subject to those configurations where the Fe-O is sufficiently elongated by thermal motion to successfully weaken it. In any case, the extended Fe-O bond of about 3.2 Å alone is not sufficient to enable the conversion.

Experimentally, this is hard to verify, since the process is fast and mostly involves hydrogen bonds. However, there may be a way to indirectly evaluate the process outlined above by analysing the different coordinating geometries for the surface iron site the ^{I}O is bonded to. CTR experiments are more sensitive to positions and coordination of both iron and oxygen than to hydrogen atoms. Moreover, the iron site and the ^{III}O sites hardly migrate and, therefore, are more closely following the underlying assumptions of the experimental analysis than the ^{I}O sites involved. In Figure 3.37, the bond lengths for the average coordination of the iron site are compared for an attempt and for an actual switch. For two bonds, the distance is very close to the bulk equilibrium structure and well within the typical thermal fluctuations around this structure. For four bonds, however, this is essentially different. Firstly, the bond to the ^{I}O site (labelled *F*) in Figure 3.37 is much more extended as compared to the bulk configuration (2.77 Å as compared to 2.42 Å). Consequently, the opposite bond to the oxygen site below the iron atom is shortened for a successful switch (1.89 Åas compared to 1.95 Å). For the remaining two bonds of the sixfold coordination, the changes shown in Figure 3.37 generally move the iron site closer to the surface plane. While the extended Fe-O bond to the ^{I}O site is not sufficient to enable the switching event, the coordination of the iron site may be a useful indicator to search for this process in



▲ **Figure 3.38** – Ordering of the steps involved in a conversion of a ^{I}O site into adsorbed water from a 1-1-1 initial protonation scheme. Left: Time series of the OH distances involved, shows bond formation and breaking. Right: order of events that eventually lead to the ^{I}O oxygen site leaving the surface (step 4).

Calculation(s) ee5, 89a, 0c9, 26a

experimental assessment of the interface in question.

In literature, there is a debate as to which protonation state is stable, although the 2-1-0 state is the most popular.[29] To investigate the relative stability, independent systems with different water solvation structures and different protonation states (2-1-0 and 1-1-1) have been simulated. On the time scale of the 2-1-0 simulation, the average number of protons per oxygen is 1.999 for ^{I}O, 0.994 for ^{II}O, and 0.008 for ^{III}O. This means that the protonation state is not changing much. However, it is important to note that the number of ^{I}O sites decreases quickly by converting ^{I}O groups into adsorbed water. For the 1-1-1 setup, the picture is essentially different. After only 7 ps, the average number of protons per oxygen is 1.197 for ^{I}O, 1.006 for ^{II}O, and 0.813 for ^{III}O. In this case, nearly 20% of the protons from the ^{III}O sites have travelled to the ^{I}O site. This remarkable change can be traced in AIMD simulation.

Figure 3.38 shows both the initial conversion of a 1-1-1 protonation for a single surface unit cell into a 2-1-0 protonation followed by the separation of the then doubly protonated ^{I}O site to become adsorbed water. First, the singly protonated ^{II}O site forms a hydrogen bond to the surface ^{I}O site that (at that time) is singly protonated. This hydrogen bond is the precursor for a doubly protonated ^{I}O site. Right after the proton has switched from the ^{II}O site to the ^{I}O site, the single proton from the ^{III}O is passed on to the ^{II}O oxygen, which now is singly protonated again and leaves the ^{III}O site without proton. All this happens on the sub-picosecond time scale. As for the detaching of the doubly protonated ^{I}O site,

step 3 to 5 prepare the geometry shown in Figure 3.37. First, the hydrogen bond network is reconfigured (step 3), then a solvation water molecule pulls the $^{\mathrm{I}}$O site outwards (step 4) followed by a retraction from the surface of the iron site the $^{\mathrm{I}}$O oxygen is bonded to (step 5).

In summary, this means that the 1-1-1 protonation scheme (where all oxygen atoms $^{\mathrm{I}}$O to $^{\mathrm{III}}$O are singly protonated) is not stable, but the 2-1-0 scheme alone is not stable either. The long-term stable termination seems to be a 2-1-0 protonation where the $^{\mathrm{I}}$O sites continuously interconvert with adsorbed water molecules as shown in the atom number density profile in Figure 3.34.

3.4 Charged Interfaces

Starting from the charge-neutral interfaces described in the previous section, charged interfaces have been investigated to inspect the electronic structure upon electron removal or injection. Both negatively and positively charged interfaces are important in applications of the hematite/liquid water interface, e.g. in water splitting or in crystal growth. A better understanding of the atomistic charge localisation for the charged interfaces can help understanding experimental results for the interface in question.

Practically, this meant setting the system charge to $+e$ or $-e$ and continuing from one of the previous neutral simulations, essentially using them as extended pre-equilibration. Again, two different runs have been performed that originally have started from different surface solvation structures. Changing the total charge of the system leaves the nuclei in a configuration that is not longer their equilibrium. To observe the relaxation of which is part of the following analysis. In total, 15 ps of surface dynamics have been collected.

For each atom of the system, the instantaneous charge is known from the electron density. This requires some kind of charge fitting mechanism that relates the electron density at each point in space to a single atom.[265] Here, Hirshfeld[266] charges have been used for this purpose. Now for each atom in the neutral setup, one can calculate the time-average of said charges which will be henceforth referred to as reference charges.

To get an idea how the reaction of the system to the vertical excitation looks like, a new

Year	Author	Final State	Comments
1986	Fujimori[161]	(d^4) and $d^5\underline{L}$	
1988	Lad[268]	d^4 and $d^5\underline{L}$	$(01\bar{1}2)$ surface. States closer than in Fujimori[161]

▲ **Table 3.5** – Expected XPS final states for the bulk hole.

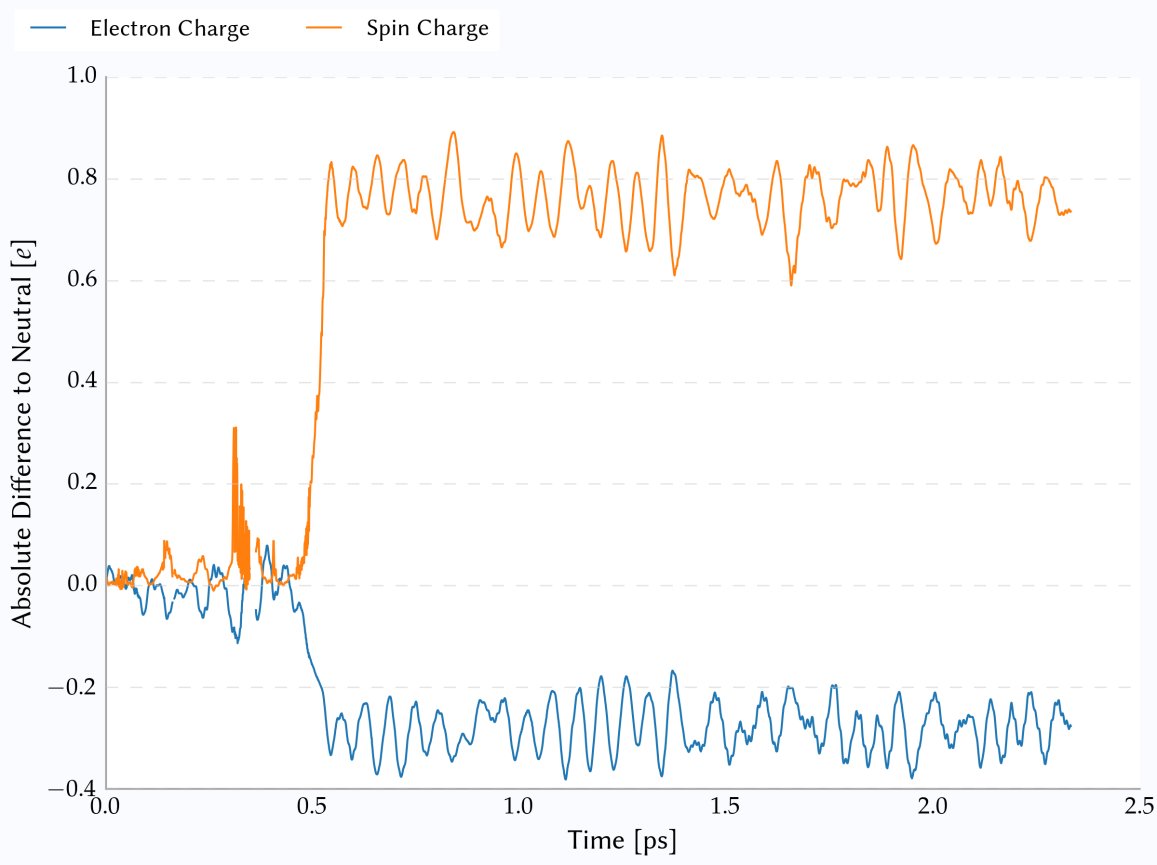
set of charges was derived as follows

$$\hat{q}_i := q_i^C - \langle q_i^N \rangle \tag{3.34}$$

In particular, this definition should help to trace the hole or the excess charge on the system as the MD evolves. While in the beginning of the positively charged simulation the hole should roughly resemble the HOMO of the neutral calculation, this is not necessarily true for the simulation based on this initial frame, as the PES are completely different. The same argument applies to the excess charge that initially resembles the LUMO of the neutral calculation.

In both positively charged calculations that have been carried out, a particular reaction of the system upon the oxidation could be observed. After 250-500 fs from the oxidation, one iron atom carried a charge that is well below the average charge from the neutral simulation, as shown in Figure 3.39. In both calculations that originally started from a different water solvation structure and have performed a neutral MD before the oxidation, the iron atom in question always was the same (up to geometric symmetries), although in a different unit cell. This hints towards this particular site having a supporting environment for charge localisation. This particular site is a sub-surface iron atom. Monitoring the charges from the surface oxygen layers and the iron layer this particular site is found in gives a clear picture of the involved charge transfer from the theoretical side. Experimental results are usually interpreted using semi-empirical methods.[194,267]

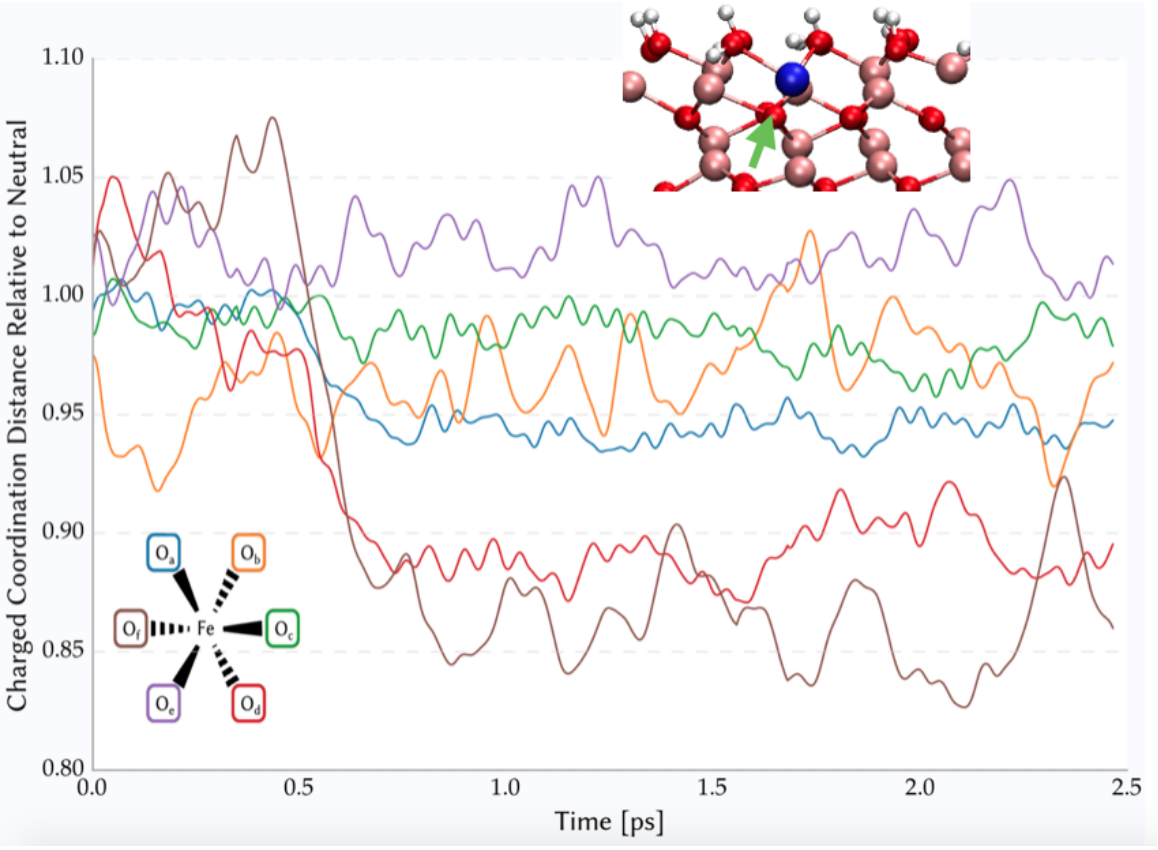
To understand the charge transfer, a description of the electronic structure for the neutral system is helpful. Taking three layers of the hematite structure parallel to the (001) surface, namely two iron layers and the single oxygen layer they enclose, gives a spin-neutral setup. Considering iron $3d$ and oxygen $2p$ only, the energy levels of the iron sites are split due to the octahedral coordination into e_g and t_{2g} with the former being higher in energy. Each of the five iron $3d$ orbitals is singly occupied with the two iron layers having different spin orientation due to the antiferromagnetic ground state of hematite.



▲ **Figure 3.39** – Instantaneous deviation of the positively charged simulation setup value from the average neutral calculation for a single iron site. Both electron charges and spin moments are calculated from the electron density using the Hirshfeld charge fitting method. Oxidation at $t = 0$.

Calculation(s) 3f1, 22c

All oxygen 2p orbitals are fully occupied. Upon removal of a single spin-down electron from an oxygen 2p orbital, the now unpaired spin-up electron in oxygen 2p is transferred to the neighbouring spin-down iron layer, thereby reducing an iron site to Fe^{2+} with the oxygen now missing two electrons compared to the bulk structure. This would constitute a ligand to metal charge transfer (LMCT) as observed for other iron active sites.[269] The charge transfer mentioned above coincides with a change in geometry. In particular, the distances between the iron site in question and the six surrounding oxygen atoms change significantly. As Figure 3.40 shows, two of the distances are reduced by 10-15 % within 0.2 ps. This could be potentially interesting as this change of geometry should be accessible by experiments. This could offer a link between charge localisation from simulations and structural measurements from experiment.[270] Comparing this to the charge picture in Figure 3.39 however makes this result counter-intuitive, since the Fe^{II} site should yield



▲ **Figure 3.40** – Distances to the nearest oxygen atoms for a specific iron atom in the outermost iron layer of the hematite-water layer for the charged system. All distances relative to the average distance for the same atom pairs in the neutral setup. Time series smoothed by exponential window averaging with a half-life time of 60 fs. Inset shows the topology of the coordinating atoms in the hematite surface as seen from the water layer. Inset top right shows a snapshot with the iron site in question highlighted in blue. The green arrow in the inset points towards the oxygen atoms which get closer to the iron site.

Calculation(s) 3f1, 22c

increasing bond distances for the coordinating oxygen atoms. This could potentially be a consequence of the interplay of an electrostatic potential from the periodic crystal and a local conformational change facilitated by the more flexible atoms near the slab surface.

However, the addition of an electron on the iron sites essentially creates a Fe^{II} site, while experiments suggest a Fe^{IV} site for vertical bulk configurations (see Table 3.5). This unexpected different behaviour prompted an investigation of the impact of the functional choice on the results, while a comparison to relaxed hole experiments would be helpful as well. While the initial AIMD calculations employed HSE06 with a HFX share reduced to 12% from the original 25 %, other separate calculations starting from the same initial structure have been started for the excess hole configuration. This time, the HFX share was increased to 25% and 50%, since a higher share of HFX should yield a stronger localisation.

The question these simulations were supposed to answer is whether the charge distribution depends significantly on the functional employed. While short simulations have confirmed that the charge localisation is stable even with higher shares of HFX (up to 50 %), these calculations would show whether this behaviour is observed spontaneously for the other functionals, as well. At least in the literature, there are hints towards charge transfers happening between nearest-neighbour sites of iron atoms.[33]

When comparing the relaxed hole for the hematite bulk (see page 46) to the relaxed interface hole, the same difference becomes apparent as with the comparison to the vertical experimental configuration: the relaxed bulk hole has Fe^{IV} character, which coincides with the experimental vertical hole, while the vertical bulk hole is delocalised over both O and Fe with nearly equal contributions. To resolve these incoherent results, it would be crucial to have experimental time-resolved measurements where the relaxation of the nuclei could be traced and compared to the simulation.

Figure 3.41 shows how the average charge localisation changes upon removal or addition of a single electron to the neutral c-cut setup analysed in the beginning of this chapter. The overall procedure to obtain the data was as follows. First, a regular AIMD of the neutral system was calculated. Along the trajectory, the instantaneous charge on each atom was calculated using the Hirshfeld partition scheme. This value then was averaged over the neutral calculations to obtain the reference value. Now for each of the four independent trajectories for the charged systems, the same method has been applied, yielding an average Hirshfeld charge per simulation and atom. Now the difference between this charge and the reference value from the corresponding neutral calculation can be evaluated for each individual atom. Finally, these per-atom charge differences have been aggregated along separate subsystems: individual atom kinds, bulk hematite layers (both oxygen and iron layers), as well as the entire bulk water section. The graphs in Figure 3.41 show these final charge differences. In this analysis, it is crucial to obtain all the reference data with the same functional, since the difference in electron density between functionals is significantly larger than the difference resulting due to charging the total system.

First of all, the overall picture between the two independent calculations for each case is very much consistent. Not only the overall charge distribution, but also the projections on subsystems yields the same picture for the independent trajectories. It is interesting

▲ **Figure 3.41** – Averaged charge on subsystems for two independent excess hole (top two rows) calculations and two independent excess electron (bottom two rows) calculations. Each row compares the averaged Hirshfeld charge per subsystem (layer or atom kind) to the same averaged Hirshfeld charge on the same subsystem from the neutral calculations. In each row, the left panel compares the atom kinds, while the right panel compares the individual layers of bulk hematite and all aggregated bulk water.

Calculation(s) ea7, b82, 2a2, 0e8, 9fb, 227, 5d8, 19b, 3f1, 446, 7a3

#	Vertical charge @Fe	Vertical charge difference @Fe	Average absolute spin moment @Fe
0	1.01	1.31	3.93
1	0.34	0.12	4.01
2	0.24	1.16	3.95
3	0.65	0.60	3.91
4	0.55	0.27	3.97
5	0.46	1.12	3.91
6	0.22	1.25	3.95
7	0.54	0.52	3.98
8	0.35	0.11	4.00
9	0.29	0.11	4.00

▲ **Table 3.6** – Electronic properties of the vertical hole along a neutral MD calculation of the hematite/water interface: the charge on the iron subsystem in e, the maximum charge difference at a single iron site in e and the average absolute spin moment on the iron subsystem.

Calculation(s) e19

to see that in all cases about two thirds of the excess charge is located on the hematite slab, no matter what the nature of the excess charge. In any case, within hematite the surface oxygen atoms play a crucial role and contribute the majority of the total charge of the hematite layer, which is sensible since the HOMO of the neutral interface has mostly oxygen $2p$ character. It is likely that this is a consequence of the surface geometry where the interlayer spacing of the surface is greatly different from the bulk structure. This not only applies to the layer spacing, but also to the placement of individual atoms within the layer plane, which in turn influences the coordination geometry of the subsurface iron atoms. These subsurface iron layers however, overcompensate the charge difference in the excess hole case. This indicates that electron charge is transferred from the neighbouring oxygen atom layers to the iron layer.

In order to evaluate whether the charge localisation found for the excess hole simulation mentioned above is due to the relaxation of the system as response to the hole and how this compares to the vertical hole in the interface, equidistant snapshots from the neutral AIMD calculation have been selected for calculation of the doublet state. The results are shown in Table 3.6. It is clearly visible that the total charge on the iron subsystem for the vertical hole is significantly higher than the average charge from the relaxed MD run as shown in Figure 3.41. This means that during the MD with an excess hole, configurations become stabilised that reduce the total hole charge on the iron subsystem as compared to

the configurations sampled in a neutral MD. This is surprising, since the calculations for bulk hematite in this work showed that the relaxed hole is localised on an iron site. Two potential solutions exist: either the geometry optimisation for the bulk setups consistently converges to a local minimum or the effect observed for the interface is a genuine result of the presence of the interface. One way to investigate this would be to analyse the configurations sampled in the calculations in greater detail. In particular, the coordination of the subsurface iron sites in the charged MD setups as compared to the bulk hole geometry would be of interest and warrant future work.

3.5 Conclusion

In this chapter, various aspects of the atomistic picture of the hematite/liquid water interface have been investigated. Having covered a selection of crystal surfaces and different surface protonations, the results presented cover the majority of experimentally observed and investigated crystal terminations. In particular the information regarding the surface protonation is valuable, since the proton dynamics at the surface is not directly observable by experiment.

In any case, the results are limited by the computational costs, which grow prohibitively large for properties that require longer sampling or larger systems: diffusion, ion-solvent interaction, finite size effects, long-range solvation effects, relaxation processes of surface protonation, or pKa values for a variety of surface protonations to quantify relative stability. To enable further calculations covering larger systems and / or longer time scales, a new force screening method has been developed which drastically reduces the computational costs for this and similar systems. The next chapter describes the underlying analysis, the resulting method and demonstrates its applicability to a variety of systems.

Chapter 4

Accelerating the Calculation of HFX Forces in Hybrid DFT-based MD

Density functional theory (DFT) has become a standard tool for investigating physical properties for a wide range of materials and systems when information about the electronic structure is required. Depending on the material in question, different levels of theory are to be applied in order to properly describe the electronic structure. While the Generalized Gradient Approximation (GGA) approach gives good results for lattice constants and absorption energies for many materials,[271,272] obtaining band gaps or electronic coupling matrix elements has been proven to be more difficult.[273,274] Besides general corrections like the Perdew-Zunger scheme[275] or constrained DFT (CDFT),[276] some materials benefit from adding a Hubbard-U term[105,107] to mitigate the self-interaction error (SIE).[277] While the (effective) U value which is typically found empirically by reproducing the band gap or other electronic properties[278,279] can also be found from first-principles calculations,[114,280] it still remains only one additional parameter that can be used for tuning the electronic properties and may not be able to reproduce several of them with one value.[104,281,282] For a case in point, see Figure 2.5. This can be further improved by employing hybrid functionals that mix Hartree-Fock-Exchange (HFX) into the functional.[206,283–285] While this can give results superior to DFT+U,[286,287] it also increases the computational costs significantly due to the (usually quartic[288,289]) scaling of the Hartree-Fock (HF) method with respect to the number of basis functions.

Several schemes are already in use to mitigate the high costs resulting from the HF

method, typically employed in a staggered combined approach. Firstly, integral screening based on the Schwarz inequality[290] typically improves the scaling from exponent 4 to 2.2-2.3.[289] For non-conducting materials, the density matrix falls off exponentially[291] with distance, which means that any two or four-center integrals are dominated by local interactions. Exploiting this improves performance not only on short-range.[292] While not the motivation[293] for range-separated hybrid functionals, this effect is used in these functionals which can be also applied to conducting materials[127] and–together with Schwarz screening–offer linear scaling.[294] While the range separation introduces empirical parameters again, their precise value is rarely critical to the results[295] and can obtained from first principles in some formulations.[296]

Besides the functional itself, another factor contributes to the balance of result quality and computational effort, namely the basis set.[297] In many cases, converging HFX results with respect to basis set size requires prohibitively large basis sets.[298] In these cases approaches like the Auxiliary Density Matrix Method[134] allow for the required quality of the basis set while limiting the computational effort.

All these approaches ultimately can be combined since they offer different angles of attack for the same problem. While this renders hybrid functional molecular dynamics (MD) feasible for extended systems, the overall costs still can be significantly larger than for standard DFT calculations. In particular for strongly correlated materials, where employing HFX drastically improves the results,[299] the distance-based range-separation approach becomes less effective and the forces required for MD and calculated via the molecular orbital derivatives[134] become the dominating cost for the overall calculation. The interface calculations presented in the previous chapter took about a month of walltime on 32 nodes with 24 cores each. This makes it highly desirable to reduce the computational cost to extend the time scale that can be covered. In order to improve the computational efficiency for the force evaluation without sacrificing accuracy, we suggest an interaction screening method exploiting the proximity of consecutive MD steps in phase space as outlined in the next section.

The main reason for the computational cost of HFX calculations are the four-centre integrals that appear in the energy expression, which is shown here for the atom centred

orbital representation

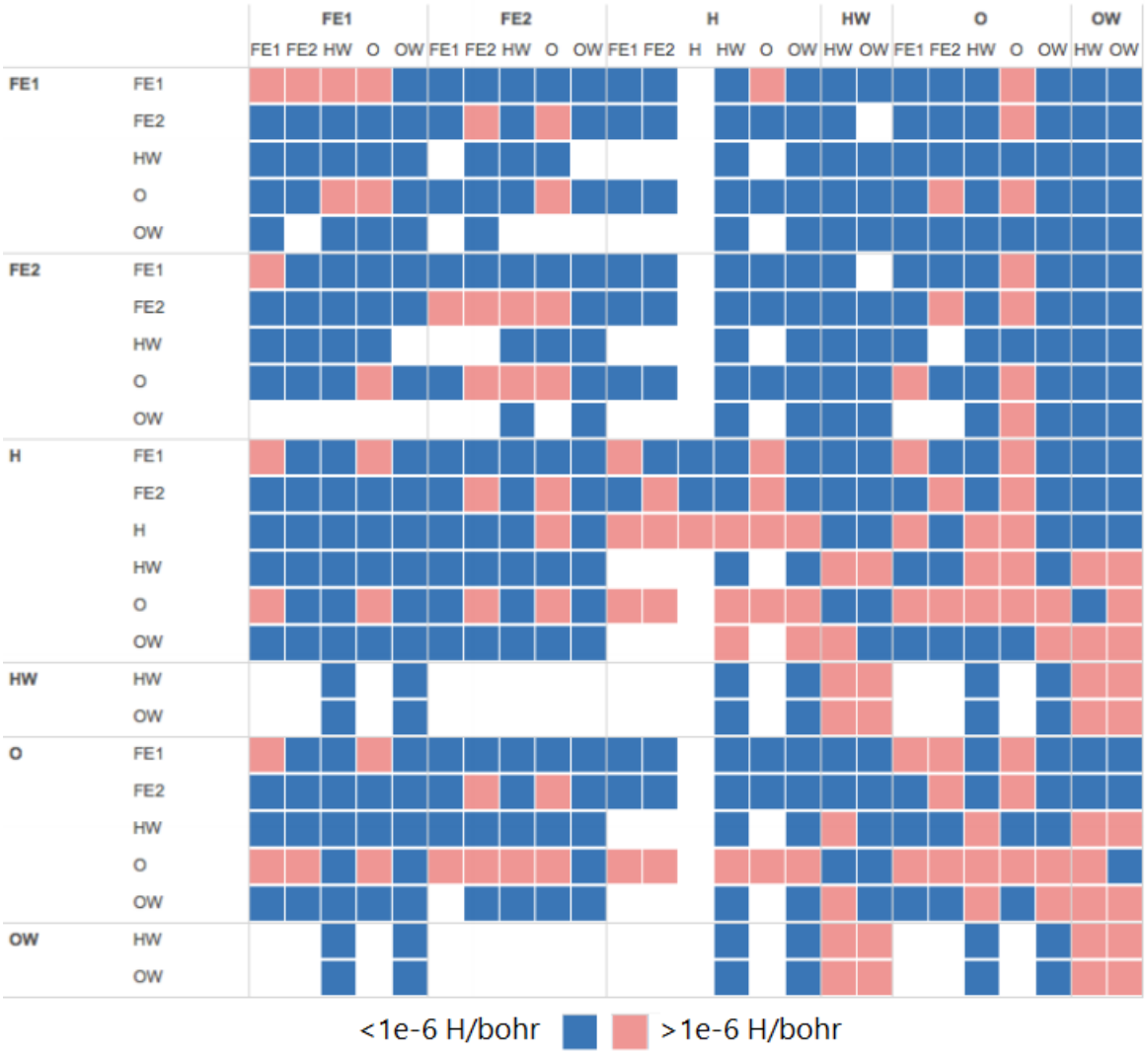
$$E_{\mathrm{X}}^{\mathrm{HF}} = -\frac{1}{2}\sum_{\lambda\sigma\mu\nu}\left[\sum_{i} C_{\mu i}C_{\sigma i}\right]\left[\sum_{j} C_{\nu j}C_{\lambda j}\right]\int d\mathbf{r_1}\int d\mathbf{r_2}\frac{\phi_\mu(\mathbf{r_1})\phi_\nu(\mathbf{r_1})\cdot\phi_\lambda(\mathbf{r_2})\phi_\sigma(\mathbf{r_2})}{|\mathbf{r_2}-\mathbf{r_1}|} \tag{4.1}$$

where C_{ij} are the density matrix coefficients, and ϕ_i represent the atomic orbitals. From this expression, it is clear that the higher the number of atomic orbitals with significant mutual overlap, the more expensive this expression. For the HSE functional, only the short range part contains HFX meaning that HSE calculations will be most expensive for systems that have a high local electron density. While Schwartz screening acts on the overlap of the orbitals alone and ADMM reduces the basis set size, the method proposed here is based on screening the residual total HFX force contributions, thus further exploiting the sparsity of the density matrix. Transition metal oxides with their strongly correlated electron systems are a case which strongly benefits from the presented method.

4.1 Analysis

Before any changes to the current calculation can be made, one has to understand the computational complexity of the current method. This involves profiling the components of the calculation to identify the most expensive parts thereof. As a sample system, the (001) surface of hematite in contact with water as outlined in the other chapters has been chosen. This system consists of 430 atoms with about 2,000 explicit electrons or about 10,000 basis functions. This system has been calculated using CP2K on 32 nodes of ARCHER where each node has 24 cores. For a typical MD calculation, the machine needs about 60 seconds for a single point calculation (SPC) while the force evaluation required for propagation of the atoms takes about 225 seconds. In both cases, the calculation was parallelised via MPI only. In each MD step, $1.2 \cdot 10^{12}$ electron repulsion integrals (ERI) were calculated.

Given the high number of electron repulsion integrals to be calculated, it would be surprising if all of them would contribute similarly. To investigate whether some of the electron repulsion integrals are more important than others and whether there is a specific pattern in the relative contributions, all relevant information has been recorded for a single snapshot of the hematite/liquid water interface. For each of the four centre integrals, the involved atoms, the quantum numbers (n, l, m, and s) of the four involved orbitals and the final force components have been stored only if the force components had an absolute

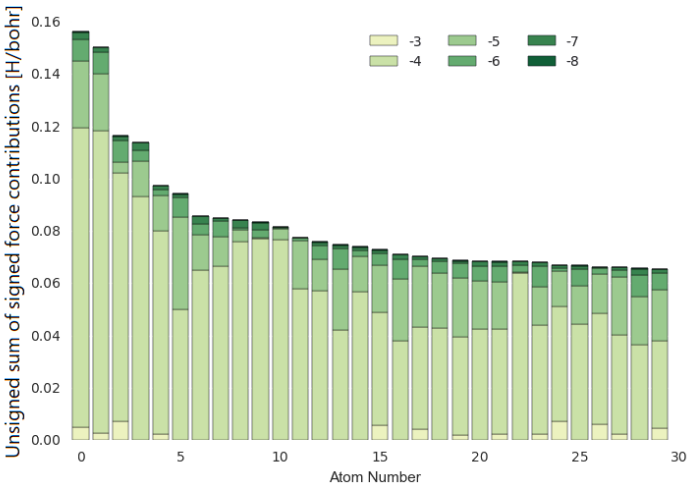


▲ **Figure 4.1** – Maximum absolute force contribution resulting per atom kinds of the four orbitals involved. In the setup in question, there are six atom kinds: Fe1 and Fe2 (iron atoms that differ by their majority spin in the antiferromagnetic spin pattern of hematite), H and O for hydrogen atoms and oxygen atoms within or at the surface of hematite, as well as HW and OW for hydrogen atoms and oxygen atoms in bulk water. Coloured cells denote combinations where the maximum force contribution is above (red) or below (blue) the threshold of 10^{-6}H/bohr. Empty or missing cells mark combinations that did not occur in this particular MD step, likely due to exclusion during the overlap screening or due to symmetry considerations. All calculations in the blue cells include 23% of the total ERI recorded.

Calculation(s) e81

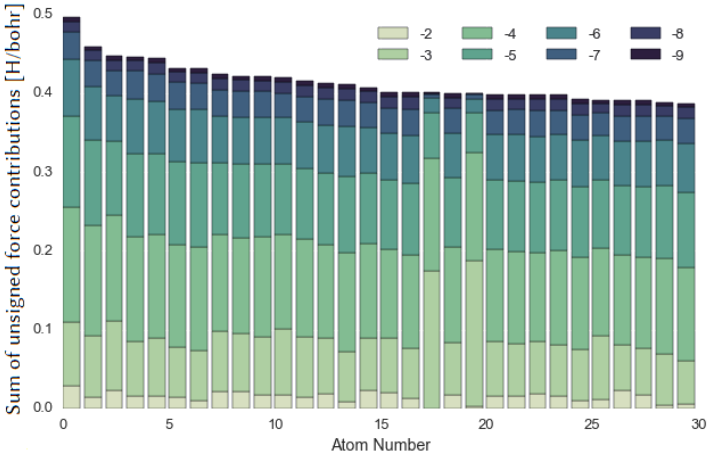
value of more than 10^{-10} H/bohr (data in simulation e81). For this snapshot randomly drawn from an equilibrated MD trajectory, a total of $13.7 \cdot 10^9$ forces resulting from $6.9 \cdot 10^9$ individual ERI have been recorded and subsequently analysed.

On the highest level, the orbitals can be grouped by the atom kind they belong to. This has been done by taking every recorded force component, identifying the four orbitals ϕ_i involved and looking up the atom kind of the atom where these orbitals are located. As



▶ **Figure 4.2** – Unsigned sum of signed force contributions for several atoms from a single snapshot of the hematite/water interface. Forces are grouped by the magnitude of their absolute value. Any bars below 10^{-8} H/bohr are too small to be visible. Underlying data included up to 10^{-10} H/bohr. This diagram shows contributions from HFX forces only, no other force components included. Image reused from my publication[300] with permission from ACS.

Calculation(s) e81

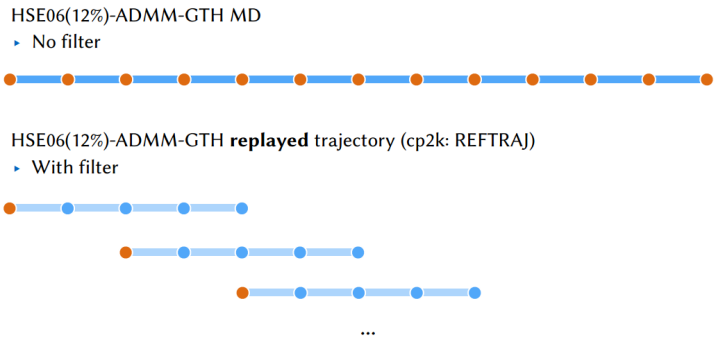


▶ **Figure 4.3** – Sum of unsigned force contributions for several atoms from a single snapshot of the hematite/water interface. Forces are grouped by the magnitude of their absolute value. Any bars below 10^{-8} H/bohr are too small to be visible. Underlying data included up to 10^{-10} H/bohr. This diagram shows contributions from HFX forces only, no other force components included. Image reused from my publication[300] with permission from ACS.

Calculation(s) e81

shown in Figure 4.1, already on that level many of the combinations exhibit only small contributions to the resulting force vectors. This means that for many combinations of atom groups, not a single one of the individual forces obtained from single four centre integrals has an absolute value larger than 10^{-6} H/bohr. While the graph in Figure 4.1 suggests that this already allows to greatly reduce the number of ERI to evaluate, one has to note that the cells with a maximum force contribution below the threshold of 10^{-6}H/bohr hold only 23% of the total recorded number of ERI. This means that selecting orbital combinations to be excluded purely based on their atom kind is too coarse and another method of distinguishing important from unimportant orbital combinations is required.

To justify the exclusion of groups of four centre integrals, it has to be shown that the many small contributions do not sum up to something significant. If they would, the individual small contributions would have to be calculated or approximated. Figure 4.2 shows



◀ **Figure 4.4** – Schema of evaluation runs to compare force approximations to actual force vectors. First, a reference trajectory (top) is created. Afterwards, equally spaced runs are branched off this trajectory following exactly the coordinates from the reference trajectory. For these replay trajectories, all but the very first step are approximated using a filter list built in the first step.

the contribution to the force vectors of single atoms in the system grouped by the individual contributions' order of magnitude. It is clearly visible that the contributions vanish for small absolute forces. Figure 4.3 illustrates that the contributions cancel each other out for small orders of magnitude. Since the data in Figure 4.2 is aggregated on a per-atom basis, an exclusion list on this level of detail is justified. In the following, all calculations have been performed with a filter mask either including or excluding all four centre integrals originating from all orbitals from a set of four not necessarily distinct atoms. Wherever force vectors are mentioned, this means only to the HFX components of the total force vectors. All other contributions are not considered since they are invariant to the algorithm changes proposed in this chapter.

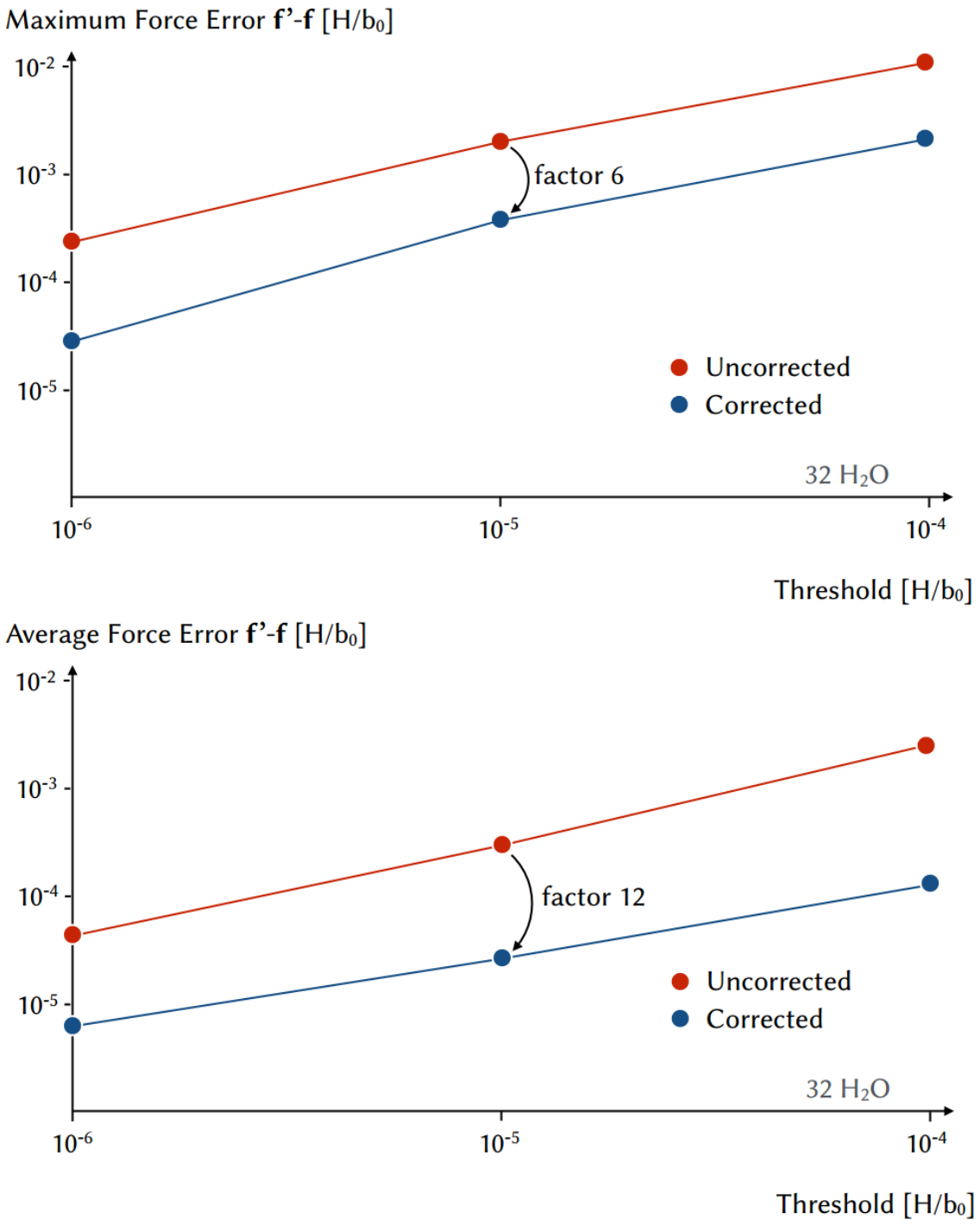
Given the applicability of this approximation, the overall concept can be outlined as follows: perform a full HFX calculation and identify small force contributions. Then the small contributions – i.e. contributions below a certain threshold – are neglected in the MD calculation for a certain duration until the exclusion list is updated by a full HFX step. This can be motivated by the fact that MD calculations move only slowly through phase space, meaning that subsequent steps have close geometries giving rise to similar importance of individual orbital combinations. The method itself has the advantage that it is an additional screening step that operates on top of Schwartz screening or ADMM.[134]

To be able to compare the approximated forces with the exact ones including all interactions, the following setup was used. First, an equilibrated MD run without any changes to the force evaluation has been performed. During this run, all configurations have been stored and used as input for subsequent runs that started the calculation at time steps shifted along the trajectory, as illustrated in Figure 4.4. In steps of the reference trajectory, the unfiltered (original) HFX force vectors have been recorded alongside the total number

of four centre integrals that have been evaluated. In all the shifted calculations, the filtered HFX force vectors have been stored together with the number of four centre integrals in each step. This process has been repeated for several different thresholds typically for each order of magnitude between 10^{-2} H/bohr and 10^{-6} H/bohr.

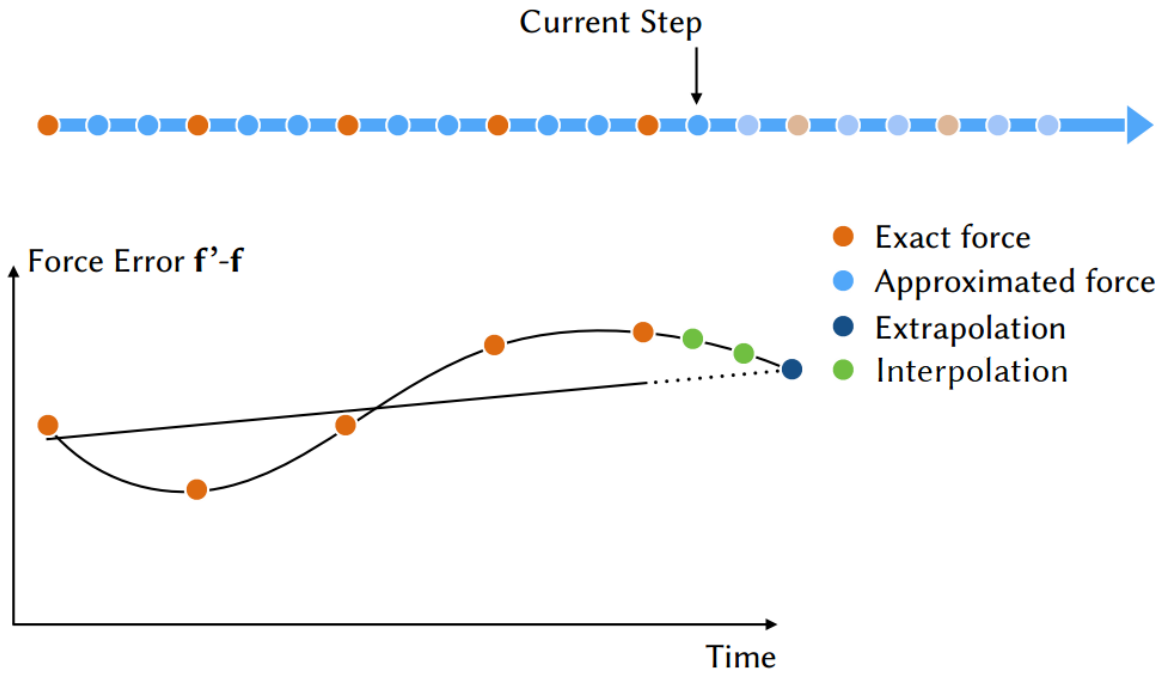
With this data at hand, it became possible to investigate the degree to which force evaluations at subsequent geometries from a single contiguous trajectory differ. Figure 4.5 shows the uncorrected results for a simple system, a periodic water box. As it turns out, increasing the threshold of individual force components that are neglected lead to both a higher average and maximum deviation from the reference forces. While this general trend is expected, the quantitative picture is more important here. It shows that constraining the maximum force error to a common convergence criterion for geometry optimisations of $5 \cdot 10^{-4}$ H/bohr requires a fairly small threshold of 10^{-6} H/bohr for the individual components. This low threshold, consequently, is unable to provide a significant speed-up, since almost all of the four centre integrals have to be evaluated in this case. Two potential strategies could improve the situation: either switch the exclusion list from working on atoms to working on a smaller entity (until ultimately reaching a single orbital) or add a correction mechanism that reduces the overall error that is to be expected. The first case would drastically increase the memory requirements of this method, since the exclusion list with all combinations of the entities to ignore has to be stored. With increasing level of detail, more combinations would have to be recorded, which also increases the lookup times when querying the exclusion list. Moreover, the lookups would have to happen more often, since the individual entity is smaller and each entity has to be checked separately. Therefore, the second approach, the correction scheme, was pursued.

For the correction schemes, the basic approach was as follows. Input to the correction scheme is part of the history of the trajectory until a full update step (where all HFX contributions are evaluated and a new exclusion list is prepared). Output is a correction to the HFX forces of a certain atom for the subsequent time steps where only part of the HFX force contributions are calculated. Based on the reference data outlined above, various correction schemes have been evaluated. Autoregressive methods like ARIMA were able to predict the overall time-evolution of the forces, likely due to the autocorrelation thereof. However, the prediction was never close enough for the systems under investigation to warrant go-



▲ **Figure 4.5** – Maximum (left) and average (right) force error depending on the threshold. Calculated comparing the true HFX forces **f** to the approximated HFX forces **f**′. Underlying data from a MD simulation of a periodic box of 32 water molecules over 1,000 steps. Image reused from my publication[300] with permission from ACS.

Calculation(s) 70a



▲ **Figure 4.6** – Correction scheme applied to reduce the maximum force error due to the approximation. Details described in the main text. Image reused from my publication[300] with permission from ACS.

ing forward with these class of statistical models. In particular, since all the models from this family have high requirements to the underlying data series which render them hardly applicable to the problem at hand. More precisely, the required stationarity can rarely be guaranteed, since e.g. rotation of the physical system produces non-stationary projections along the axes of the coordinate system. Similar restrictions apply to the heteroscedasticity: changing the chemical environment or bond configuration during the MD run gives rise to fluctuations of different extent.

The scheme that finally was successful makes use of physical information and is described in the following section.

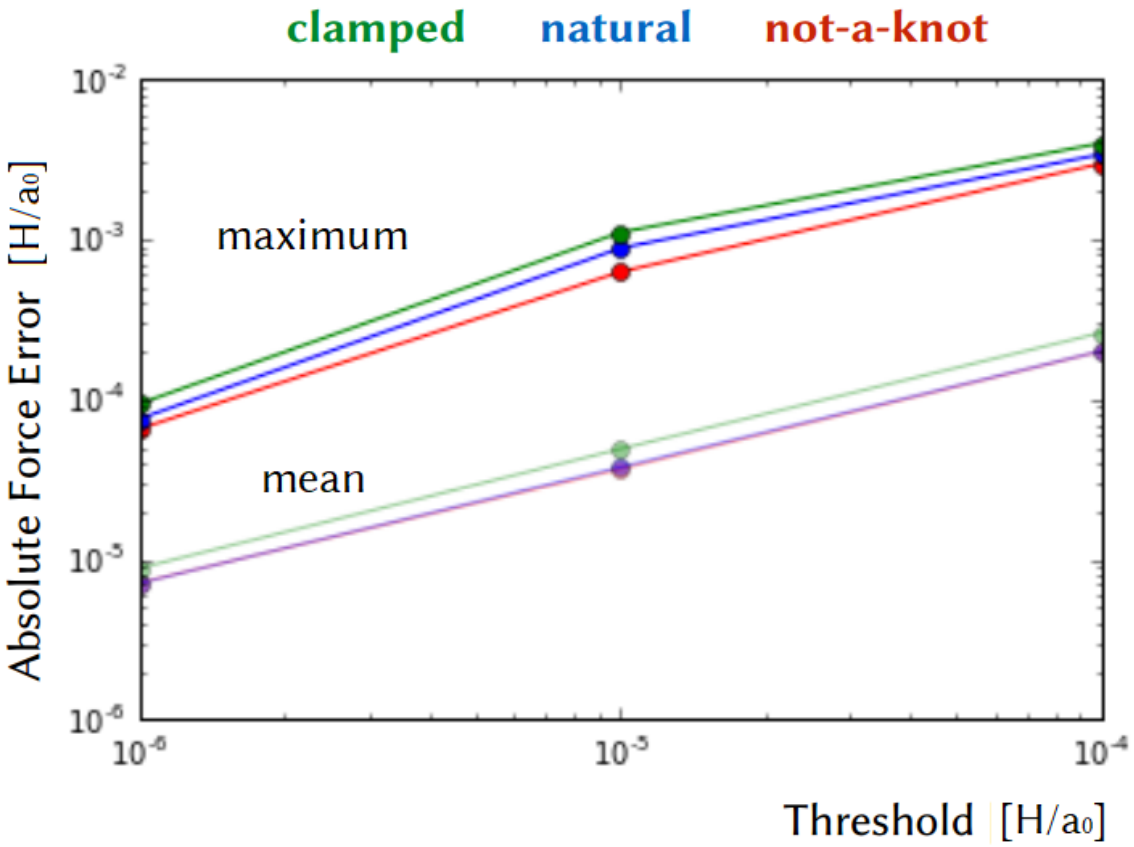
4.2 Final Method

Figure 4.6 illustrates the process for a single axis of the coordinate system and a single atom. In a stable equilibrated MD calculation, forces are smooth in times and, therefore, have no sudden jumps. The same applies if the underlying interactions are slightly perturbed by neglecting part of the ERI. The difference of two smooth functions is a smooth function again, so the error between approximation and reference should be a smooth function in

time. At regularly spaced intervals, the exact error made by the approximation is known: each step that updates the exclusion list has perfect information and can quantify the approximated forces, although the exact forces are used for propagation. This error in forces for each component of the coordinate system and every atom can be recorded and extrapolated. Extrapolating is done based on the last five steps in two steps: first, the linear trend found by linear regression is extrapolated for the next full update step. Secondly, the value at the approximated steps is estimated by cubic spline interpolation between the past five and the one future data point.

For the spline interpolation involved, three flavours have been investigated, which differ only by the boundary conditions. The boundary conditions can be important, since the only area of interest, the estimate of future errors, is in the last segment of the spline interpolation, hence, the error estimates are influenced by the right boundary condition. Generally, the condition of a continuous second derivative generates $n - 2$ equations for the n data points in the interpolation range. This requires two additional conditions to generate a system of linear equations that can be solved to obtain the coefficients of the cubic polynomials over the intervals between the n data points. The first way of adding these conditions in question is called natural spline, since it is inspired by a flexible ruler being bent such that it follows the individual data points. Therefore, the condition is that the second derivative has to be 0. This way, the natural spline becomes linear at the boundaries. Another variant, the clamped spline, requires a specific non-zero slope to be defined. Finally, the not-a-knot spline has been evaluated where the third derivatives are required to be continuous at the second and second-to-last points. Other boundary conditions such as the periodic one are not applicable and, hence, have not been evaluated. Figure 4.7 shows a direct comparison of the aforementioned cases. The suggested method uses the non-a-knot variant, since it consistently produces lower errors.

Revisiting the results in Figure 4.5, it becomes immediately apparent that the correction method outlined above is crucial to realise the potential efficiency gain. The correction mechanism brings the average error down by more than one order of magnitude. This allows for a threshold about one order of magnitude higher for individual force contributions which in turn allows for more aggressive approximations. With fewer ERI to evaluate, the overall performance is greatly improved without sacrificing the accuracy of the

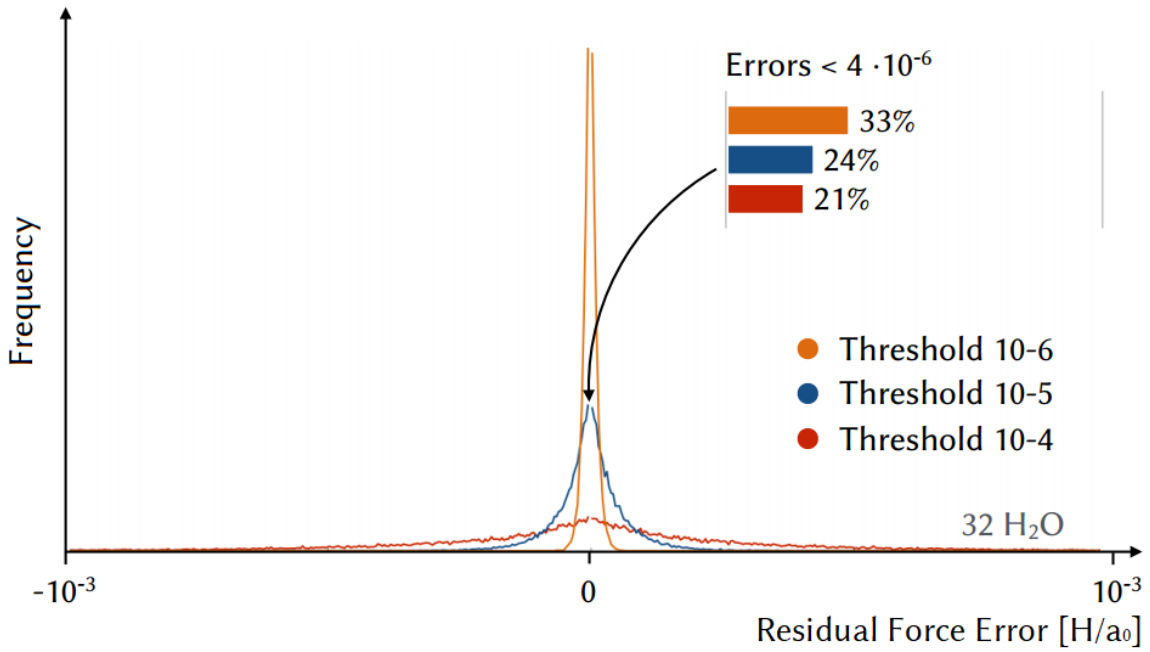


▲ **Figure 4.7** – Absolute force error depending on the choice of the boundary conditions for the spline used in the error correction scheme. Maximum force error and minimum force error curves are shown for a MD simulation of a periodic box of 32 water molecules over 1,000 steps. Data obtained using replays and not via actual MD calculations following the schema in Figure 4.4. Definitions of the boundary conditions employed are given in the main text. Note that the mean error for natural cubic splines and not-a-knot cubic splines coincide.

Calculation(s) 70a

computation.

Figure 4.5 however holds only aggregate data and does not show the distribution of individual force components. This is required, since there could be a systematic bias in the resulting forces. If this would be the case, the rotational invariance would be compromised, since the bias would be different subject to rotations of the system. Moreover, a drift motion of the centre of mass would be a potential result, removing energy from the kinetic degrees of freedom in the system, effectively cooling the MD simulation. Figure 4.8 shows the distribution of the residual force errors after the correction scheme has been applied. It is striking that for a threshold of 10^{-6} H/bohr one third of the forces differ by less than $4 \cdot 10^{-6}$ H/bohr from the corresponding accurate reference force. This is surprising because the omission of individual contributions of the order of 10^{-7} H/bohr does not produce an error



▲ **Figure 4.8** – Signed residual force error after the correction scheme has been applied for different threshold values. The x-axis is linear but the middle bin of the histogram has been omitted for clarity. The values of which are shown as the inset bar chart. Underlying data is from a MD simulation of a periodic box of 32 water molecules over 1,000 steps. Image reused from my publication[300] with permission from ACS.

Calculation(s) 70a

larger than one order of magnitude higher. Moreover, the force error is largely symmetric, as can been seen from the histograms in Figure 4.8. This holds for both aggressive and conservative thresholds.

The method as outlined above has three free parameters that need to be fixed in order to obtain a stable and generic method that does not require manual work upon a change of the underlying physical system of interest. The first parameter n is the number of steps between full update steps where the exclusion list is updated, while the second parameter f_T is the threshold below which individual force contributions are considered to be small and, hence, negligible. Finally, the number of history points k considered in the force error correction scheme has to be pinned.

The repeat frequency n affects the performance strongly for small n. In particular, n should be as large as the correction scheme allows it to be. The overall speedup S is given as

$$S = \frac{nm}{m+n-1} \tag{4.2}$$

where m is the ratio of the CPU times for a step with a full HFX force update and a step with a screened HFX force calculation (i.e. $m \geq 1$). From this equation, it is clear that the increase of speed-up going from $n = 2$ to $n = 3$ is larger than going from $n = 3$ to $n = 4$, because the cost of the full step is distributed over n steps. The maximum theoretical speedup is $S = m$.

Regarding performance, the force threshold is more critical than n. A high threshold means higher performance gains, as a higher number of ERI is excluded from evaluation for intermediate steps. However, f_T has to be small enough to reproduce accurate forces. This value could be optimized for each system separately. However, a value of 10^{-5} H/bohr was found to reproduce the forces with a maximum error of $5 \cdot 10^{-4}$ H/bohr for a wide class of systems as outlined in the next section.

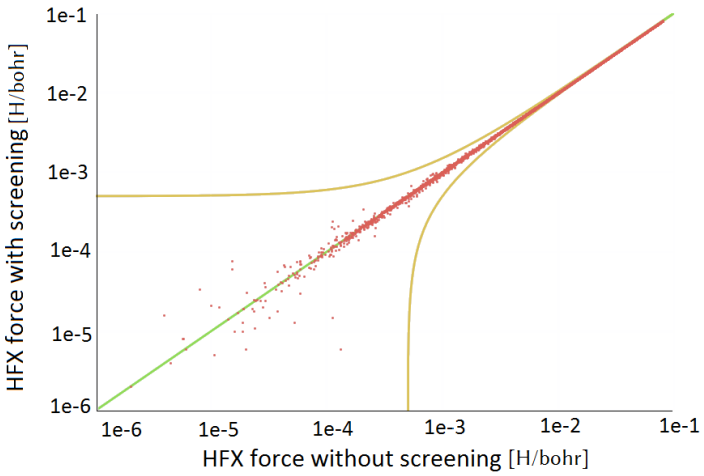
For the last parameter, the number of history points k, there is a physical justification for its value. The HFX forces are particularly sensitive to changes of the orbital overlap (this is what screening mechanisms typically exploit). This is why they fluctuate on the time scale of vibrations, the fastest of which typically is the OH stretch motion. If the history covers much more than one oscillation period, the linear regression will give nearly zero slope, i.e. it will not follow the instantaneous trend and underestimate the variance of the forces. If the history is too short, i.e much shorter than one single oscillation, the error estimate near the turning points of the time-dependent HFX force will severely over- or undershoot the actual errors. For a time step of 0.5 fs, a value of $k = 5$ has been found to be a good compromise between these effects.

4.3 Results

To validate the method for real systems of interest, several (potential) issues have to be addressed: force accuracy, energy conservation, collective effects, and performance. Each of which has been addressed by a separate subsection.

4.3.1 Force accuracy

Since the aim of this method is to recover the original force vectors while reducing the computational cost, it is only natural to compare the individual force components. In order to do that, a 32 molecule water box has been prepared. From the same initial configuration

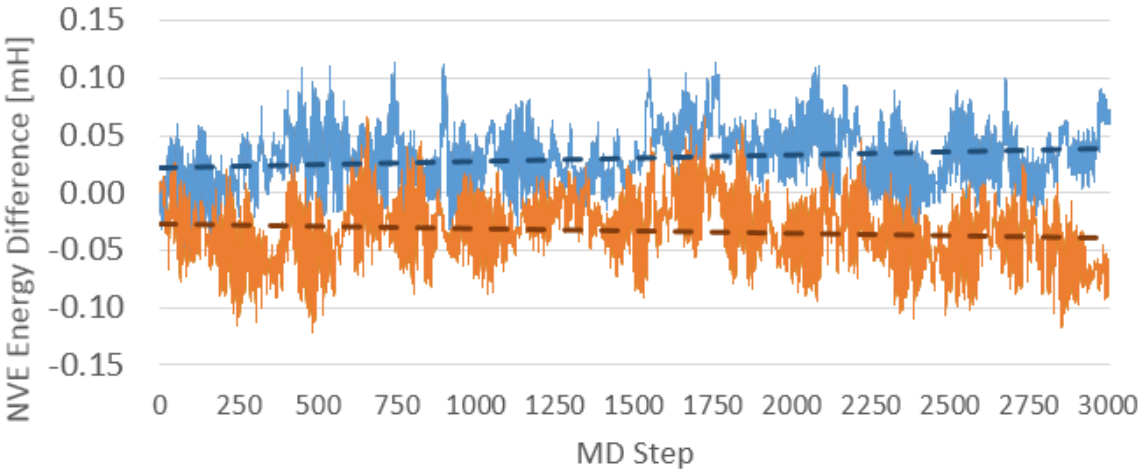


◀ **Figure 4.9** – Force calculated with screening over force calculated without screening. All forces are unsigned. Every scatter point (red) is for one atom and one axis in one of 1,000 MD time steps for a fully periodic box of 32 water molecules. Diagonal (green) shows the ideal correlation. Curved lines (yellow) denote the desired accuracy which is at most the convergence criterion around the ideal force. Image reused from my publication[300] with permission from ACS.

Calculation(s) 70a

and initial velocities, a molecular dynamics run of 1,000 steps was started with the original implementation of HSE06. In this case, the initial converged electron density has been used as initial guess for the two subsequent runs (with and without the approximation outlined in this chapter). The same initial guess is important, since it may introduce small differences in the forces, since the density is converged subject to a finite convergence criterion in any case. The common reference trajectory is required, since otherwise minor differences in the force vectors would give rise to different configurations being sampled for the two methods, which would make the trajectories slowly diverge in phase space similar following the Lyapunov exponent.[301] For both the original method and the approximative scheme, the simulation has been performed with ADMM and GTH pseudopotentials as well as D3 dispersion correction.

Figure 4.9 shows a direct comparison of the force components obtained using the aforementioned technique. It is clearly visible that the resulting forces are very close to the original ones. For large absolute force values, the variance is very small. For small absolute force values, there is a visible spread away from the ideal values. However, this variation is well within the convergence criterion typically employed for geometry optimisations, $5 \cdot 10^{-4}$ H/bohr. This means that this is the expected error of the reference force values as compared to the full HFX forces in the limit of an infinite SCF convergence. It is to be noted that the data shown in Figure 4.9 comes from a MD run where the exclusion list of the four-centre integrals has been updated every five time steps and, hence, gives a realistic picture of the accuracy of the force approximation during a regular production run.



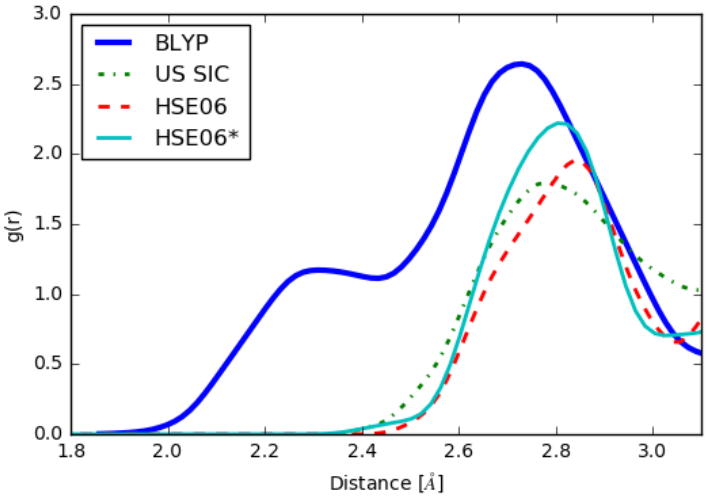
▲ **Figure 4.10** – Energy (conserved quantity) from a water dimer in NVE ensemble. Bottom curve without, top curve with HFX integral screening for a HFX forces threshold of $f_T = 10^{-5}$H/bohr and a time step of 0.5 fs. Energy drift as determined by linear regression is $4.7 \cdot 10^{-7}$H/atom/ps with screening compared to $3.2 \cdot 10^{-7}$H/atom/ps. Image reused from my publication[300] with permission from ACS.

Calculation(s) 1b4

4.3.2 Energy conservation

Altering the way how forces are calculated may give rise to an increased energy drift that would slowly destabilise the MD calculation. In a BOMD calculation, reusing the electron density from previous steps (as done in CP2K using the ASPC[302] method) in itself already introduces a time-dependency and therefore (unless in special cases[303]) makes the calculation non-reversible, which in turn gives rise to a small energy drift.[147,304,305] This energy drift typically is in the order of 10^{-6} H/atom/ps[306] for NVE ensembles. In NVE ensembles, there is no thermostat or barostat that could remove the energy drift of the system. Due to this property of the NVE ensemble, the same system (a water dimer) has been simulated for both the full method as well as for the approximated forces.

Figure 4.10 shows the energy drift for a water dimer in direct comparison between a MD calculation with and without the force approximation in NVE ensemble. The energy drift is clearly visible on the time scale covered and is about 45% higher for the approximated calculations. While a higher energy drift is expected in general, since all the aspects contributing to the energy drift of the unmodified force evaluation also apply to the modified version thereof, the difference resulting from the approximation of the forces is still small, keeping the energy drift at a typical level for DFT calculations.[306]



◀ **Figure 4.11** – Pair correlation function between the oxygen atom of the OH radical and all oxygen atoms of the solvating water box. The hemibonded peak at 2.25 Å as obtained from BLYP is not present in HSE06 after 2 ps equilibration. Accelerating the forces as presented in this work does not alter this dynamic picture (labelled HSE06*). Data for BLYP and US SIC from VandeVondele and Sprik.[307] Image reused from my publication[300] with permission from ACS.

Calculation(s) 045, fef

4.3.3 Collective Effects

With the demonstrated small errors on the force components and the energy conservation outlined above, one potential issue is a structural bias, e.g. discriminating between forces along axes and in-between axes. To assess whether this effect is present in the outlined method, a system of particular sensitivity to HFX forces has been chosen. The OH radical solvated in water exhibits a (spurious) hemibonded peak in the O_{OH}–O_{HOH} RDF when a GGA functional alone is used. This is the direct consequence of a solvation structure which is too close to water.[307]

A MD calculation of one OH radical solvated in 31 water molecules has been performed at HSE06-GTH-ADMM level with and without the approximation scheme. In both cases, the initial structure comes from a snapshot where the coordination of the solvated OH radical is still close to the water solvation structure.

Figure 4.11 compares the RDF for the solvated OH radical from fully periodic calculations. The spurious hemibonded peak has vanished in both the calculations with and without the force approximation, which means that the overall interaction between the sites is correctly covered in both cases. By and large, the shape of the RDF for the approximated method and the full calculation are close. The deviation left is attributed to finite sampling.

Material	UC Volume [Å^3]	Units per UC	Electrons per UC	Electron density [e·Å^{-3}]
$InVO_4$	332.7[308]	4	200	0.62
$InNbO_4$	143.5[308]	$InNbO_2$	38	0.26
$InTaO_4$	143.7[308]	$InTaO_2$	38	0.26
$BiVO_4$	309.6[308]	6	252	0.81
MnO_2	55.6[309]	2	54	0.97
CoO	81.5[310]	4	92	1.13

▲ **Table 4.1** – Number of explicit electrons for GTH pseudopotentials for various transition metal oxides. All data is given for a single unit cell (UC).

4.3.4 Performance and Scaling

When assessing the performance of the method in question, it is important to select systems that are genuinely of interest and not only systems where the method performs well. Since the focus of this thesis is a transition metal oxide (TMO), several compounds from that family have been considered as sample material. TMO are of interest not only to theory due to their strongly correlated electrons,[311] but also to applications e.g. water splitting.[39,312] In total, the method was applied to cobalt(II) oxide,[313] hematite (α-Fe_2O_3),[314] a water dimer, bulk water, on OH radical in bulk water[307,315] and a pre-equilibrated hematite/liquid water interface.[238] From the materials in Table 4.1, CoO has been chosen as benchmark compound, because of the comparably high local (explicit) electron density, which makes it expensive. Moreover, CoO has a small number of atoms per unit cell, which makes it more suitable to scaling tests where the system size is increased and the performance is measured as long as the compute resources allow for a system of that size.

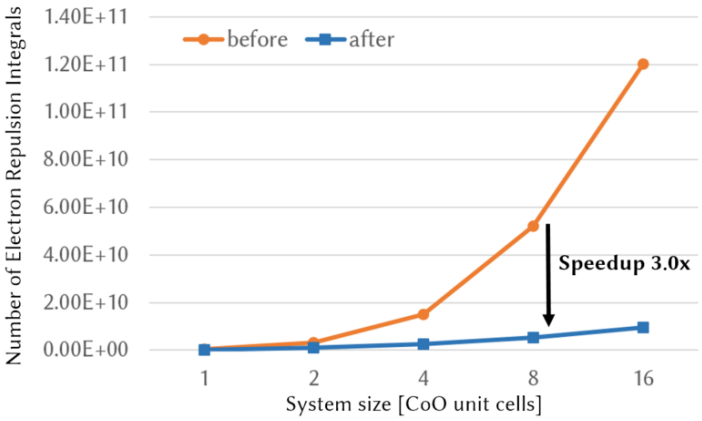
Table 4.2 shows the effective speed-up that could be realised. Speed-up values are obtained by performing MD calculations of different length with and without the force approximation and comparing the total runtime with the exception of the first 25 steps which are identical in both methods. Timings include all steps (including the full evaluation where the exclusion list is updated) and, therefore, are the true timings relevant for production runs. All the benchmarks have been performed on a Microsoft Azure G5 instance with 32 cores and 448 GB of memory. This shows that for real systems of interest, a speed-up of about three can be realised.

Moreover, for larger systems, significantly fewer four-centre integrals have to be evaluated. For the $4 \times 2 \times 2$ CoO setup, only 8% of the ERI have to be evaluated to recover the

System	Number of ERI	Included ERI	Speedup
CoO bulk $1 \times 1 \times 1$ (92 electrons)	$4.2 \cdot 10^8$	55%	1.4
CoO bulk $2 \times 1 \times 1$ (184 electrons)	$3.1 \cdot 10^9$	30%	2.0
CoO bulk $2 \times 2 \times 1$ (368 electrons)	$1.5 \cdot 10^{10}$	16%	2.7
CoO bulk $2 \times 2 \times 2$ (736 electrons)	$5.2 \cdot 10^{10}$	10%	3.0
CoO bulk $4 \times 2 \times 2$ (1.472 electrons)	$1.2 \cdot 10^{11}$	8%	2.9
Water dimer (16 electrons)	$8.9 \cdot 10^6$	89%	1.1
32 water box (256 electrons)	$1.5 \cdot 10^9$	32%	1.3
OH radical in 31 water box (255 electrons)	$2.1 \cdot 10^9$	50%	1.3
Hematite bulk $1 \times 1 \times 1$ (300 electrons)	$1.6 \cdot 10^{10}$	54%	2.4
Hematite bulk $2 \times 1 \times 1$ (600 electrons)	$4.7 \cdot 10^{10}$	28%	–
Hematite bulk $2 \times 2 \times 1$ (1.200 electrons)	$1.1 \cdot 10^{11}$	21%	–

▲ **Table 4.2** – Speed-up obtained for model systems with number of electron repulsion integrals (ERI) originally needed and the share of ERI included in the simplified calculation. Speed-up of 2 would mean twice as fast MD as without this method. All calculations done with $n, k = 5, f_T = 10 \cdot 10^{-6}$ H/bohr.

Calculation(s) 4a4, 045, 2ba, b34, 9be



◀ **Figure 4.12** – Scaling of the performance improvement with system size.

Calculation(s) 4a4

full forces. This highlights the screening nature of this approach. While Schwarz screening works on the integral overlap only[290] and improves the scaling from the naive N^4 to $N^{2.3}$,[289] this approach takes the density matrix coefficients in the expression for the forces into account. While this could be thought of as making the HFX interaction more short-ranged, this is not the correct picture: four-centre integrals that contribute significantly to the overall forces are still taken into account no matter what value the distance or orbital overlap. Since the total share of ERI to be kept for the intermediate approximate time steps decreases, the method generally improves the scaling of the HFX force evaluation, thereby pushing the limit of calculations that are feasible for a given budget. The scaling improvement based on the data from Table 4.2 is shown in Figure 4.12.

4.4 Implementation Details

In terms of implementation, my own work consisted of deriving the method, building a prototype in python and performing all the evaluation calculations outlined before. The base implementation of the exclusion list as a dense matrix in the Quickstep[316] module of CP2K following the python prototype has been done by Rasmus Jakobsen[300] under my supervision. Afterwards, I extended his implementation to work with a sparse matrix representation rather than the dense matrix that has been used for simplicity. Finally, I changed the load balancing scheme for HFX calculations in CP2K to deal with the filtered approach.

All implementations have been tested in simulations (a2d, 7ce) to reproduce the original forces of CP2K up to machine precision.

Initially, a dense matrix has held the exclusion list, which in turn consists of the informations about the four atoms the interactions of which are to be ignored and the affected spin channel. This scales to the fourth power of the number of atoms in the system and has to be available on each MPI rank separately, which typically limits the system size to less than 100 atoms. In the base implementation, this data structure was a matrix of double precision floating point numbers, since the value held the maximum force component observed on any rank had to be kept until the full update step has been completed. With 8 bytes per matrix cell, two spin channels, 100 atoms and 32 MPI ranks, this implementation required $8\,\mathrm{B} \cdot 2 \cdot 100^4 \cdot 32 \simeq 48\,\mathrm{GB}$ of memory just for bookkeeping.

This exclusion matrix however, is sparse similar to a contact matrix or adjacency matrix. This is a direct consequence of the short-range HFX interaction in the HSE functional, but also emerges naturally for other functionals, since the orbital overlap decreases exponentially with distance.[294] This sparsity can be exploited by finding a data structure that omits all the entries that are not used at all, thus not only reducing memory requirements, but also CPU and communication overhead.

In order to keep track of those electron repulsion integrals (ERI) that have been observed and are considered to be too small be warrant an additional evaluation in the upcoming steps, a table or sparse matrix implementation is employed as follows. The HFX module of CP2K splits the interactions into tasks in an automated fashion. Each MPI process then get one of these tasks assigned. Since there are much more tasks than MPI pro-

cesses, this method is also used for load balancing. In the implementation, each MPI process still works on the tasks assigned. However, each process keeps track of two lists: the list of observed but negligible ERI (described by the four atom kind indices and the spin channel) and the list of forces below the threshold (described by the four atom kind indices, the atom the forces act on and the spin channel). Both lists originally are implemented as sorted linked lists, i.e. each element except for the last one contains a pointer to the next element. This allows fast recording of the entries in a unordered fashion.

The first list, the one of the observed but negligible ERI, is merged across the nodes using a Map-Reduce approach similar to MPI_Gather that also works efficiently when the number of processes is not a power of two.[317] For this purpose, the linked lists are converted to a fixed-size Fortran ALLOCATABLE (map), then sent via the network and inserted into the linked list of the reducing process (reduce). This is repeated until only one final linked list is available on the master node. This is the exclusion list for the upcoming MD steps and must be present on all processes. Therefore, this list is finally converted into an allocatable and replicated on all nodes.

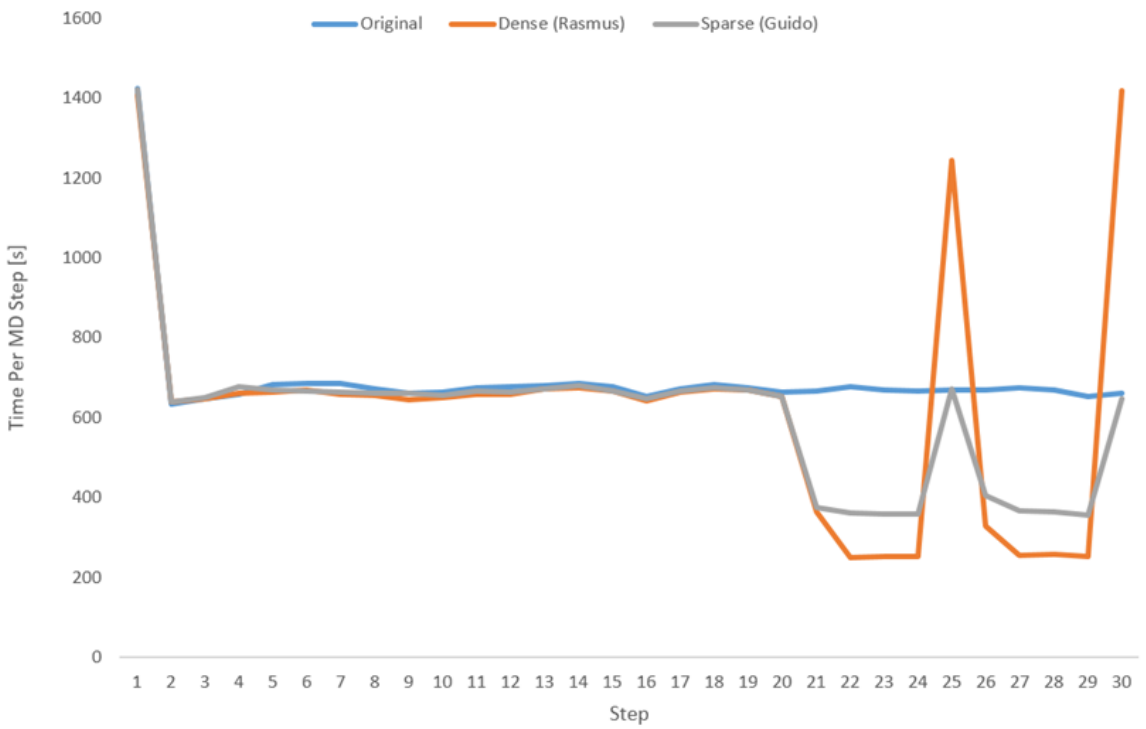
The second list is required to calculate the force error of the method during the update steps, which in turn is the basis for the force correction in subsequent MD steps. This list is kept as linked list and only traversed once in the end. This list is not replicated on separate processes, since the data does not have to be available globally. Based on the exclusion list received from the master process, each MPI process calculates its own contribution to the total force error vector. These vectors which scale linear with system size are finally added up across all processes, followed by a deallocation of this list on each process.

In order to reduce the memory consumption of both lists, two shortcuts are used when inserting an element that already exists. Since memory always holds sorted linked lists, duplicates can be detected in constant time during an insertion. For the list of observed but negligible ERI, the entry is only updated and for the force error history, the new entry is added to the existing force vector.

The modified way of calculating the four-centre integrals creates a problem with the load balancing. For the HFX part of the system, load balancing for both the energy and the force evaluation is done[318,319] by separating the nested loops over the four orbitals (i, j, k, l) into chunks (called either *work packages* or *bins* in the source code). These chunks

are defined by a range of values for i and j. Chunks always include the two loops over k and l. This means that the enumeration of the individual ERI is always contiguous for each chunk. Initially, the cost of each of the chunks is estimated by the number of four-centre integrals belonging to it and is subject to empirical weighting of the ERI to adjust for the different computational complexity for different orbital combinations. In subsequent steps, this estimate is replaced by the actual walltime of the calculation. The underlying idea obviously is that the total cost of all ERI varies only slowly over time. Now with the method outlined in this chapter, this assumption does not hold any more, since a good part of the ERI are neglected all together. CP2K uses the cost estimate to derive which chunks have to be calculated by which MPI rank such that the overall cost is as evenly distributed as possible. This means that the wrong (or more precisely: outdated) walltime measurements for the cost of the chunks lead to unequally distributed overall cost. This imbalance degrades performance, which is why changes were necessary. I changed the code to keep two estimates for each chunk: one for the full step (any initial MD step until the history of force errors is built) while the other one is applicable to any step where approximate forces are calculated. In the first approximate step, where no such information is available from previous calculations, the estimated duration of the full force evaluation is rescaled by the share of ERI to include in this particular chunk. To this end, the number of ERI is counted for each chunk.

Figure 4.13 compares the time per MD step required for three versions of the code for the same problem. The first MD step is more expensive than the others, because converging the initial guess of the density matrix takes significantly more SCF iterations than converging the density from the last MD step on the new atom coordinates. In the further progress, the initial steps have all nearly identical costs. Starting from step 20, a pattern repeats every five steps. The first step in the group is the step where the exclusion list is updated. The higher cost in the dense matrix implementation comes from the lack of improved load balancing. This can be seen from the fact that step 20 and step 25 perform roughly the same work, with the difference being the load balancing between MPI ranks which is not optimal for step 25. In the sparse matrix method, the cost of the update steps is nearly identical to the cost in the unmodified code, thanks to the load balancing. All intermediate steps are significantly cheaper than the reference run. The dense matrix implementation is



▲ **Figure 4.13** – Time per MD step required for fully periodic system of a $1 \times 1 \times 1$ CoO super cell for the original routines in CP2K (blue), the dense matrix implementation (orange) and the sparse matrix version (grey). Calculations performed on an Microsoft Azure DS15 v2 instance (20 cores, 140 GB memory). Time is measured in walltime, not CPU time. With the exception of the keyword enabling the force filtering from this work, the inputs are identical to all the versions of CP2K.

Calculation(s) d0f

slightly faster here, because lookups in a static dense matrix are constant time operations, while lookups in the linked list representation of a sparse matrix are linear in the number of non-sparse elements.

4.5 Conclusion

In this chapter, a new force screening approach has been proposed. It has been shown that this method has little impact on the resulting forces but can significantly improve the computational efficiency. The method conserved the energy of a MD run to a similar extent as regular MD and potentially can be extended to act on the energy calculation, as well. This would reduce the computational cost even further and likely improve the energy conservation as the energy and forces then would be evaluated based on exactly the same orbital interactions.

Chapter 5

pKa Calculation

Determining the protonation state of a surface is challenging both experimentally and computationally. In experiment, the low scattering cross section of protons makes their detection hard, while the associated time scales are hard to cover in simulations of sufficient level of theory.* While the work in previous chapters has shown that in some cases the picosecond time scale is long enough to capture some of the deprotonation events at the hematite/water interface, this is not generally the case. One way to capture these effects that can span multiple orders of magnitude in their time scale is the calculation of the acid dissociation constant pK_a. For the interface in question, this is another way of writing the free energy difference between a surface oxygen being protonated and the surface oxygen being deprotonated at the expense of a solvated H^+ in bulk water.

The acid dissociation constant in the Brønsted definition is given as

$$pK_a \equiv -\log\left(\frac{[A^-][H_3O^+]}{[HA][H_2O]}\right) \tag{5.1}$$

with the equilibrium concentrations $[\cdot]$ of the compounds in the following equilibrium reaction

$$HA + H_2O \rightleftharpoons A^- + H_3O^+ \tag{5.2}$$

Moreover, the Gibbs free energy difference for this reaction is given as

$$\Delta G = -k_B T \ln\left(\frac{[A^-][H_3O^+]}{[HA][H_2O]}\right) = \ln(10) \cdot k_B T \cdot pK_a \tag{5.3}$$

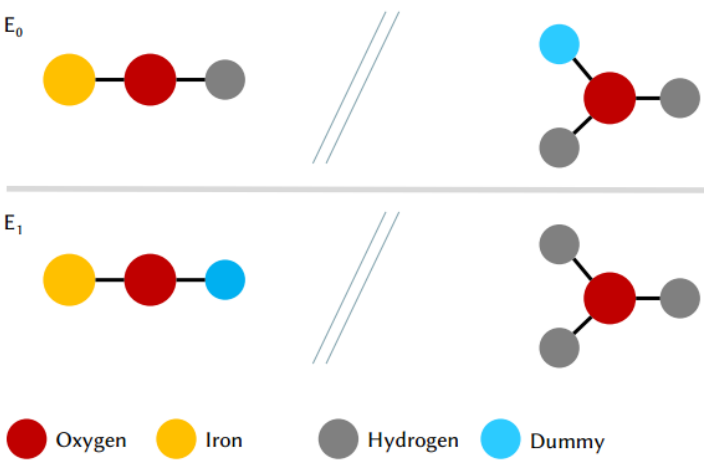
* This work ignores quantum-nuclear effects in the AIMD simulations. These nuclear effects are likely to play an important role.[320]

with the temperature T. This last relation is the basis to the computational estimation of the pK_a as applied here.[320] In this approach, the free energy difference is evaluated using thermodynamic integration.

5.1 Theory of the pKa Calculation

Based on the aforementioned definition of the pK_a, its value is estimated using dummy sites as follows.[320] In the hematite/liquid water interface setups discussed in earlier chapters of this work, there is always a (protonated) hematite slab and a bulk water slab. The free energy difference between the surface site being protonated and finding the proton solvated as hydronium in bulk water is calculated by means of an alchemical transformation, where the energy difference between these two states is evaluated along a trajectory propagated on either of the two energy surfaces or a mixture thereof. Since the vertical energy gap is only applicable to the same nuclear coordinates, the point in space where to move the proton in question to has to be known. This point is calculated during the MD run by a weakly interacting placeholder site. To this end, a dummy site is introduced as shown in Figure 5.1. This dummy site is a classical particle and carries no charge and does not interact with the DFT part of the system. In the reactant state (E_0 in Figure 5.1), the surface OH group is protonated and the dummy proton is added to an arbitrarily selected bulk water molecule to form a geometry that resembles the hydronium (H_3O^+) configuration. The dummy atom is held in this configuration by classical harmonic bond and angle potentials with this particular water molecule only. For the product state, the explicit proton and the dummy site are exchanged, which forms a DFT hydronium site in bulk water and a deprotonated OH group at the surface. The dummy site at the surface is kept in place using classical harmonic bond and angle potentials.

In both reactant and product states, all harmonic bond and angle terms are included in the propagation of the nuclei. For the evaluation of the vertical energy gap however, those classical terms are included in the energy of the states E_0 and E_1 that describe the bonding potential of the dummy site in the respective propagation. If we split the classical contributions in those for the surface where the dummy contributes E_S^{MM}, those for the hydronium ion in bulk water where the dummy contributes E_H^{MM}, and those which are only used to ensure the correct selection of configuration space E_C^{MM}, then the energies of



▶ **Figure 5.1** – Visualisation of the two topologies employed in the alchemical calculation of the free energy difference. Bulk hematite and bulk water omitted for clarity. For the E_0 PES, the explicit proton is located at the surface of bulk hematite and the dummy atom is located at a purely DFT water. For the E_1 PES, these two sites are exchanged such that an explicit hydronium ion is formed in the middle of the bulk water layer. Black lines denote bonds in the classical MD sense and, hence, are subject to classical forces.

the states i are given with the DFT total energy E_i^{QM} as

$$E_0 \equiv E_0^{\mathrm{QM}} + E_{\mathrm{H}}^{\mathrm{MM}} \tag{5.4}$$

$$E_1 \equiv E_1^{\mathrm{QM}} + E_{\mathrm{S}}^{\mathrm{MM}} \tag{5.5}$$

The terms for the correct selection of configuration space are the classical bonded and angle potentials on the hydronium site that involve only the explicit protons of the selected water molecule. They have to be imposed on the system to prevent deprotonation of the hydronium site during the MD run, but do not contribute to the energy of the dummy atom itself. While not necessary for this crystal surface, other crystal surfaces may require additional terms to stabilise the geometry that all would be included in $E_{\mathrm{H}}^{\mathrm{MM}}$.

These two different topologies also form different PES E_0 and E_1, because the explicit proton in the DFT evaluation is located at a different position. Just as outlined in the introduction of this chapter, one can interpolate between these two similar but distinct PES by introducing a mixing parameter η

$$E_\eta(\mathbf{R}) = (1-\eta)E_0(\mathbf{R}) + \eta E_1(\mathbf{R}) \tag{5.6}$$

where $\mathbf{R}$ is the position of the nuclei including the dummy site. Along this linear interpolation, the energy difference between the PES (the vertical energy gap ΔE) has to be evaluated such that the ensemble average thereof can be integrated using the thermodynamic integration method.

$$\Delta E(\mathbf{R}) \equiv E_1(\mathbf{R}) - E_0(\mathbf{R}) \tag{5.7}$$

This shows that the explicit dummy site is necessary to provide the coordinates of the proton insertion after the proton removal at the surface when going from E_0 to E_1. The thermodynamic integral

$$\Delta G \equiv G(\eta = 1) - G(\eta = 0) = \int_0^1 \left\langle \frac{\partial H(\eta)}{\partial \eta} \right\rangle_{\eta'} d\eta \tag{5.8}$$

then gives the exact free energy difference between the reactant and product states.[320]

The ensemble average of this vertical energy gap ΔE is calculated by running a AIMD simulation for all chosen values of η. It is important to perform the sampling of configurations on the correct PES, since e.g. the solvation structure depends on the location of the explicit proton, which is why the conformational space sampled in E_1 can be essentially different from the one sampled in E_0. The same argument holds for the intermediate (artificial) PES. If the trajectories obtained this way are long enough, the variance of the vertical energy gap ΔE can be estimated. This is of interest, because the variance of the raw time series directly relates to the derivative of the ensemble vertical energy gap w.r.t. the mixing parameter[321]

$$\frac{\partial}{\partial \eta} \langle \Delta E \rangle_\eta = -\frac{\sigma_\eta^2}{k_B T} = -\frac{\langle (\Delta E_\eta - \langle \Delta E \rangle_\eta)^2 \rangle}{k_B T} \tag{5.9}$$

This allows for additional information from the MD data. Depending on the relative convergence speed of the ensemble average of ΔE and the variance thereof, this may contribute to reduce the statistical uncertainty associated with the final estimate.

To relate the thermodynamic integral (eqn 5.8) to the pK_a, a number of correction terms are necessary.[320] First, one needs to estimate the free energy to remove the artificial dummy sites from reactant and product states. This contribution, ΔA_{DI}, can be estimates as

$$\Delta A_{DI} \equiv k_B T \ln \frac{q^{\text{int}}_{\text{Ad}^-} q^{\text{int}}_{\text{H}_2\text{O}}}{q^{\text{int}}_{\text{A}^-} q^{\text{int}}_{\text{H}_2\text{Od}}} \tag{5.10}$$

Additionally, the resulting thermodynamic integral has to be corrected for treating the dummy at classical level but having an explicit quantum proton at DFT level (the quantum correction ΔA_{QM}):

$$\begin{aligned} \Delta A_{QM} &= -\Delta A_{\text{qc}}(\text{AH}) + \Delta A_{\text{qc}}(\text{H}_3\text{O}^+) \\ &\equiv \sum_{i=1}^{6} \frac{h}{2} \nu_i' - \sum_{i=3}^{3} \frac{h}{2} \nu_i'' + \beta^{-1} \ln \left(\prod_{i=1}^{6} \frac{1}{h\nu_i' \beta} \prod_{i=1}^{3} \frac{\nu_i''}{\nu_i} \prod_{i=1}^{3} \frac{1 - \exp(-h\nu_i \beta/2)}{\exp(-h\nu_i \beta/2)} \right) \end{aligned} \tag{5.11}$$

where ν_i are the DFT vibrational frequencies for the proton at the surface, ν_i' are the DFT vibrational frequencies for the hydronium ion, and ν_i'' are the DFT frequencies for a water molecule.

The final correction consists of the difference between the Brønsted and Arrhenius picture of the proton in bulk water. During the calculation, an explicit hydronium ion is used rather than a solvated proton which is a migrating charged defect in water. The free energy difference ΔA_{DH} between the hydronium ion and the solvated proton has been estimated to be

$$\Delta A_{DH} \equiv k_B T \ln c^\circ \Lambda^3_{\mathrm{H^+}} \tag{5.12}$$

Including all corrections, the expression for the pK_a reads[320]

$$pK_a = \frac{1}{\ln 10 k_B T}\left(\int_0^1 \langle \Delta E \rangle_\eta d\eta + k_B T \ln \frac{q^{\mathrm{int}}_{\mathrm{Ad^-}} q^{\mathrm{int}}_{\mathrm{H_2O}}}{q^{\mathrm{int}}_{\mathrm{A^-}} q^{\mathrm{int}}_{\mathrm{H_2Od}}} - \Delta A_{\mathrm{qc}}(\mathrm{AH}) + \Delta A_{\mathrm{qc}}(\mathrm{H_3O^+}) + k_B T \ln c^\circ \Lambda^3_{\mathrm{H^+}} \right) \tag{5.13}$$

While the original work suggests to use two half-reactions[320] in order to re-use part of the simulation for the variety of compounds they analyse, this offers no benefit for a single configuration as in this work. Therefore, the direct conversion from an explicit proton at the surface to an explicit proton in bulk water is calculated in this work.

The first term of eqn 5.13 is the aforementioned thermodynamic integral which is obtained by weighted summation over the intermediate points η. The second term gives the free energy of dummy insertion and can be obtained from the classical vibrational frequencies for the protonated surface, the classical vibrational frequencies for the hydronium molecule and the DFT frequencies of a single water molecule. Since an evaluation of the vibrational frequencies of the surface proton would require a relaxed interface geometry first, a slightly different approach is used for the hematite surface. Instead of evaluating global modes, their local equivalent is calculated by drawing a snapshot from the equilibrated MD trajectory, freezing all atoms except for the surface proton, followed by a geometry optimisation (which then only affects the surface proton). For the resulting geometry, the Hessian and, thereby, the vibrational frequencies can be extracted.

The next correction terms, $-\Delta A_{\mathrm{qc}}(\mathrm{AH}) + \Delta A_{\mathrm{qc}}(\mathrm{H_3O^+})$, are the quantum corrections which account for the difference arising from the classical treatment of the proton instead

of treating it fully at DFT level. To this end, the DFT frequencies for all relevant subsystems (surface proton, hydronium ion and water) are used.

The last correction is the dissociation free energy accounting for the difference in energy between a hydronium ion (the description used in the simulation) and a solvated proton with a water molecule. Here, Λ is the thermal wavelength of the proton at simulation temperature.

5.1.1 Thermodynamic Integration

This method is not specific to the estimation of pK_a, but pertains to calculations of free energy differences in general.[322] The general approach is to first define a mixing parameter or reaction coordinate η connecting the two system states the free energy difference of which is to be estimated. Then, a series of calculations is performed for various values of η to sample a superposition of initial and final system. Finally, the values for the different calculations η are combined by integration. For this last step, there are multiple ways. Some of which will be explained below.

Generally, the free energy G depends on the mixing parameter η indirectly via the partition sum Z or the Hamiltonian H

$$G(\eta) = -k_B T \ln Z(\eta) \Rightarrow \frac{\partial G(\eta)}{\partial \eta} = \left\langle \frac{\partial H(\eta)}{\partial \eta} \right\rangle_{\eta'} \tag{5.14}$$

where $\langle \cdot \rangle_\eta$ denotes the ensemble average for the mixed systems for a given parameter η. Conventionally, η is chosen such that the initial state has $\eta = 0$ and the final state has $\eta = 1$. In this case, the free energy difference ΔG is given by

$$\Delta G \equiv G(\eta = 1) - G(\eta = 0) = \int_0^1 \left\langle \frac{\partial H(\eta)}{\partial \eta} \right\rangle_{\eta'} d\eta \tag{5.15}$$

This resulting integral can be evaluated in several ways: performing MD calculations for several fixed values of η,[320] slowly increasing η over the course of a long MD calculation,[322] quickly changing η but going forth and back between initial and final state multiple times,[323] performing a forward and backward sweep of η followed by averaging,[324] adding intermediate values of η on an as-needed basis,[325] tailoring the η values to the physical interaction,[326] sampling η via replica exchange MC calculations,[327] and propagating in η.[328] In all these cases, the integral is approximated by a sum over finite values of η and some kind of integration scheme has to be applied.

The integration scheme to be used depends on the underlying physical quantity and system under investigation. If the values for η are equally spaced over the interval $[0, 1]$, then the Newton-Cotes method can be applied, given that the number of values of η are sufficient to capture any (potential) curvature of the integrand over the interval. In case of non-equally spaced values, either composite Newton-Cotes can be applied to sub-intervals of equal spacing or other more complex numerical integration algorithms have to be chosen. This is the reason why equally spaced intervals of η are common. In some cases, it may be helpful to include the derivative of the integrand at the points of fixed η as well. Whether this is feasible and brings an advantage depends on the physical property and the simulation method. One application where this is helpful is the estimation of pK_a values, as will be shown below.

5.2 Integration Methods

For the different integration methods, different uncertainty estimates for the final integral arise. They can be derived by applying propagation of uncertainty methods to the integration formulas. It is however crucial to emphasise that the discussion in this subsection only applies to the statistical *uncertainty*, not to the systematic *error* which is inherently defined by the approach, the force field, the level of theory for the AIMD calculation and the finite sampling. In this chapter, the term *uncertainty* will be used to refer to the statistical component and *error* will refer to the deviation from the true value.

5.2.1 Newton-Cotes

While the underlying theory is valid for all numbers of intermediate points of η, this work will focus on up to five total points of η, since this is the integration done later on to determine the pK_a uses this many points.

One method is composite application of Simpson's rule:

$$I^{\mathrm{S}} = \frac{1}{6}\left(\langle Q\rangle_0 + \langle Q\rangle_1\right) + \frac{2}{3}\langle Q\rangle_{0.5} \tag{5.16}$$

where Q is the integrand. While Simpson's rule only uses three points, the composite version thereof consists of splitting the overall integral in two intervals $\eta \in [0, 0.5]$ and

$\eta \in [0.5, 1]$. Then, the total integral I^{SC} can be expressed as

$$I^{SC} = \frac{1}{12}\left(\langle Q\rangle_0 + 2\langle Q\rangle_{0.5} + \langle Q\rangle_1\right) + \frac{1}{3}\left(\langle Q\rangle_{0.25} + \langle Q\rangle_{0.75}\right) \tag{5.17}$$

in this case, the uncertainty for the simple Simpson's rule as calculated from propagation of uncertainty is

$$\Delta I^{S} = \sqrt{\sum_i \left|\frac{\partial I^{S}}{\partial \langle Q\rangle_i}\right|^2 \Delta(\langle Q\rangle_i)^2} = \sqrt{\frac{1}{36}\left[\Delta(\langle Q\rangle_0)^2 + \Delta(\langle Q\rangle_1)^2\right] + \frac{4}{9}\Delta(\langle Q\rangle_{0.5})^2} \tag{5.18}$$

$$\Rightarrow \Delta I^{SC} = \sqrt{\frac{\Delta(\langle Q\rangle_0)^2}{144} + \frac{1}{9}\Delta(\langle Q\rangle_{0.25})^2 + \frac{\Delta(\langle Q\rangle_{0.5})^2}{36} + \frac{1}{9}\Delta(\langle Q\rangle_{0.75})^2 + \frac{\Delta(\langle Q\rangle_1)^2}{144}} \tag{5.19}$$

where Δ denotes the absolute uncertainty and i runs over all discrete values of η. This assumes that the individual observations of $\langle Q\rangle_i$ are uncorrelated and, hence, $\mathrm{Cov}(\langle Q\rangle_i, \langle Q\rangle_j) = 0$ for all $i \neq j$. The integration error (i.e. the error between the true integral and the approximation) is proportional to the fourth derivative of the integrand (at one point within the interval) and proportional to the fifth power of the interval width. Since pK_a curves of the integrand often exhibit cubic curvature at most,[329] this error term is zero or at least very close to zero.

Instead of composite Newton-Cotes, the direct equation for five points, Boole's rule, can be used:

$$I^{B} = \frac{1}{90}\left[7(\langle Q\rangle_0 + \langle Q\rangle_1) + 32(\langle Q\rangle_{0.25} + \langle Q\rangle_{0.75}) + 12\langle Q\rangle_{0.5}\right] \tag{5.20}$$

Straightforward application of the propagation of uncertainty yields

$$\Delta I^{B} = \frac{1}{15}\sqrt{\frac{49}{36}\left[\Delta(\langle Q\rangle_0)^2 + \Delta(\langle Q\rangle_1)^2\right] + \frac{256}{9}\left[\Delta(\langle Q\rangle_{0.25})^2 + \Delta(\langle Q\rangle_{0.75})^2\right] + 4\Delta(\langle Q\rangle_{0.5})^2} \tag{5.21}$$

with the same assumption on the covariance of the individual observations. Here, the integration error scales with the sixth derivatives of the integrand at a point within the interval and the seventh power of the integration range. Therefore, the aforementioned argument still holds that this error is vanishingly small for the integrands of interest.

Only in the linear response model, the initial and final ensembles are the only ones to be considered. In this case, the Trapezoid rule gives an integral value

$$I^{TR} = \frac{\langle Q\rangle_0 + \langle Q\rangle_1}{2} \qquad \Delta I^{TR} = \sqrt{\frac{\Delta(\langle Q\rangle_0)^2 + \Delta(\langle Q\rangle_1)^2}{4}} \tag{5.22}$$

5.2.2 Hermite Splines

Another integration method is based on Hermite splines for equidistant values. In this case, every segment of the interval is interpolated using cubic splines with neighbouring intervals serving as mutual constraint on the derivatives to ensure smooth derivatives across the whole integration interval. This approach allows for inclusion of the derivatives of the integrand in the fit. The cubic splines c_i for the segments i are defined using the Hermite basis functions h_{ii}

$$c_i(x) = h_{00}(t_i)\langle Q\rangle_i + h_{10}(t_i)h\frac{\partial\langle Q\rangle_i}{\partial\eta} + h_{01}(t_i)\langle Q\rangle_{i+1} + h_{11}(t_i)h\frac{\partial\langle Q\rangle_{i+1}}{\partial\eta} \tag{5.23}$$

$$t_i(x) \equiv \frac{x - \langle Q\rangle_i}{\langle Q\rangle_{i+1} - \langle Q\rangle_i} \tag{5.24}$$

$$h_{00}(t_i) = 2t_i^3 - 3t_i^2 + 1; \qquad h_{10}(t_i) = t_i^3 - 2t_i^2 + t_i \tag{5.25}$$

$$h_{01}(t_i) = -2t_i^3 + 3t_i^2; \qquad h_{11}(t_i) = t_i^3 - t_i^2 \tag{5.26}$$

where h is the width between consecutive data points (so $h = 0.25$ for five points), and t linearly covers a $[0,1]$ between two consecutive data points. These cubic splines can be piecewise integrated over the interval they are defined on. For a single interval $[x_i, x_{i+1}]$, the integral of $c_i(x)$ is

$$\int_{x_i}^{x_{i+1}} c_i(x)dx = \underbrace{(x_{i+1} - x_i)}_{\equiv h}\int_0^1 c_i(t')dt' = h\left(\frac{\langle Q\rangle_i + \langle Q\rangle_{i+1}}{2} + \frac{h}{12}\left(\frac{\partial\langle Q\rangle_i}{\partial\eta} - \frac{\partial\langle Q\rangle_{i+1}}{\partial\eta}\right)\right) \tag{5.27}$$

which would be applicable for cases where no intermediate points are taken into consideration (but now including the derivatives):

$$I^{\mathrm{SHS}} = \frac{\langle Q\rangle_0 + \langle Q\rangle_1}{2} + \frac{1}{12}\left(\frac{\partial\langle Q\rangle_0}{\partial\eta} - \frac{\partial\langle Q\rangle_1}{\partial\eta}\right) \tag{5.28}$$

$$\Delta I^{\mathrm{SHS}} = \sqrt{\frac{\Delta(\langle Q\rangle_0)^2 + \Delta(\langle Q\rangle_1)^2}{4} + \frac{1}{144}\left(\Delta\left(\frac{\partial\langle Q\rangle_0}{\partial\eta}\right)^2 + \Delta\left(\frac{\partial\langle Q\rangle_1}{\partial\eta}\right)^2\right)} \tag{5.29}$$

Addition of all segments i spanning the integral range $[\eta = \eta(i), \eta = \eta(i+1)]$ yields the total integral

$$I^{\mathrm{HS}} = \sum_i \int_{x_i}^{x_{i+1}} c_i(x)dx \tag{5.30}$$

$$= \frac{1}{4}\left(\frac{\langle Q\rangle_0 + \langle Q\rangle_1}{2} + \langle Q\rangle_{0.25} + \langle Q\rangle_{0.5} + \langle Q\rangle_{0.75}\right) + \frac{1}{192}\left(\frac{\partial\langle Q\rangle_0}{\partial\eta} - \frac{\partial\langle Q\rangle_1}{\partial\eta}\right) \tag{5.31}$$

Name		$\partial\langle Q\rangle_0/\partial\eta$	$\langle Q\rangle_0$	$\langle Q\rangle_{0.25}$	$\langle Q\rangle_{0.5}$	$\langle Q\rangle_{0.75}$	$\langle Q\rangle_1$	$\partial\langle Q\rangle_1/\partial\eta$
Trapezoid rule	I^{TR}		x				x	
Simpson's rule	I^{S}		x		x		x	
Composite Simpson's rule	I^{SC}		x	x	x	x	x	
Boole's rule	I^{B}		x	x	x	x	x	
Single Hermite spline	I^{SHS}	x	x				x	x
Hermite spline	I^{HS}	x	x	x	x	x	x	x

▲ **Table 5.1** – Integration formulas and which information they require on the integrand $\langle Q\rangle_\eta$.

because all intermediate derivatives cancel each other out. For the statistical uncertainty, this means

$$\Delta I^{\mathrm{HS}} = \frac{1}{4}\left[\left(\frac{\Delta(\langle Q\rangle_0)^2+\Delta(\langle Q\rangle_1)^2}{4}+\Delta(\langle Q\rangle_{0.25})^2+\Delta(\langle Q\rangle_{0.5})^2+\Delta(\langle Q\rangle_{0.75})^2\right)+\frac{1}{2304}\left(\Delta\left(\frac{\partial\langle Q\rangle_0}{\partial\eta}\right)^2+\Delta\left(\frac{\partial\langle Q\rangle_1}{\partial\eta}\right)^2\right)\right]^{1/2} \tag{5.32}$$

All these integration methods require different information of the integrand and are compared in Table 5.1. The main difference is the different weighting of the various pieces of information. Since the statistical uncertainty of the final estimate highly depends on the weighting of the individual contributions, it is generally advisable to choose a integration method that reduces the uncertainty as efficiently as possible with increasing budget spent on the calculations. In the context of the pK_a calculations carried out in this work, this means that methods are to be avoided where the uncertainty depends on the intermediate trajectories (where $1 \neq \eta \neq 0$), since they require twice as much computational effort. Some of the integration formulas include derivative information of the integrand, which may increase accuracy, if and only if the statistical uncertainty of the derivative information decreases quickly enough. However, the prefactors of the terms suppress the derivative contribution to the statistical uncertainty significantly. Since the derivative information is related to the second moment of the time series and second moments generally converge slower than the plain average, these form competing effects. This is discussed later in this work.

5.3 Optimal Sampling Method for Thermodynamic Integration

When performing MD calculations to obtain the time series data that is required to evaluate the ensemble average of any quantity (not only the vertical energy gap), there is always the question of convergence. While it can be beneficial to plot the value of the observable over time, there needs to be hard criterion separating equilibration phase from the production part of the recorded data. This is particularly challenging for short trajectories, which is the typical case for the AIMD simulations carried out in this work. With short trajectories, a slight change of the choice of the extent of the production time series may yield significantly different results, which would contribute to the systematic error of the work. Therefore, it is desirable to devise a way of automatically decide on the extent of equilibration in a time series – based just on the original properties of the time series itself. This is particularly helpful, because it allows automation of this decision and enables to employ different equilibration times for different quantities, since their underlying physical equilibration processes may complete on different times. Also, an automated decision supports reproducibility.

In a bid to base the selection of the production part of a time series of an observable, it is suggested to follow the reverse cumulative average approach RCA.[330] Their overall approach is to analyse the reversed time series. If a time series is converged, then this is the case towards the end of it, not at the beginning. They then analyse the time series using block averaging and detect starting from where the mean and variance of the individual blocks do not follow a normal distribution any more. This point then marks the end of the (reversed) equilibrated data. The detailed formalised approach will be shown in the following. However, the original approach[330] has been extended in this thesis to deal with non-equally spaced data due to missing information, and frequency weighting based on Boltzman reweighting.

5.3.1 Proposed Method

Given is a non-stationary time series Q_i of an observable $Q(t)$ for a ordered set of N irregularly spaced times t_i fully spanning the interval $[0, T]$. Now the goal is to obtain an estimate for the equilibrated continuous subset of data points Q_i that then can be used to calculate an estimate for the true ensemble average (or expectation value of the time series), $\langle Q \rangle$,

the variance thereof, $\mathrm{Var}(\langle Q \rangle)$, the variance of the time series, $\mathrm{Var}(Q(t))$, and the variance thereof, $\mathrm{Var}(\mathrm{Var}(Q(t)))$. Each data point has an associated weighting factor $w_i(Q_i)$.

The first step is to reverse the overall time series Q_i:

$$Q_i''(t_i') \equiv Q_{N-i+1}(t_{N-i+1}) \qquad \text{with} \qquad t_i' \equiv T - t_i \tag{5.33}$$

$$w_i'(t_i') \equiv w_{N-i+1}(t_{N-i+1}) \tag{5.34}$$

This step is equivalent with the time-reversal of the MD calculation. After this step, the linear trend that is likely to be present is removed by linear regression using the OLS estimator. The resulting data is referred to as $Q_i'(t_i')$. Now the data is divided into $n \equiv \lfloor T/\tau \rfloor$ blocks of duration τ each. This means that the blocks all have the same duration in the time series no matter how many actual observations are part of them. Data from an incomplete last block is discarded. For each of these blocks, there is one set (in set-builder notation)

$$B_j^\tau \equiv \{Q_i'(t_i') : (j-1)\tau > t_i' > j\tau\} \tag{5.35}$$

With these definitions, the weighted average value $\overline{Q_j^\tau}$ and the frequency weighted sample variance $\overline{V_j^\tau}$ for the frequency weighted data within a block is given as

$$\overline{Q_j^\tau}' \equiv \left[\sum\{w_i' : Q_i' \in B_j^\tau\}\right]^{-1} \sum\{w_i' Q_i' : Q_i' \in B_j^\tau\} \tag{5.36}$$

$$\overline{V_j^\tau}' \equiv \left[\sum\{w_i' : Q_i' \in B_j^\tau\} - 1\right]^{-1} \sum\{w_i'(Q_i' - \overline{Q_j^\tau}')^2 : Q_i' \in B_j^\tau\} \tag{5.37}$$

Here and in all subsequent steps, empty sets are not taken into account when performing the next step in the analysis. These resulting averages depend on the parameter τ, which will be chosen such that the n individual blocks of length τ become statistically independent. This is estimated using the concept of statistical inefficiency:[330] the one estimator (for the ensemble property $\langle Q \rangle$) that needs the fewest observations to obtain a given uncertainty is called the most efficient one. The statistical inefficiency Φ is defined as

$$\Phi(\tau) \equiv \frac{\tau \cdot \mathrm{Var}(\{\overline{Q_j^\tau}'\})}{\mathrm{Var}(\{Q_i'\})} \tag{5.38}$$

For large blocks (i.e. large durations τ), this should converge to a fixed value, since the statistically independent chunks estimate the same variance as for the whole set of data with the exception of the finite size, i.e. the difference between sampling and population variance. Since $\Phi(\tau)$ is subject[330] to noise itself (in particular for large τ such that there are only few blocks in the whole time series), the statistical inefficiency will be considered

locally stationary w.r.t. τ according to the augmented Dickney Fuller (ADF) test, a standard tool in statistics. Due to the finite length of the underlying trajectory, there can only be a limited number of possible values of τ. Therefore, $\Phi(\tau)$ can be evaluated for all possible values of τ. Practically, there are two limits that can be exploited here: τ has to be at least the time distance between the closest samples Q_i in the time series and can be $T/4$ at most. The latter limit is a constraint that arises from the normality tests outlined below. With the results from the ADF, the set τ^Φ of all possible values of τ for which Φ is locally stationary and, therefore, the individual blocks are independent, is defined.

At this point in the analysis, it is only known which block sizes τ produce statistically independent blocks. However, the average values $\overline{Q_j^\tau}'$ of these blocks have to follow a normal distribution if they truly estimate the same equilibrated time series. Note that this is not an assumption about the distribution of the individual values Q_i (which in fact can be arbitrarily distributed), but only about the average values of the blocks. If τ is large enough, the central limit theorem states that the average has to follow a normal distribution regardless of the underlying data series. If, however, a contribution of the equilibration part of the time series is still present, this gives rise to a different average value. This different average value would not match the normal distribution of the equilibrated part of the time series.

The crucial step to determine whether the time series is equilibrated is now to follow the blocks in the order of the reversed time series and collect their averages. The Shapiro-Wilk test then shows whether the average of the next block matches the assumption of a normal distribution. This test, however, requires at least three data points, which is the reason for the aforementioned limit on potential block sizes. This way, for each block size τ, the boundary of the equilibration region can be defined based on the averages of blocks which follow a normal distribution $\mathcal{N}$ in reverse order of the original trajectory:

$$C(\tau) \equiv T - \tau j \qquad \text{where} \quad j : \forall i \leq j : \{\overline{Q_i^\tau}'\} \in \mathcal{N} \wedge \{\overline{Q_i^\tau}' : i \leq j+1\} \notin \mathcal{N} \wedge j \geq 3 \quad (5.39)$$

Practically, this only has to be calculated for all block lengths τ from τ^Φ, since for all other block durations, the statistical inefficiency is not locally stationary, which consequently does not rule out correlation between the individual blocks.

Now the final equilibrated value and the statistical uncertainty thereof can be calculated using the original time series. This is done by building the same blocks as for the reverse

time series just for the forward data, which not has been altered in any way:

$$D_j^\tau \equiv \{Q_{N-i+1}(t_{N-i+1}) : Q_i'(t_i') \in B_j^\tau\} \tag{5.40}$$

For each of these blocks, the following quantities are required for further analysis: average $\overline{Q_j^\tau}$, the frequency weighted variance $\overline{V_j^\tau}$, and the variance of this variance $\overline{W_j^\tau}$:

$$\overline{Q_j^\tau} \equiv \left[\sum\{w_i : Q_i \in D_j^\tau\}\right]^{-1} \sum\{w_i Q_i : Q_i \in D_j^\tau\} \tag{5.41}$$

$$\overline{V_j^\tau} \equiv \left[\sum\{w_i : Q_i \in D_j^\tau\} - 1\right]^{-1} \sum\{w_i (Q_i - \overline{Q_j^\tau})^2 : Q_i \in D_j^\tau\} \tag{5.42}$$

$$\overline{W_j^\tau} \equiv \frac{\sum\{w_i (Q_i - \overline{Q_j^\tau})^4 : Q_i \in D_j^\tau\}}{|D_j^\tau| \left(\sum\{w_i : Q_i \in D_j^\tau\} - 1\right)} - \frac{\overline{V_j^\tau}^2 (|D_j^\tau| - 3)}{|D_j^\tau|(|D_j^\tau| - 1)} \tag{5.43}$$

With these quantities, all individual components can be pieced together: potential solutions τ^S are given by only those block sizes τ for which the statistical inefficiency $\Phi(\tau)$ is locally stationary, and where a normal distribution of the block averages has been found are to be considered. All in all:

$$\tau^S \equiv \{\tau \in [\tau_m, T/4] : \forall i, j : \tau_m \leq |t_i - t_j| \wedge \tau \in \tau^\Phi \wedge C(\tau) \in [0, T]\} \tag{5.44}$$

Based on the individual blocked averages and variances, the estimated properties of the whole equilibrated part of the trajectory can be estimated for the case of block lengths larger than the autocorrelation time of the time series. Note that in this case the variances of the blocks are not frequency information but rather uncertainties. Therefore, the following expressions are to be used:

$$\langle Q \rangle_\tau \equiv \frac{\sum_j \overline{Q_j^\tau}\, \overline{V_j^\tau}^{-1}}{\sum_j \overline{V_j^\tau}^{-1}} \tag{5.45}$$

$$\mathrm{Var}(\langle Q \rangle)_\tau \equiv \left[\sum_j \overline{V_j^\tau}\right]^{-1} \tag{5.46}$$

$$\mathrm{Var}(Q(t))_\tau \equiv \frac{\sum_j \overline{V_j^\tau}\, \overline{W_j^\tau}^{-1}}{\sum_j \overline{W_j^\tau}^{-1}} \tag{5.47}$$

$$\mathrm{Var}(\mathrm{Var}(\langle Q \rangle))_\tau \equiv \sum_j \overline{W_j^\tau} \left[\sum_j 1\right]^{-1} \tag{5.48}$$

$$\tag{5.49}$$

However, for a converged trajectory, all block sizes that fulfil these criteria have to yield a statistically compatible estimate. If and only if the following conditions hold, the

trajectory can be considered to have an equilibrated section:

$$\tau^S \neq \emptyset \tag{5.50}$$

$$\forall \tau_i, \tau_j \in \tau^S : |\langle Q \rangle_{\tau_i} - \langle Q \rangle_{\tau_j}| \leq 2 \cdot \mathrm{Var}(\langle Q \rangle)_{\tau_j} \tag{5.51}$$

If the first condition is not met, then there is no equilibrated section at all. If the latter condition is not met, this means that there are statistically incompatible estimates for different block sizes τ.

The final estimates for the underlying physical process are for the one $\tau_i \in \tau^S$ that minimises the variance $\mathrm{Var}(\langle Q \rangle)_\tau$. The estimated mean (with the statistical uncertainty) then is

$$\langle Q \rangle = \langle Q \rangle_{\tau_i} \quad \pm \quad \sqrt{\mathrm{Var}(\langle Q \rangle)_{\tau_i}} \tag{5.52}$$

In the same way, the variance of the physical process (with the associated statistical uncertainty) is obtained as

$$\mathrm{Var}(Q(t)) = \mathrm{Var}(Q(t))_{\tau_i} \quad \pm \quad \sqrt{\mathrm{Var}(\mathrm{Var}(\langle Q \rangle))_{\tau_i}} \tag{5.53}$$

5.3.2 Efficiency and Automation

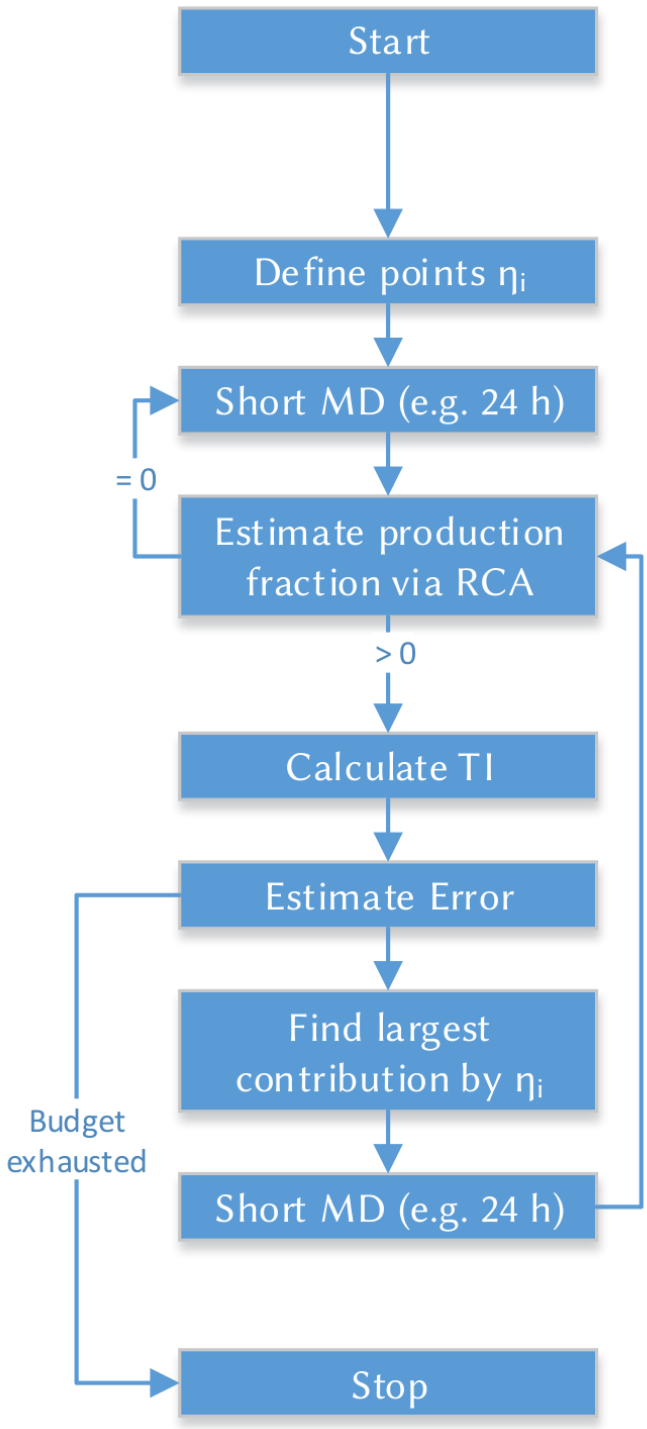
The method proposed above to select the equilibrated part of a time series can be integrated in a guided thermodynamic integration approach in order to yield reliable estimates of the pK_a value with the least required budget and without the arbitrariness of human decision. Figure 5.2 shows the general scheme to be used. Initially, a number of mixing parameters η_i are selected based on the expected curvature of the vertical energy gap over the interval $\eta \in [0, 1]$. The only condition here is that the initial and final state have to be sampled with their respective trajectories. Based on these pre-defined points η_i, short MD trajectories are performed for each η_i. Short trajectories in this context represents a meaningful length of an AIMD run such that not too much time is spend on start-up and initialisation of the MD. For the systems at hand, one day of walltime seems to be reasonable. After there is MD data for each η_i, the aforementioned method selects whether all simulations are equilibrated. If not, then the non-equilibrated trajectories are extended in increments of these short runs. Once the MD has been determined to be equilibrated for all values η_i,

the thermodynamic integral can be evaluated using one of the methods outlined at the beginning of this chapter. The required data points and their statistical uncertainty is a result of the equilibrated trajectories. This gives not only an initial value for the pK_a, but also the statistical uncertainty thereof. The individual contributions to these uncertainties are available analytically for the individual values η_i which means that the one single η_j can be determined which has the largest contribution to the overall error estimation. This single trajectory η_j can then be extended by a short MD run. Afterwards, the new equilibrated region of that η_j has to be obtained using the block sampling method outlined above. With this new set of information, another estimate of the integral and, hence, the pK_a can be calculated. This process should be repeated until either the budget is exhausted or the statistical uncertainty for the pK_a is sufficiently small.

This has the desirable property that there is a hard criterion suggesting which trajectory to extend. This criterion is based on the relative quality of all trajectory data available so far and chooses the values of η_i that are most likely to reduce the statistical uncertainty. While human interaction may be helpful to control the suggestions of the algorithm, this procedure greatly reduces the manual effort related to the calculation.

5.4 Computational Details

When preparing the AIMD simulation, a pre-equilibrated snapshot from the interface calculations has been used. However, there are additions required for several reasons. First, the dummy site has to be introduced (simulation a03). This is easily done by selecting a random water molecule in the middle of the water layer of the setup and roughly positioning the dummy atom at a location close to the configuration of a hydronium ion. This dummy site has to be held in place (relative to the water molecule in question) by classical force field terms, such that at any given time, there is a well-defined insertion position for the explicit proton when evaluating the vertical energy gap. However, once this explicit proton is switched to that dummy site, over the course of the E_1 trajectory, other protons of that water molecule may dissociate, since it is a regular AIMD simulation. To prevent this from happening, the same classical terms that keep the dummy proton in place are added to the other protons of the same water molecule. All other water molecules remain unchanged. In this work, the force constants as parametrised in literature have been



▲ **Figure 5.2** – Schema on obtaining the pK_a value via thermodynamic integration in an automated fashion.

	k_r [au]	r_{eq} [Å]	k_θ [au]	θ_{eq}
Hematite (surface O) - (surface H)	0.2	1.00		
Hematite (surface O) - (surface O) - (surface H)			0.00063043	90
Hydronium[320]	0.2	1.00	0.2	111

▲ **Table 5.2** – Classical force field parameters (force constants k, equilibrium distance r_{eq}, and equilibrium angle θ_{eq}) used in the pK_a calculation.

Calculation(s) 952

used.[320]

Secondly, the dummy terms that would be required for the E_1 trajectory have to be added. While the physical reasoning is the same as for the aforementioned classical terms on the water molecule, these potentials are added to the surface OH group. The force constants here have been derived from the thermal fluctuations of the equilibrated interface AIMD data.

Thirdly, auxiliary terms are required to prohibit those relaxation processes of the system where unprotonated sites would become protonated by migrating protons at the surface. Since the c-cut surface does not exhibit deprotonation events during the course of the neutral trajectories as outlined in previous chapters, there are no such term in this specific setup, but they are required for future investigation of e.g. the r-cut, where fast proton movement has been observed.

Table 5.2 lists the force constants as used for the description for the dummy particle. The values for the O-H bond and H-O-H angle have been taken from literature,[320] while the very weak O-O-H force constant has been used to avoid that the dummy site penetrates the surface and reaches unphysical configurations. The force constant has been chosen such that the dummy proton samples a similar configuration space as the explicit proton.

5.5 Results

Based on the aforementioned setup procedure, the pK_a of the single proton on the hematite c-cut surface has been evaluated. For this evaluation, the estimation of the statistical uncertainty has been done according to the scheme outlined in this chapter. The same applies to the selection of the equilibrated section of the time series. In total, five equidistant values of η have been chosen: 0, 0.25, 0.5, 0.75, 1. While the MD simulations and the calculation

of the time series of the vertical energy gap including the local mode calculations has been done by the student Oliver Gittus under my supervision, the analysis presented here is solely my own work. All the estimators and algorithms in this chapter have been implemented by me. The python code employs the external packages *numpy*, *scipy*, *pandas*, and *statsmodels*.

For all values of η, both the ensemble average vertical energy gaps have been determined over trajectories of 29 ps in total:

$$\langle \Delta E \rangle_{\eta=0.00} = (0.259 \pm 0.013)\,\text{Hartree}$$
$$\langle \Delta E \rangle_{\eta=0.25} = (0.172 \pm 0.004)\,\text{Hartree}$$
$$\langle \Delta E \rangle_{\eta=0.50} = (0.013 \pm 0.008)\,\text{Hartree}$$
$$\langle \Delta E \rangle_{\eta=0.75} = (-0.071 \pm 0.004)\,\text{Hartree}$$
$$\langle \Delta E \rangle_{\eta=1.00} = (-0.165 \pm 0.007)\,\text{Hartree}$$

For the initial and final trajectory, the partial derivatives have been obtained as well:

$$\partial \langle \Delta E \rangle_{\eta=0} / \partial \eta = (-0.5735 \pm 0.0048)\,\text{Hartree}$$
$$\partial \langle \Delta E \rangle_{\eta=1} / \partial \eta = (-0.1376 \pm 0.0017)\,\text{Hartree}$$

For each of the trajectories, the vertical energy gap has been evaluated in a strided pattern rather than on each MD frame. The striding steps of 25 fs have been obtained from the decay time of the autocorrelation function ACF of the continuously sampled vertical energy gap on the $\langle \Delta E \rangle_0$ surface. For the initial and final states, a replay of the corresponding PES was required: since the vertical energy gap is $E_1 - E_0$, both PES need to be calculated. For all intermediate states, a replay was not necessary, because the individual energy components already have been calculated during the MD to allow for PES interpolation using the mixing parameter η.

With the calculated frequencies in Table 5.3, the correction terms in eqn. 5.13 can be calculated as follows:

$$k_B T \ln \frac{q^{\text{int}}_{\text{Ad}^-} q^{\text{int}}_{\text{H}_2\text{O}}}{q^{\text{int}}_{\text{A}^-} q^{\text{int}}_{\text{H}_2\text{Od}}} = 50.9\,\text{meV} \tag{5.54}$$

$$-\Delta A_{\text{qc}}(\text{AH}) + \Delta A_{\text{qc}}(\text{H}_3\text{O}^+) = 39.9\,\text{meV} \tag{5.55}$$

$$k_B T \ln c^\circ \Lambda^3_{\text{H}^+} = -190.9\,\text{meV} \tag{5.56}$$

	Literature QM [cm^{-1}]	Literature MM [cm^{-1}]	QM [cm^{-1}]	MM [cm^{-1}]
H_3O^+	1032	1450	1149	1672
	1668	1810	1514	1830
	1692	1920	1754	2365
	3449	3340	3458	2504
	3523	3660	3741	4380
	3540	3760	3762	4459
AH	577	1186	922	376
	1141	1797	1406	398
	3446	2290	3094	2290
H2O	1610		1652	1776
	3650		3698	2345
	3750		3771	2380

▲ **Table 5.3** – Frequencies for inclusion in the correction terms for the pK_a expression in comparison with literature values.[320]

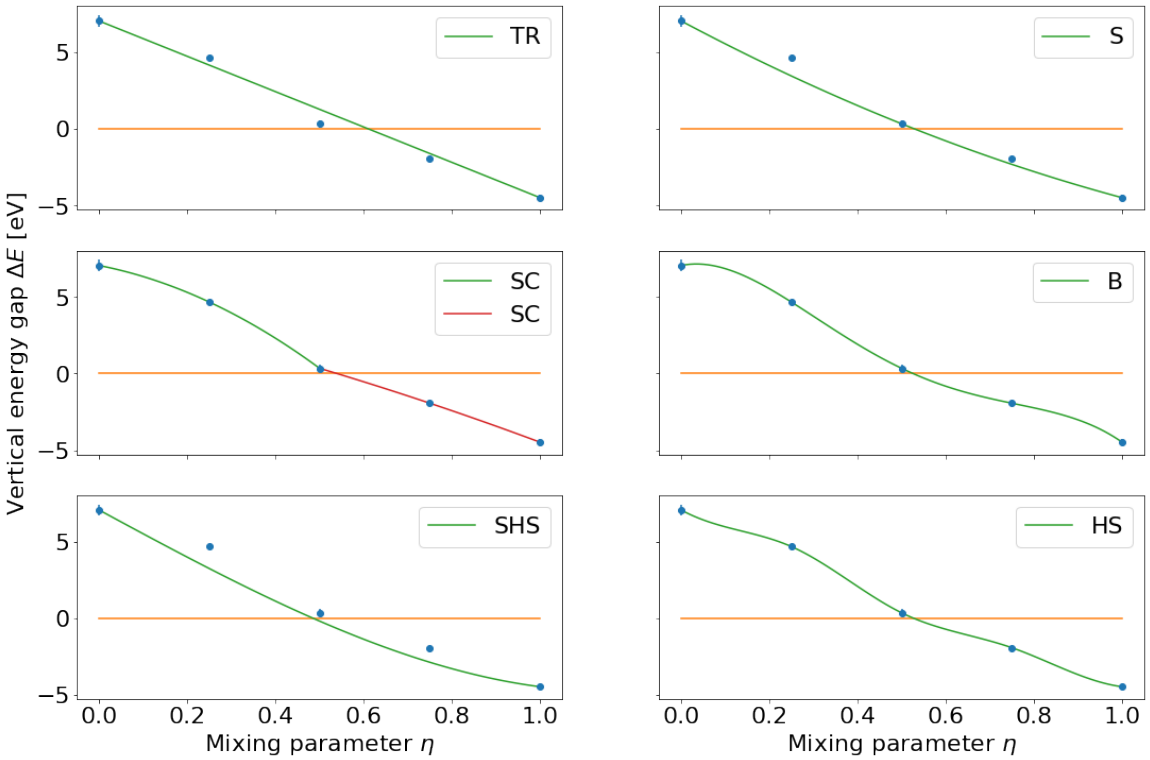
Calculation(s) 952

Name		ΔA [eV]	pK_a
Trapezoid rule	I^{TR}	1.289 ± 0.206	20.0 ± 3.2
Simpson's rule	I^{S}	0.659 ± 0.174	9.4 ± 2.5
Composite Simpson's rule	I^{SC}	1.184 ± 0.076	18.2 ± 1.2
Boole's rule	I^{B}	1.219 ± 0.074	18.8 ± 1.2
Single Hermite spline	I^{SHS}	0.300 ± 0.206	3.4 ± 2.3
Hermite spline	I^{HS}	1.031 ± 0.086	15.6 ± 1.3

▲ **Table 5.4** – Total free energy difference and pK_a values for the surface OH of the hematite c-cut as obtained by different integration methods and the resulting pK_a value including all corrections.

Calculation(s) 4c5, 9cf

Insertion of these numbers into the various integration formulas outlined in the beginning of this chapter yields the integral values in Table 5.4. Two values stand out: the free energy difference calculated using Simpson's rule and for the single Hermite spline is significantly lower (and unphysical) than the other values. The reason here is that the curvature at the endpoints and the curvature in the middle of the integration domain do not give a consistent result. This could only be improved by longer trajectories for the initial and final state. The agreement of the value for the Trapezoid rule with the other estimates with more data points is fortuitous, as can be seen by the large statistical uncertainty. All values with the exception of Simpson's rule and the single Hermite spline are statistically compatible with each other, with three of them having a small statistical uncertainty: the

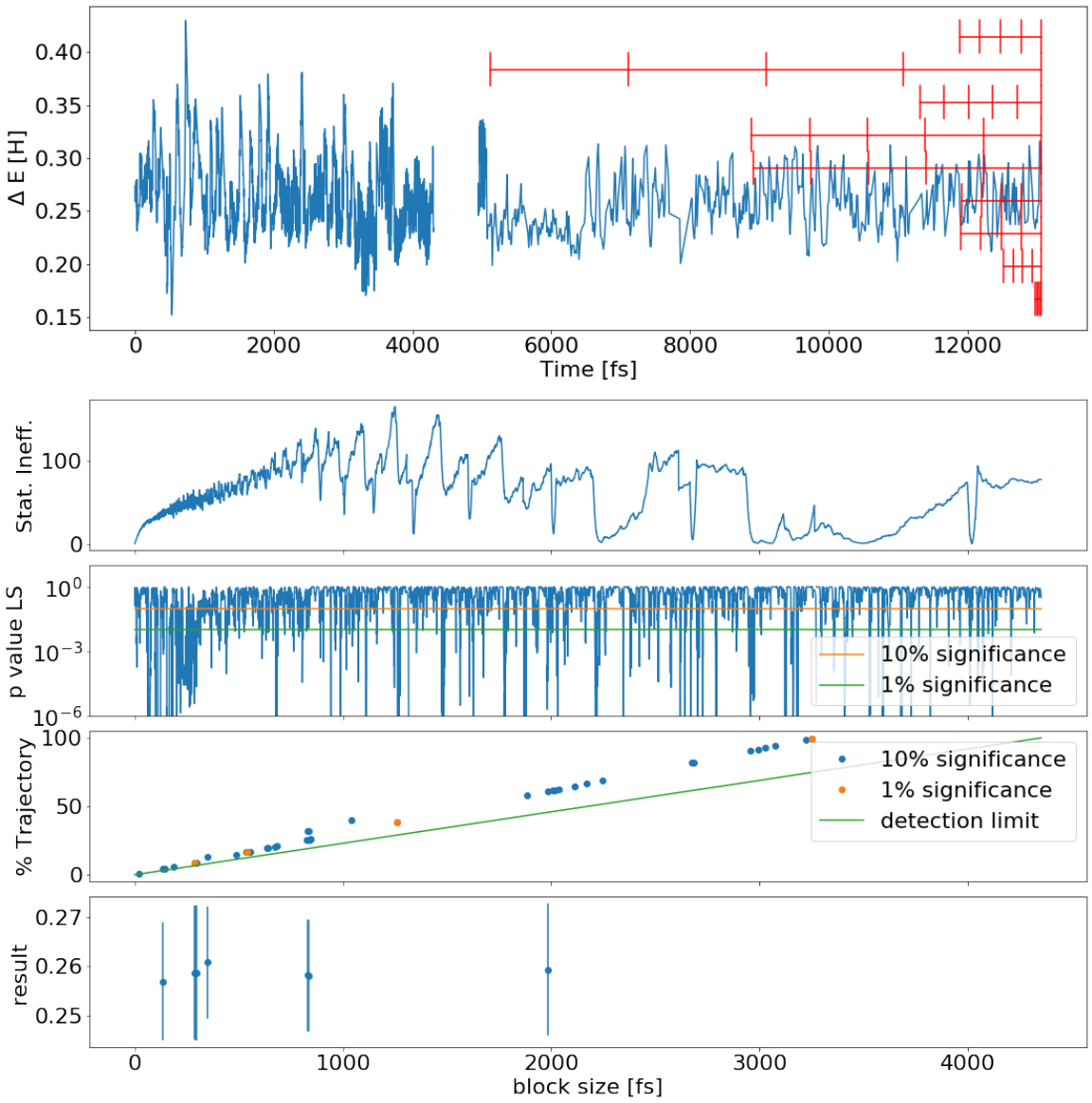


▲ **Figure 5.3** – Comparison of the raw data points (including error bars) and the integrand shapes as created to obtain values from Table 5.4. For the Simpson's composite rule, both segments are differently coloured. Labels denote Trapezoid rule, Simpson's rule, Composite Simpson's rule, Boole's rule, single Hermite spline and Hermite splines.

Calculation(s) 4c5, 9cf

composite Simpson's rule, Boole's rule and Hermite splines. Of these, it is interesting that the inclusion of the derivatives for the Hermite splines does not significantly increase the uncertainty despite the fact that the derivatives as obtained from the second moment of the vertical energy gap distribution are more prone to slow convergence. Since this offers additional physical insight into the system, the inclusion of the derivatives is most reasonable, but only once the trajectories are sufficiently long to capture converged variances of the vertical energy gap. Before this point is reached, Boole's rule is the best option. This means that the pK_a value determined in this work for the c-cut is 18.8 ± 1.2.

However, all three of these methods with a small statistical error require the five data points, i.e. three intermediate calculations. Since these intermediate ones are the most expensive ones, it is desirable to identify a method that works well when just given information about initial and final state MD. From the two options, the trapezoid rule and the single Hermite spline, the latter is expected to be more robust for non-linear integrands, since at least the derivative at the endpoints is correct. For the single Hermite spline which



▲ **Figure 5.4** – Example of the determination of the equilibrated region of the original E_0 time series (top panel, ΔE). The panels below show the statistical inefficiency Φ, the p value of having local stationarity in the statistical inefficiency, the percentage of the trajectory that can be considered to be equilibrated based on the Shapiro-Wilk test. For each block size τ where both the statistical inefficiency is locally stationary and the Shapiro-Wilk test suggests Gaussian distributed block averages, the bottom panel shows the resulting estimated mean average and uncertainty thereof. All block sizes and the corresponding equilibrated regions are the red segments shown in the top panel. The section of the trajectory covered by the red bars is considered to be equilibrated. For a detailed discussion of the method, please refer to the beginning of the chapter.

Calculation(s) 4c5, 9cf

with the current data gives a particularly low estimate, this is due to the steep derivative of the integrand at the endpoints. Since this derivative is obtained from the variance of the time series, which only converges to the true underlying physical variance for long trajectories, the single Hermite spline can only be expected to be accurate for longer trajectories.

For the method evaluation, Figure 5.4 shows the region in the original data series that has automatically been found to be equilibrated. While there are several block sizes τ that fulfilled all criteria for convergence, their estimate is consistent with each other, as expected for an equilibrated trajectory. The graph shows that the condition of local stationarity for the statistical inefficiency Φ is often met over the course of the trajectory, even at a 1 % significance level. The criterion of Gaussian distribution for the block averages seems to be met less often, but this is likely to depend on trajectory length. The picture for the other four trajectories analysed this way is very much similar, which means that this method may be a helpful tool when reliably calculating thermodynamic integrals. To fully make it a black box with very little human interaction, the estimation method should be extended such that the input data may be autocorrelated. Currently, this is ensured by spacing the input data such that consecutive data points are not correlated. With this extension, the raw time series from the AIMD trajectory can be directly fed into this mechanism.

With this method tested on the more stable c-cut, the next step would be to perform similar calculations for the more labile and reactive r-cut, where the question of the surface protonation pK_a is still open. This is particularly interesting, because different empirical models based on bond-valence sum considerations yield highly incompatible predictions.[260]

5.6 Conclusion

In this chapter, the thermodynamic integration of the vertical energy gap has been used for quantifying the pK_a value of the single proton of the c-cut (001) surface. This is a good test system, because the surface structure is stable on the picosecond time scale in terms of the oxygen occupancy and the protonation state, as shown in previous chapters. For the r-cut however, to achieve the same result will be much harder, since the surface aquo groups frequently interchange with bulk water and since the protonation state is much more labile. Since there are three distinct sites of interest for the r-cut, the methodology outlined in this chapter as to how to determine whether the run has been equilibrated and how to minimise the sampling uncertainty, will help to reduce the computational cost. Those results could help evaluating the competing models for estimation of the pK_a which currently give incompatible results.

6 Conclusion

In this work, three major topics have been discussed: the hematite / liquid water interface, accelerating the calculation of Hartree-Fock exchange forces, and efficient thermodynamic integration.

For hematite, common approaches in literature (GGA+U) and hybrid functionals have been compared for both bulk and interface structures. A reliable method to obtain the interface structure has been presented and the finite size effects thereof have been estimated. Based on the resulting AIMD data, the structure of the interface has been obtained for different crystal cuts and terminations, as observed in experiment. The dynamic patterns arising from the dynamics of the surface protonation have been described and in some cases (e.g. for missing chained hydrogen bonds) explained with joint geometric and electronic arguments. The equilibration process of the surface protonation has been described for the more reactive crystal cut under investigation. The obtained structural information was linked to experimental information about the solvation dynamics and hydrogen bond lifetimes. With the solvation structure at the interface, proton movement like the hydronium assisted transport have been shown and discussed. With the detailed information about interconversion of doubly protonated surface groups to adsorbed or even bulk water, experimental results and simulations could be brought into excellent agreement. Finally, the localisation of excess charges at the interface could be described and the structure thereof has been compared to the charge neutral interface.

As for the Hartree-Fock exchange forces, an in-depth analysis of the individual contributions arising from single four-centre integrals has been presented. Based on this information, a scheme for accelerating the force evaluation has been proposed, implemented in CP2K and evaluated on a variety of systems. This scheme makes a MD calculation at hybrid functional level about three times faster, while keeping the same accuracy in forces and a comparable energy drift.

Regarding the thermodynamic integration, a method to automatically detect the equilibrated region of the time series has been suggested and tested with the pK_a calculation of one hematite/water interface. For this interface, several integration formulas have been discussed. This method allows for Boltzmann reweighting (if necessary) and provides reliable statistical uncertainties for the final physical values. This enables more reproducible

and efficient simulations while reducing the impact of potential human bias.

The work outlined above so far has been published during my PhD.[206,238,260,300]

Outlook

In all three major topics of this work, several applications and research questions are of future interest.

For the hematite/liquid water interface, the charge transfer at the surface and across the surface to solvent or adsorbed molecules is discussed in the literature without a common picture. High-level simulations like the ones carried out in this work could contribute to elucidating this aspect. Potential methods here include charge-constrained DFT, which recently has seen a new implementation in CP2K. Also, a more accurate estimate of the pK_a of the r-cut interface could help settle a dispute between competing experimental approaches and could help understanding protonation dynamics at the interface.

While the forces have been screened in the acceleration of Hartree-Fock exchange calculations, it would be desirable to apply the same mechanism to the energy evaluation, such that there is a consistent picture of the potential energy surface the simulation is following and the forces that act on the atoms. This would likely give another speed-up and reduce the energy drift. Furthermore, better extrapolation schemes may be helpful and could be tested together with multiple time step methods to improve sampling at a comparably high level of theory.

The guided thermodynamic integration suggested in this work would benefit from support for autocorrelated data. In this case, the overall system would work as a black box in terms of efficiently bringing down the statistical uncertainty of thermodynamic integrals. Moreover, it would be interesting to benchmark this method for more systems and different properties accessible by thermodynamic integration, not only the pK_a.

Acknowledgements

I gratefully acknowledge a PhD studentship co-sponsored by University College London and Pacific Northwest National Laboratory (PNNL) through its BES Geosciences program supported by the U.S. Department of Energy's Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences and Biosciences Division. MD simulations with-

out screening were carried out on ARCHER, the UK national HPC facility (Edinburgh), to which access was granted via the ARCHER Leadership pilot call and the Materials Chemistry Consortium (EPSRC grant EP/L000202). Data analysis and calculations with screening were carried out with computing resources provided through two Microsoft Azure Sponsorships as well as AWS Cloud Credits for Research and supported by software made available by Tableau Inc.

I thank my supervisors, Jochen Blumberger (UCL), Alex Shluger (UCL), and Kevin Rosso (PNNL) for their support and time. I also would like to thank the following people for discussing aspects of this work with me: Ellen Backus (MPI Mainz), Iain Bethune (EPCC), Eric Bylaska (PNNL), Michele Ceriotti (EPFL), Oliver Gittus (UCL), Jürg Hutter (UZH), Rasmus Jakobsen (UCL), Sebastien Kerisit (PNNL), Sanliang Ling (UCL), Todd Martinez (Stanford), Martin McBriarty (PNNL), Angelos Michaelides (UCL), Daniel Sebastiani (MLU Halle), Sameer Shende (UOregon), Kamelia von Rudorff (FU Berlin), Matt Watkins (Lincoln), Brian Wylie (VI-HPS), past and present members of the Blumberger Research Group (Antoine Carof, Peter Cooke, Zdenek Futera, Fruzsina Gajdos, Samuele Giannini, Jacob Spencer, Benjamin Rosseau, Laura Scalfi, Xiuyun Jiang, Hui Yang) at UCL, and the support staff at ARCHER.

Finally, I thank all contributors to the open source software this work relied on and without which many ideas would not been feasible to explore: bokeh, biopython, cp2k, git, jupyter, matplotlib, MDAnalysis, mdtraj, MongoDB, the numpy ecosystem, openbabel, pandas, phabricator, python, scipy, statsmodels, and VESTA.

Appendix A

List of Abbreviations

ACF	Auto-correlation Function
ADF	Augmented Dickey-Fuller
ADMM	Auxiliary Density Matrix Method[134]
AIMD	ab-initio molecular dynamics
AM	Air Mass
ARIMA	Autoregressive Integrated Moving Average
ASCII	American Standard Code for Information Interchange
ASPC	Always Stable Predictor-Corrector
AWS	Amazon Web Services
BES	Basic Energy Sciences
BJ	Becke-Johnson Damping[143]
BLYP	Becke-Lee-Yang-Parr Correlation Functional[331]
BOMD	Born-Oppenheimer Molecular Dynamics
CC BY	Creative Commons By-Attribution License
CDFT	Constrained Density Functional Theory
COM	Center of Mass
CP2K	Software Package for ab inito Molecular Dynamics
CPMD	Car-Parinello Molecular Dynamics / Software Package for ab inito Molecular Dynamics
CPU	Central Processing Unit
CSVR	Canonical Sampling Through Velocity Rescaling[332]
CTR	Crystal Truncation Rod[213]
D3	Grimme Dispersion Correction[146]
DFT	Density Functional Theory
DFTMD	Molecular Dynamics based on Density Functional Theory methods
DoE	Department of Energy
DOS	Density of States
DZVP	Double Zeta Valence Polarized
EPSRC	Engineering and Physical Sciences Research Council
ERI	Electron Repulsion Integral
EWMA	Exponential Window Moving Average
GGA	Generalised Gradient Approximation

GGA+U | Generalised Gradient Approximation with Hubbard-U
GTH | Goedecker-Teter-Hutter[333]
HAB11 | Electronic Coupling Database[334]
HB | Hydrogen Bond
hcp | hexagonal close packing
HF | Hartree-Fock
HFX | Hartree-Fock Exchange
HK | Hohenberg-Kohn
HOMO | Highest Occupied Molecular Orbital
HPC | High-performance Computing
HSE03 | Heyd-Scuseria-Ernzerhof[127]
HSE06 | Heyd-Scuseria-Ernzerhof[124]
HSE | Heyd-Scuseria-Ernzerhof[127]
IASD | Integrated Absolute Spin Density
IPC | Interprocess Communication
kAU | ARCHER Budget Unit
KDE | Kernel Density Estimation
KS | Kohn-Sham
LJ | Lennard-Jones potential[335]
LMCT | Ligand-Metal Charge Transfer
LR | Long Range
LSDA | Local Spin Density Approximation
LUMO | Lowest Unoccupied Molecular Orbital
MCC | Materials Chemistry Consortium
MC | Monte Carlo
MD | Molecular Dynamics
MIC | Minimum Image Convention
MLE | Maximum Likelihood Estimator
MLLE | Maximum Log-Likelihood Estimator
MO | Molecular Orbital
MP2 | Second Order Møller–Plesset Perturbation Theory[336]
MPI | Message Passing Interface
MSD | Mean-squared Deviation
MT | Multi-Threading
NPTF | Ensemble with constant number of particles, barostat and thermostat with a flexible unit cell
NVE | Ensemble with constant number of particles, constant volume and constant energy
NVT | Ensemble with constant number of particles, constant volume and thermostat
OLS | Ordinary Least Squares Estimator
PBC | Periodic Boundary Conditions
PBE0 | Parameter-Free Perdew-Burke-Ernzerhof[130]

PBE | Perdew-Burke-Ernzerhof[128]
PDOS | Projected Density of States
PES | Potential Energy Surface
PNNL | Pacific-Northwest National Lab
PW | Plane Wave
PZC | Point of Zero Charge
RCA | Reverse Cumulative Average
RDF | Radial Distribution Function
RMSD | Root Mean Squared Deviation
RMSE | Root Mean Square Error
RMS | Root Mean Squared
SCF | self-consistent field
SIE | self-interaction error
SMP | Symmetric Multiprocessing
SPC | Single Point Calculation
SQL | Structured Query Language
SR | Short Range
SVD | Singular Value Decomposition
SW | Scattered Wave
TD-DFT | Time-Dependent Density Functional Theory
TIP3P | Three-Site Transferrable Intermolecular Potential[199]
TI | Thermodynamic Integration
TMO | Transition Metal Oxide
UC | Unit Cell
+U | including Hubbard-U
UK | United Kingdom
US SIC | Spin Density Self-Interaction Correction Method[337]
vdW | van der Waals
XPS | X-ray Photoelectron Spectroscopy

List of Calculations

045 | OH radical in 31 water molecules with code A
0c9 | Hematite r-cut $2 \times 2 \times 1$ interface with 1-1-1 protonation scheme with code B
0da | Hematite $2 \times 2 \times 1$ iron terminated c-cut with code B
0db | Hematite r-cut $2 \times 2 \times 1$ setup with code B
0e8 | Neutral c-cut interface with 25% HFX with code C
10d | Hematite $2 \times 2 \times 1$ iron terminated c-cut with code B
129 | $2 \times 2 \times 1$ interface 30Å water layer with code D
140 | $2 \times 2 \times 1$ interface 30Å water layer with code E
19b | Positive c-cut interface with 50% HFX with code C
1b4 | Energy conservation testing with code F
227 | Positive c-cut interface with 12% HFX with code C

22c | $2 \times 2 \times 1$ interface with 30Å water layer with code E
253 | $4 \times 2 \times 1$ hematite cell with water layer performance benchmark with code E
257 | CDFT Ethylene dimer with periodic box with code G
26a | Hematite r-cut $2 \times 2 \times 1$ interface with 1-1-1 protonation scheme with code B
283 | Box with 32 water molecules with code C
29a | Hematite r-cut $2 \times 2 \times 1$ setup with code B
2a2 | Positive c-cut interface with 25% HFX with code C
2b2 | Hematite r-cut $2 \times 2 \times 1$ setup with code B
2ba | Water dimer and hematite bulk systems with code A
2cc | Hematite $2 \times 2 \times 1$ r-cut equilibration with code B
318 | Hematite r-cut $2 \times 2 \times 1$ setup with code B
36e | $2 \times 2 \times 1$ hematite cell with code D
3dd | 30 atom hematite geometry optimization with code H
3f1 | Positive c-cut interface with 12% HFX with code C
42e | 270 atom hematite PDOS calculation with code I
446 | Negative c-cut interface with 12% HFX with code C
448 | 30 atom hematite cell with 96 water molecules geometry optimisation with code E
4a4 | CoO unit cells of variable size with code A
4c4 | Hematite $2 \times 2 \times 1$ r-cut equilibration with code B
4c5 | Hematite c-cut pKa for E1 with code G
4c8 | Water box using HSE with code C
4d9 | Maximum-Likelihood estimator and pattern generator (own code)
520 | 120 atom hematite cell with 384 water molecules performance benchmark with code E
523 | Hematite $2 \times 2 \times 1$ iron terminated c-cut with code B
55e | Hematite c-cut $3 \times 3 \times 1^*$ interface with code J
568 | Hematite r-cut $2 \times 2 \times 1$ setup with code B
590 | $2 \times 2 \times 1$ interface 30Å water layer with code D
5d8 | Negative c-cut interface with 12% HFX with code C
5eb | $1 \times 1 \times 1$ hematite cell with code C
633 | Hematite r-cut $2 \times 2 \times 1$ setup with code B
634 | Hematite $2 \times 2 \times 1$ c-cut interface with code K
65b | HSE water box without dispersion correction with code C
689 | 30 atom hematite DOS calculation with code L
692 | Hematite bulk $3 \times 3 \times 1$ setup with code K
696 | Hematite c-cut $3 \times 3 \times 1^*$ interface with code J
6a0 | Hematite $2 \times 2 \times 1$ iron terminated c-cut with code B
6cf | Hematite $2 \times 2 \times 1$ iron terminated c-cut with code B
6f2 | $2 \times 2 \times 1$ interface with code M
6fd | Hematite r-cut $2 \times 2 \times 1$ setup with code B
70a | 32 molecule water box and OH radical with code G
782 | Hydrogen bond analysis (own code, contributions Rasmus Jakobsen)
7a3 | Positive c-cut interface with 50% HFX with code C
7a7 | $1 \times 1 \times 1$ hematite cell with code C

7ab | 30 atom hematite single point calculation with code N
7b0 | Interface equilibration with code C
7ce | Water dimer regression test with code G
82e | Classical solvation sampling for 1-1-1 protonation with code G
87b | Hematite $2 \times 2 \times 1$ iron terminated c-cut with code B
893 | 270 atom hematite cell optimization with code O
89a | Hematite r-cut $2 \times 2 \times 1$ interface with 1-1-1 protonation scheme with code B
952 | pka Frequencies with code G
9be | Cobalt oxide scaling scan with code F
9cf | Hematite c-cut pKa for E0 with code G
9d1 | Helium dimer, positively charged with code P
9fb | Negative c-cut interface with 25% HFX with code C
a03 | Test case for pKa setups with code G
a2d | Water dimer with code Q
a63 | Hematite $2 \times 2 \times 1$ r-cut equilibration with code B
ab3 | Hematite $2 \times 2 \times 1$ iron terminated c-cut with code B
acc | $2 \times 2 \times 1$ interface 30Å water layer with code D
af9 | 120 atom hematite PDOS calculation with code R
b0f | Scaling of water box with code C
b34 | Cobalt oxide HFX force comparison with code S
b42 | 120 atom hematite geometry optimization and PDOS calculation with code T
b7b | Hematite bulk $4 \times 4 \times 1$ setup with code K
b82 | Negative c-cut interface with 50% HFX with code C
c92 | Water box using HSE with code C
cd |
d0e | Hematite c-cut $3 \times 3 \times 1^*$ interface with code J
d0f | CoO single unit cell with code U
d41 | CDFT Ethylene dimer with code G
d9b | 270 atom hematite geometry optimization and PDOS calculation with code V
dc9 | Hematite c-cut $3 \times 3 \times 1^*$ interface with code J
e19 | Hematite c-cut $2 \times 2 \times 1$ interface with code W
e47 | Hematite $2 \times 2 \times 1$ iron terminated c-cut with code B
e49 | $2 \times 2 \times 1$ interface with 30Å water layer with code E
e81 | Hematite/water interface single point with code X
e9a | Hematite c-cut $2 \times 2 \times 1$ setup with code B
ea6 | Classical water box with code C
ea7 | Neutral c-cut interface with 50% HFX with code C
eb1 | $2 \times 2 \times 1$ interface 30Å water layer with code D
ec8 | Hematite r-cut $2 \times 2 \times 1$ setup with code B
ee5 | Hematite r-cut $2 \times 2 \times 1$ interface with 1-1-1 protonation scheme with code B
f23 | $2 \times 2 \times 1$ hematite cell with code D
f8d | $2 \times 2 \times 1$ interface 30Å water layer with code D
fa9 | 270 atom hematite geometry optimization with code T

fd5	$4 \times 4 \times 2$ hematite cell with water layer performance benchmark with code E
fef	OH radical in 31 water molecules with code A
ffe	Hematite $2 \times 2 \times 1$ r-cut equilibration with code B
code A	CP2K development versions (4f54120,f4b564a)[126,128,134,135,137,146,211,302,316,318,319,333,338–343]
code B	CP2K 3.0[126,128,134,135,137,146,211,316,318,319,333,338–343]
code C	CP2K 2.7[107,108,128,135,211,302,316,333,338–343]
code D	CP2K 2.7[126,128,134,135,137,146,211,316,318,319,333,338–343]
code E	CP2K 2.6 (rev. 14482)[126,128,134,135,137,146,211,316,318,319,333,338–343]
code F	CP2K development version 27793ce[126,128,134,135,137,146,211,302,316,318,319,333,338–343]
code G	CP2K 4.0[126,128,134,135,137,146,211,302,316,318,319,333,338–343]
code H	CP2K 2.3 (rev. 12205)[128,135,302,316,333,339–343]
code I	CP2K 2.5 (rev. 13632)[128,135,211,302,316,333,339–343]
code J	CP2K 3.0[126,128,134,135,137,146,211,316,318,319,333,338–343]
code K	CP2K 4.1[126,128,134,135,137,146,211,302,316,318,319,333,338–344]
code L	CP2K 2.6 (rev. 14482)[128,135,211,302,316,333,339–343]
code M	CP2K 2.7[107,108,128,135,137,146,211,316,333,338–343]
code N	CP2K 2.3 (rev. 12205)[128,135,316,333,339–343]
code O	CP2K 2.6 (rev. 14482)[126,128,134,135,137,146,211,302,316,318,319,333,338–343]
code P	CP2K 2.7[128,211,316,333,338–343]
code Q	CP2K development versions (f4b564a, 8a77588, 90974b1)[126,128,134,135,137,146,211,302,316,318,319,333,338–343]
code R	CP2K 2.5 (rev. 13632)[128,135,211,302,316,333,339–342,342,343]
code S	CP2K development version f4b564a[126,128,134,135,137,146,211,302,316,318,319,333,338–343]
code T	CP2K 2.6 (rev. 14482)[126,128,134,135,211,302,316,318,319,333,338–343]
code U	CP2K development versions (f4b564a, 4f54120, b2b6330)[126,128,134,135,137,146,211,302,316,318,319,333,338–343]
code V	CP2K 2.6 (rev. 14482)[126,128,134,135,211,302,316,318,319,333,338–343]
code W	CP2K 4.1[126,128,134,135,137,146,211,302,316,318,319,332,333,338–344]
code X	a development version of CP2K[126,128,134,135,137,146,211,302,316,318,319,333,338–343]

Literature

1 | Hamann, T. W. Splitting water with rust: hematite photoelectrochemistry. *Dalton Transactions*, 41(26), **2012**:7830.

2 | Chatman, S., Zarzycki, P., and Rosso, K. M. Surface potentials of (001), (012), (113) hematite (α-Fe_2O_3) crystal faces in aqueous solution. *Physical Chemistry Chemical Physics*, 15(33), **2013**:13,911.

3 | Freyria, F. S., et al. Hematite Nanoparticles Larger than 90 nm Show No Sign of Toxicity in Terms of Lactate Dehydrogenase Release, Nitric Oxide Generation, Apoptosis, and Comet Assay in Murine Alveolar Macrophages and Human Lung Epithelial Cells. *Chemical Research in Toxicology*, 25(4), **2012**:850–861.

4 | Lin, Y., et al. Hematite-based solar water splitting: challenges and opportunities. *Energy and Environmental Science*, 4(12), **2011**:4862.

5 | Gilbert, B., et al. Band-gap measurements of bulk and nanoscale hematite by soft x-ray spectroscopy. *Physical Review B*, 79(3), **2009**:035,108.

6 | Yanina, S. V. and Rosso, K. M. Linked Reactivity at Mineral-Water Interfaces Through Bulk Crystal Conduction. *Science*, 320(5873), **2008**:218–222.

7 | Kämpf, N. and Schwertmann, U. Goethite and hematite in a climosequence in southern Brazil and their application in classification of kaolinitic soils. *Geoderma*, 29(1), **1983**:27–39.

8 | Philippe, A. and Schaumann, G. E. Interactions of Dissolved Organic Matter with Natural and Engineered Inorganic Colloids: A Review. *Environmental Science & Technology*, 48(16), **2014**:8946–8962.

9 | Mylon, S. E., Chen, K. L., and Elimelech, M. Influence of Natural Organic Matter and Ionic Composition on the Kinetics and Structure of Hematite Colloid Aggregation: Implications to Iron Depletion in Estuaries. *Langmuir*, 20(21), **2004**:9000–9006.

10 | Ravishankara, A. R., Rudich, Y., and Wuebbles, D. J. Physical Chemistry of Climate Metrics. *Chemical Reviews*, 115(10), **2015**:3682–3703.

11 | Shi, L., et al. Respiration of metal (hydr)oxides byShewanellaand-Geobacter: a key role for multihaemc-type cytochromes. *Molecular Microbiology*, 65(1), **2007**:12–20.

12 | Breuer, M., et al. Thermodynamics of Electron Flow in the Bacterial Decaheme Cytochrome MtrF. *Journal of the American Chemical Society*, 134(24), **2012**:9868–9871.

13 | Breuer, M., et al. Multi-haem cytochromes in Shewanella oneidensis MR-1: structures, functions and opportunities. *Journal of The Royal Society Interface*, 12(102), **2014**:20141,117–20141,117.

14 | Blumberger, J. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chemical Reviews*, 115(20), **2015**:11,191–11,238.

15 | Reddy, M. V., Subba Rao, G. V., and Chowdari, B. V. R. Metal Oxides and Oxysalts as Anode Materials for Li Ion Batteries. *Chemical Reviews*, 113(7), **2013**:5364–5457.

16 | Sivula, K., Le Formal, F., and Grätzel, M. Solar Water Splitting: Progress Using Hematite (α-Fe_2O_3) Photoelectrodes. *ChemSusChem*, 4(4), **2011**:432–449.

17 | Lützenkirchen, J., et al. Structure-charge relationship - the case of hematite (001). *Faraday Discuss.*, 180, **2015**:55–79.

18 | Kuhlenbeck, H., Shaikhutdinov, S., and Freund, H.-J. Well-Ordered Transition Metal Oxide Layers in Model Catalysis – A Series of Case Studies. *Chemical Reviews*, 113(6), **2013**:3986–4034.

19 | Herrera, F., et al. Catalytic combustion of Orange II on hematite. *Applied Catalysis B: Environmental*, 29(2), **2001**:147–162.

20 | Jung, J.-E., et al. First-Principles Investigation of Mercury Adsorption on the α-Fe2O3(1-102) Surface. *The Journal of Physical Chemistry C*, 119(47), **2015**:26,512–26,518.

21 | Dallongeville, S., et al. Proteins in Art, Archaeology, and Paleontology: From Detection to Identification. *Chemical Reviews*, 116(1), **2016**:2–79.

22 | Cheng, J. and Sprik, M. Acidity of the Aqueous Rutile TiO2(110) Surface from Density Functional Theory Based Molecular Dynamics. *Journal of Chemical Theory and Computation*, 6(3), **2010**:880–889.

23 | Cheng, J. and Sprik, M. Acidity of the Aqueous Rutile TiO2(110) Surface from Density Functional Theory Based Molecular Dynamics. *Journal of Chemical Theory and Computation*, 6(3), **2010**:880–889.

24 | Nguyen, M.-T., et al. Photo-Oxidation of Water on Defective Hematite(0001). *ACS Catalysis*, 5(2), **2015**:715–721.

25 | Liu, L.-M., et al. Structure and dynamics of liquid water on rutile $TiO_{2}(110)$. *Physical Review B*, 82(16), **2010**:161,415.

26 | Zarzycki, P., Smith, D. M., and Rosso, K. M. Proton Dynamics on Goethite Nanoparticles and Coupling to Electron Transport. *Journal of Chemical Theory and Computation*, **2015**:150317124118,003.

27 | Kerisit, S., Liu, C., and Ilton, E. S. Molecular dynamics simulations of the orthoclase (001)- and (010)-water interfaces. *Geochimica et Cosmochimica Acta*, 72(6), **2008**:1481–1497.

28 | Gaigeot, M.-P., Sprik, M., and Sulpizi, M. Oxide/water interfaces: how the surface chemistry modifies interfacial water properties. *Journal of Physics: Condensed Matter*, 24(12), **2012**:124,106.

29 | Tanwar, K. S., et al. Surface diffraction study of the hydrated hematite surface. *Surface Science*, 601(2), **2007**:460–474.

30 | Catalano, J. G. Weak interfacial water ordering on isostructural hematite and corundum (001) surfaces. *Geochimica et Cosmochimica Acta*, 75(8), **2011**:2062–2071.

31 | Trainor, T. P., et al. Structure and reactivity of the hydrated hematite (0001) surface. *Surface Science*, 573(2), **2004**:204–224.

32 | Kerisit, S. Water structure at hematite-water interfaces. *Geochimica et Cosmochimica Acta*, 75(8), **2011**:2043–2061.

33 | Kerisit, S. and Rosso, K. M. Computer simulation of electron transfer at hematite surfaces. *Geochimica et Cosmochimica Acta*, 70(8), **2006**:1888–1903.

34 | Jeon, B., et al. Nanoscale oxidation and complex oxide growth on single crystal iron surfaces and external electric field effects. *Phys. Chem. Chem. Phys.*, 15(6), **2013**:1821–1830.

35 | Kerisit, S. and Rosso, K. M. Kinetic Monte Carlo model of charge transport in hematite (α-Fe2O3). *The Journal of Chemical Physics*, 127(12), **2007**:124,706.

36 | Lo, C. S., et al. Density functional theory study of the clean and hydrated hematite(11¯02)surfaces. *Physical Review B*, 75(7), **2007**:075,425.

37 | Ovcharenko, R., Voloshina, E., and Sauer, J. Water adsorption and O-defect formation on Fe2O3(0001) surfaces. *Phys. Chem. Chem. Phys.*, 18(36), **2016**:25,560–25,568.

38 | English, N. J., et al. Photo-active and dynamical properties of hematite (Fe2O3)–water interfaces: an experimental and theoretical study. *Physical Chemistry Chemical Physics*, 16(28), **2014**:14,445.

39 | Zhang, X. and Bieberle-Hütter, A. Modeling and Simulations in Photoelectrochemical Water Oxidation: From Single Level to Multiscale Modeling. *ChemSusChem*, 9(11), **2016**:1223–1242.

40 | Nag, S., Banerjee, K., and Datta, D. Estimation of the van der Waals radii of the d-block elements using the concept of bond valence. *New Journal of Chemistry*, 31(6), **2007**:832.

41 | Pauling, L. THE PRINCIPLES DETERMINING THE STRUCTURE OF COMPLEX IONIC CRYSTALS. *Journal of the American Chemical Society*, 51(4), **1929**:1010–1026.

42 | Lin, S. Magnetic Properties of Hematite Single Crystals. I. Magnetization Isotherms, Antiferromagnetic Susceptibility, and Weak Ferromagnetism of a Natural Crystal. *Physical Review*, 116(6), **1959**:1447–1452.

43 | Sato, Y. and Akimoto, S.-i. Hydrostatic compression of four corundum-type compounds: α-Al_2O_3, V_2O_3, Cr_2O_3, and α-Fe_2O_3. *Journal of Applied Physics*, 50(8), **1979**:5285.

44 | Ovsyannikov, S. V., et al. High-pressure cycling of hematite α-Fe_2O_3: Nanostructuring, in situ electronic transport, and possible charge disproportionation. *Physical Review B*, 86(20), **2012**:205,131.

45 | Pasternak, M., et al. Breakdown of the Mott-Hubbard State in Fe_2O_3: A First-Order Insulator-Metal Transition with Collapse of Magnetism at 50 GPa. *Physical Review Letters*, 82(23), **1999**:4663–4666.

46 | Rodríguez-Fernández, A., et al. Chiral properties of hematite α-Fe_2O_3 inferred from resonant Bragg diffraction using circularly polarized x rays. *Physical Review B*, 88(9), **2013**:094,437.

47 | Srikrishna Ramya, S. and Mahadevan, C. Preparation and structural, optical, magnetic, and electrical characterization of Mn2+/Co2+/Cu2+ doped hematite nanocrystals. *Journal of Solid State Chemistry*, 211, **2014**:37–50.

48 | Chen, Y. H. and Lin, C. C. Effect of nano-hematite morphology on photocatalytic activity. *Physics and Chemistry of Minerals*, 41(10), **2014**:727–736.

49 | Kumari, S., et al. Spray pyrolytically deposited nanoporous Ti4+ doped hematite thin films for efficient photoelectrochemical splitting of water. *International Journal of Hydrogen Energy*, 35(9), **2010**:3985–3990.

50 | Zhang, Q. and Dunn, S. Room Temperature Synthesis of Crystalline α-Fe_2O_3 Nanoparticles. *Journal of Nanoscience and Nanotechnology*, 11(4), **2011**:3716–3718.

51 | Wang, H. and Turner, J. A. Photoelectrochemistry of Hematite Thin Films. *ECS Transactions*, 25(42), **2010**:49–62.

52 | Chamberlin, S. E., et al. Optical absorption and spectral photoconductivity in α-$(Fe_{\{}1 - x\}Cr_x)_2O_3$solid-solution thin films. *Journal of Physics: Condensed Matter*, 25(39), **2013**:392,002.

53 | Kim, J. Y., et al. Single-crystalline, wormlike hematite photoanodes for efficient solar water splitting. *Scientific Reports*, 3, **2013**:2681.

54 | Sorenson, S., et al. Ultrafast Carrier Dynamics in Hematite Films: The Role of Photoexcited Electrons in the Transient Optical Response. *The Journal of Physical Chemistry C*, 118(41), **2014**:23,621–23,626.

55 | Wang, H. and Turner, J. A. Characterization of Hematite Thin Films for Photoelectrochemical Water Splitting in a Dual Photoelectrode Device. *Journal of The Electrochemical Society*, 157(11), **2010**:173–178.

56 | Miller, E. L., et al. Low-temperature reactively sputtered iron oxide for thin film devices. *Thin Solid Films*, 466(1-2), **2004**:307–313.

57 | Wang, J., White, W. B., and Adair, J. H. Optical Properties of Hydrothermally Synthesized Hematite Particulate Pigments. *Journal of the American Ceramic Society*, 88(12), **2005**:3449–3454.

58 | Sarkar, D., Mandal, M., and Mandal, K. Design and Synthesis of High Performance Multifunctional Ultrathin Hematite Nanoribbons. *ACS Applied Materials and Interfaces*, 5(22), **2013**:11,995–12,004.

59 | Misho, R. and Murad, W. Band gap measurements in thin films of hematite Fe2O3, pyrite FeS2 and troilite FeS prepared by chemical spray pyrolysis. *Solar Energy Materials and Solar Cells*, 27(4), **1992**:335–345.

60 | Engel, J. and Tuller, H. L. The electrical conductivity of thin film donor doped hematite: from insulator to semiconductor by defect modulation. *Physical Chemistry Chemical Physics*, 16(23), **2014**:11,374.

61 | Shinde, S. S., et al. Physical properties of hematite α-Fe_2O_3 thin films: application to photoelectrochemical solar cells. *Journal of Semiconductors*, 32(1), **2011**:013,001.

62 | Hardee, K. L. and Bard, A. J. Semiconductor Electrodes. *Journal of The Electrochemical Society*, 123(7), **1976**:1024.

63 | Li, N., et al. Effect of La-Doping on optical bandgap and photoelectrochemical performance of hematite nanostructures. *J. Mater. Chem. A*, 2(45), **2014**:19,290–19,297.

64 | Cardillo, D., et al. Highly porous hematite nanorods prepared via direct spray precipitation method. *Materials Letters*, 117, **2014**:279–282.

65 | Mallick, P. Influence of different materials on the microstructure and optical band gap of α-Fe2O3 nanoparticles. *Materials Science-Poland*, 32(2), **2014**:193–197.

66 | Glasscock, J., et al. Structural, optical and electrical properties of undoped polycrystalline hematite thin films produced using filtered arc deposition. *Thin Solid Films*, 516(8), **2008**:1716–1724.

67 | Chen, Y.-H. and Tu, K.-J. Thickness Dependent on Photocatalytic Activity of Hematite Thin Films. *International Journal of Photoenergy*, 2012, **2012**:1–6.

68 | Hamaguchi, C. Basic Semiconductor Physics. Springer, **2010**.

69 | Tauc, J. Optical properties and electronic structure of amorphous Ge and Si. *Materials Research Bulletin*, 3(1), **1968**:37–46.

70 | Van, T.-K., et al. Nanocystals of Hematite with Unconventional Shape-Truncated Hexagonal Bipyramid and Its Optical and Magnetic Properties. *Crystal Growth and Design*, 12(2), **2012**:862–868.

71 | Kuiper, P., et al. X-ray magnetic dichroism of antiferromagnet Fe_2O_3: The orientation of magnetic moments observed by Fe 2p x-ray absorption spectroscopy. *Physical Review Letters*, 70(10), **1993**:1549–1552.

72 | Rosso, K. M., Smith, D. M. A., and Dupuis, M. An ab initio model of electron transport in hematite (α-Fe_2O_3) basal planes. *The Journal of Chemical Physics*, 118(14), **2003**:6455.

73 | Wang, X.-G., et al. The Hematite α-Fe_2O_3) (0001) Surface: Evidence for Domains of Distinct Chemistry. *Physical Review Letters*, 81(5), **1998**:1038–1041.

74 | Lemire, C., et al. Ferryl (Fe=O) Termination of the Hematite $\alpha-Fe_2O_3(0001)$ Surface. *Physical Review Letters*, 94(16), **2005**:166,101.

75 | Barbier, A., et al. Surface termination of hematite at environmental oxygen pressures: Experimental surface phase diagram. *Physical Review B*, 75(23), **2007**:233,406.

76 | Chambers, S. and Yi, S. Fe termination for α-Fe_2O_3 (0001) as grown by oxygen-plasma-assisted molecular beam epitaxy. *Surface Science*, 439(1-3), **1999**:L785–L791.

77 | Thevuthasan, S., et al. Surface structure of MBE-grown α-Fe_2O_3 (0001) by intermediate-energy X-ray photoelectron diffraction. *Surface Science*, 425(2-3), **1999**:276–286.

78 | Rochester, C. H. and Topham, S. A. Infrared study of surface hydroxyl groups on haematite. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 75, **1979**:1073.

79 | Yamamoto, S., et al. Water Adsorption on α-Fe_2O_3 (0001) at near Ambient Conditions. *The Journal of Physical Chemistry C*, 114(5), **2010**:2256–2266.

80 | Nguyen, M.-T., Seriani, N., and Gebauer, R. Water adsorption and dissociation on α-$Fe_2O_3(0001)$: PBE+U calculations. *The Journal of Chemical Physics*, 138(19), **2013**:194,709.

81 | Rohrbach, A., Hafner, J., and Kresse, G. Ab initio study of the (0001) surfaces of hematite and chromia: Influence of strong electronic correlations. *Physical Review B*, 70(12), **2004**:125,426.

82 | Cheng, J. and Sprik, M. The electric double layer at a rutile TiO2water interface modelled using density functional theory based molecular dynamics simulation. *Journal of Physics: Condensed Matter*, 26(24), **2014**:244,108.

83 | Rustad, J. R., Wasserman, E., and Felmy, A. R. Molecular modeling of the surface charging of hematite. *Surface Science*, 424(1), **1999**:28–35.

84 | Eggleston, C. and Jordan, G. A new approach to pH of point of zero charge measurement: crystal-face specificity by scanning force microscopy (SFM). *Geochimica et Cosmochimica Acta*, 62(11), **1998**:1919–1923.

85 | Chang, H.-C., Healy, T. W., and Matijević, E. Interactions of metal hydrous oxides with chelating agents. *Journal of Colloid and Interface Science*, 92(2), **1983**:469–478.

86 | Honeyman, B. D. and Santschi, P. H. Coupling adsorption and particle aggregation: laboratory studies of "colloidal pumping" using iron-59-labeled hematite. *Environmental Science & Technology*, 25(10), **1991**:1739–1747.

87 | Zhang, J. and Buffle, J. Kinetics of Hematite Aggregation by Polyacrylic Acid: Importance of Charge Neutralization. *Journal of Colloid and Interface Science*, 174(2), **1995**:500–509.

88 | Martin, G. J., et al. Bulk and key surface structures of hematite, magnetite, and goethite: A density functional theory study. *American Mineralogist*, 94(10), **2009**:1341–1350.

89 | Rollmann, G., et al. First-principles calculation of the structure and magnetic phases of hematite. *Physical Review B*, 69(16), **2004**:165,107.

90 | Catti, M., Valerio, G., and Dovesi, R. Theoretical study of electronic, magnetic, and structural properties of α-Fe_2O_3 (hematite). *Physical Review B*, 51(12), **1995**:7441–7450.

91 | Kokubun, J., et al. Chiral and magnetic effects in forbidden x-ray scattering from antiferromagnetic hematite α-Fe_2O_3 and eskolaite Cr_2O_3. *Physical Review B*, 78(11), **2008**:115,112.

92 | Bak, T., et al. Photo-electrochemical hydrogen generation from water using solar energy. Materials-related aspects. *International Journal of Hydrogen Energy*, 27(10), **2002**:991–1022.

93 | Radecka, M., et al. Importance of the band gap energy and flat band potential for application of modified TiO2 photoanodes in water photolysis. *Journal of Power Sources*, 181(1), **2008**:46–55.

94 | Dare-Edwards, M. P., et al. Electrochemistry and photoelectrochemistry of iron(III) oxide. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 79(9), **1983**:2027.

95 | Peter, L. M. Energetics and kinetics of light-driven oxygen evolution at semiconductor electrodes: the example of hematite. *Journal of Solid State Electrochemistry*, 17(2), **2012**:315–326.

96 | Guillaumont, D. and Nakamura, S. Calculation of the absorption wavelength of dyes using time-dependent density-functional theory (TD-DFT). *Dyes and Pigments*, 46(2), **2000**:85–92.

97 | Cesar, I., et al. Influence of Feature Size, Film Thickness, and Silicon Doping on the Performance of Nanostructured Hematite Photoanodes for Solar Water Splitting. *The Journal of Physical Chemistry C*, 113(2), **2009**:772–782.

98 | Pham, T. A., et al. Interfacial Effects on the Band Edges of Functionalized Si Surfaces in Liquid Water. *Journal of the American Chemical Society*, 136(49), **2014**:17,071–17,077.

99 | Hohenberg, P. and Kohn, W. Inhomogeneous Electron Gas. *Physical Review*, 136, **1964**:864–871.

100 | Platts, J. A. Theoretical electron densities in transition metal dihydrides. *Journal of Molecular Structure: THEOCHEM*, 545(1-3), **2001**:111–118.

101 | Di Fabrizio, E., et al. Spin density of ordered FeCo: A failure of the local-spin-density approximation. *Physical Review B*, 40(14), **1989**:9502–9507.

102 | Fähnle, M., et al. Magnetoelasticity of Fe: Possible failure of ab initio electron theory with the local-spin-density approximation and with the generalized-gradient approximation. *Physical Review B*, 65(14), **2002**:144,436.

103 | Tokura, Y. Orbital Physics in Transition-Metal Oxides. *Science*, 288(5465), **2000**:462–468.

104 | Wang, L., Maxisch, T., and Ceder, G. Oxidation energies of transition metal oxides within theGGA+Uframework. *Physical Review B*, 73(19), **2006**:195,107.

105 | Hubbard, J. Electron Correlations in Narrow Energy Bands. IV. The Atomic Representation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 285(1403), **1965**:542–560.

106 | Anisimov, V. I., Aryasetiawan, F., and Lichtenstein, A. I. First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA+U method. *Journal of Physics: Condensed Matter*, 9(4), **1997**:767–808.

107 | Dudarev, S. L., et al. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Physical Review B*, 57(3), **1998**:1505–1509.

108 | Dudarev, S. L., Manh, D. N., and Sutton, A. P. Effect of Mott-Hubbard correlations on the electronic structure and structural stability of uranium dioxide. *Philosophical Magazine Part B*, 75(5), **1997**:613–628.

109 | Solovyev, I. V. and Dederichs, P. H. Corrected atomic limit in the local-density approximation and the electronic structure of d impurities in Rb. *Physical Review B*, 50(23), **1994**:16,861–16,871.

110 | Gunnarsson, O., et al. Density-functional calculation of the parameters in the Anderson model: Application to Mn in CdTe. *Physical Review B*, 39(3), **1989**:1708–1722.

111 | McMahan, A. K., Martin, R. M., and Satpathy, S. Calculated effective Hamiltonian forLa2CuO4and solution in the impurity Anderson approximation. *Physical Review B*, 38(10), **1988**:6650–6666.

112 | Springer, M. and Aryasetiawan, F. Frequency-dependent screened interaction in Ni within the random-phase approximation. *Physical Review B*, 57(8), **1998**:4364–4368.

113 | Cococcioni, M. and de Gironcoli, S. Linear response approach to the calculation of the effective interaction parameters in the LDA+U method. *Physical Review B*, 71(3), **2005**:035,105.

114 | Aryasetiawan, F., et al. Calculations of Hubbard-U from first-principles. *Physical Review B*, 74(12), **2006**:125,106.

115 | Velev, J., et al. Electronic and magnetic structure of transition-metal-doped α-hematite. *Physical Review B*, 71(20), **2005**:205,208.

116 | Pentcheva, R. and Nabi, H. Interface magnetism in Fe2O3/FeTiO3 heterostructures. *Physical Review B*, 77(17), **2008**:172,405.

117 | Huang, X., Ramadugu, S. K., and Mason, S. E. Surface-Specific DFT +UApproach Applied to α-Fe2O3(0001). *The Journal of Physical Chemistry C*, 120(9), **2016**:4919–4930.

118 | Alexandrov, V., et al. Electron Exchange and Conduction in Nontronite from First-Principles. *The Journal of Physical Chemistry C*, 117(5), **2013**:2032–2040.

119 | Alexandrov, V. and Rosso, K. M. Insights into the Mechanism of Fe(II) Adsorption and Oxidation at Fe-Clay Mineral Surfaces from First-Principles Calculations. *The Journal of Physical Chemistry C*, 117(44), **2013**:22,880–22,886.

120 | Leonov, I., et al. Charge and Orbital Order in $Fe_{3}O_{4}$. *Physical Review Letters*, 93(14), **2004**:146,404.

121 | Mulakaluri, N., et al. Partial Dissociation of Water on $Fe_{3}O_{4}$(001): Adsorbate Induced Charge and Orbital Order. *Physical Review Letters*, 103(17), **2009**:176,102.

122 | Mulakaluri, N., Pentcheva, R., and Scheffler, M. Coverage-Dependent Adsorption Mode of Water on Fe_3O_4(001): Insights from First Principles Calculations. *The Journal of Physical Chemistry C*, 114(25), **2010**:11,148–11,156.

123 | Punkkinen, M. P. J., et al. Fe_2O_3 within the LSDA +U approach. *Journal of Physics: Condensed Matter*, 11(11), **1999**:2341–2349.

124 | Krukau, A. V., et al. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *The Journal of chemical physics*, 125(22), **2006**:224,106.

125 | Heyd, J., Scuseria, G. E., and Ernzerhof, M. Erratum: "Hybrid functionals based on a screened Coulomb potential" [J. Chem. Phys. 118, 8207 (2003)]. *The Journal of Chemical Physics*, 124(21), **2006**:219,906.

126 | Heyd, J. and Scuseria, G. E. Assessment and validation of a screened Coulomb hybrid density functional. *The Journal of Chemical Physics*, 120(16), **2004**:7274.

127 | Heyd, J., Scuseria, G. E., and Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics*, 118(18), **2003**:8207.

128 | Perdew, J. P., Burke, K., and Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 77(18), **1996**:3865–3868.

129 | Perdew, J. P., Burke, K., and Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Physical Review Letters*, 78(7), **1997**:1396–1396.

130 | Adamo, C. and Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics*, 110(13), **1999**:6158.

131 | Perdew, J. P., Ernzerhof, M., and Burke, K. Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics*, 105(22), **1996**:9982.

132 | VanVoorhis, T. and Scuseria, G. E. A novel form for the exchange-correlation energy functional. *The Journal of Chemical Physics*, 109(2), **1998**:400.

133 | Zhou, Z., Shi, J., and Guo, L. A comparative study on structural and electronic properties and formation energy of bulk α-Fe2O3 using first-principles calculations with different density functionals. *Computational Materials Science*, 113, **2016**:117–122.

134 | Guidon, M., Hutter, J., and VandeVondele, J. Auxiliary Density Matrix Methods for Hartree-Fock Exchange Calculations. *Journal of Chemical Theory and Computation*, 6(8), **2010**:2348–2364.

135 | VandeVondele, J. and Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics*, 127(11), **2007**:114,105.

136 | Weigend, F. and Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 7(18), **2005**:3297.

137 | Grimme, S., Ehrlich, S., and Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 32(7), **2011**:1456–1465.

138 | Chai, J.-D. and Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Physical Chemistry Chemical Physics*, 10(44), **2008**:6615.

139 | Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12), **2004**:1463–1473.

140 | Wu, Q. and Yang, W. Empirical correction to density functional theory for van der Waals interactions. *The Journal of Chemical Physics*, 116(2), **2002**:515.

141 | Mooij, W. T. M., et al. Transferable ab Initio Intermolecular Potentials. 1. Derivation from Methanol Dimer and Trimer Calculations. *The Journal of Physical Chemistry A*, 103(48), **1999**:9872–9882.

142 | Bondi, A. van der Waals Volumes and Radii. *Journal of Physical Chemistry*, 68(3), **1964**:441–451.

143 | Johnson, E. R. and Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *The Journal of Chemical Physics*, 124(17), **2006**:174,104.

144 | Finger, L. W. and Hazen, R. M. Crystal structure and isothermal compression of Fe_2O_3, Cr_2O_3, and V_2O_3 to 50 kbars. *Journal of Applied Physics*, 51(10), **1980**:5362.

145 | Blake, R. L., et al. Refinement of the Hematite Structure. *American Mineralogist*, 51, **1951**:123–129.

146 | Grimme, S., et al. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics*, 132(15), **2010**:154,104.

147 | VandeVondele, J., et al. The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. *The Journal of Chemical Physics*, 122(1), **2005**:014,515.

148 | Conesa, J. C. The Relevance of Dispersion Interactions for the Stability of Oxide Phases. *The Journal of Physical Chemistry C*, 114(51), **2010**:22,718–22,726.

149 | Pozun, Z. D. and Henkelman, G. Hybrid density functional theory band structure engineering in hematite. *The Journal of Chemical Physics*, 134(22), **2011**:224,706.

150 | Pauling, L. and Hendricks, S. B. The Crystal Structures of Hematite and Corundum. *Journal of the American Chemical Society*, 47(3), **1925**:781–790.

151 | Monkhorst, H. J. and Pack, J. D. Special points for Brillouin-zone integrations. *Physical Review B*, 13(12), **1976**:5188–5192.

152 | Sherman, D. M. The electronic structures of Fe3+ coordination sites in iron oxides: Applications to spectra, bonding, and magnetism. *Physics and Chemistry of Minerals*, 12(3), **1985**:161–175.

153 | Krén, E., Szabó, P., and Konczos, G. Neutron diffraction studies on the (1-x) Fe_2O_3-xRh_2O_3 system. *Physics Letters*, 19(2), **1965**:103–104.

154 | Güell, M., et al. Importance of the Basis Set for the Spin-State Energetics of Iron Complexes. *The Journal of Physical Chemistry A*, 112(28), **2008**:6384–6391.

155 | Leach, A. R. Molecular Modelling. Principles and Applications. Prentice Hall, **2001**.

156 | Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *Journal of Chemical Physics*, 23, **1955**:1833–1840.

157 | Wigner, E. and Seitz, F. On the Constitution of Metallic Sodium. *Physical Review*, 43(10), **1933**:804–810.

158 | Dobson, J. F., White, A., and Rubio, A. Asymptotics of the Dispersion Interaction: Analytic Benchmarks for van der Waals Energy Functionals. *Physical Review Letters*, 96(7), **2006**:073,201.

159 | Mori-Sanchez, P., Cohen, A. J., and Yang, W. Many-electron self-interaction error in approximate density functionals. *The Journal of Chemical Physics*, 125(20), **2006**:201,102.

160 | McIntyre, N. S. and Zetaruk, D. G. X-ray photoelectron spectroscopic studies of iron oxides. *Analytical Chemistry*, 49(11), **1977**:1521–1529.

161 | Fujimori, A., et al. Photoemission satellites and electronic structure of Fe_2O_3. *Physical Review B*, 34(10), **1986**:7318–7328.

162 | Shull, C., Strauser, W., and Wollan, E. Neutron Diffraction by Paramagnetic and Antiferromagnetic Substances. *Physical Review*, 83(2), **1951**:333–345.

163 | Butler, W. H., Bandyopadhyay, A., and Srinivasan, R. Electronic and magnetic structure of a 1000 K magnetic semiconductor: α-hematite (Ti). *Journal of Applied Physics*, 93(10), **2003**:7882.

164 | Zaanen, J., Sawatzky, G. A., and Allen, J. W. Band gaps and electronic structure of transition-metal compounds. *Phys. Rev. Lett.*, 55, **1985**:418–421.

165 | Zaanen, J. and Sawatzky, G. Systematics in band gaps and optical spectra of 3D transition metal compounds. *Journal of Solid State Chemistry*, 88(1), **1990**:8 – 27. ISSN 0022-4596.

166 | Zhou, Z., et al. Understanding Hematite Doping with Group IV Elements: A DFT+UStudy. *The Journal of Physical Chemistry C*, 119(47), **2015**:26,303–26,310.

167 | Hu, Y.-S., et al. Pt-Doped α-Fe_2O_3 Thin Films Active for Photoelectrochemical Water Splitting. *Chemistry of Materials*, 20(12), **2008**:3803–3805.

168 | Cesar, I., et al. Influence of Feature Size, Film Thickness, and Silicon Doping on the Performance of Nanostructured Hematite Photoanodes for Solar Water Splitting. *The Journal of Physical Chemistry C*, 113(2), **2009**:772–782.

169 | Liao, P., Keith, J. A., and Carter, E. A. Water Oxidation on Pure and Doped Hematite (0001) Surfaces: Prediction of Co and Ni as Effective Dopants for Electrocatalysis. *Journal of the American Chemical Society*, 134(32), **2012**:13,296–13,309.

170 | Bu, Q., et al. A Ni2P modified Ti4+ doped Fe2O3 photoanode for efficient solar water oxidation by promoting hole injection. *Dalton Trans.*, **2017**.

171 | Deng, J., et al. Lowering the Onset Potential of Fe2TiO5/Fe2O3 Photoanodes by Interface Structures: F- and Rh-Based Treatments. *ACS Catalysis*, **2017**:4062–4069.

172 | Cava, C. E., Roman, L. S., and Persson, C. Effects of native defects on the structural and magnetic properties of hematite α-Fe2O3. *Physical Review B*, 88(4), **2013**:045,136.

173 | Lee, J. and Han, S. Thermodynamics of native point defects in α-Fe2O3: an ab initio study. *Physical Chemistry Chemical Physics*, 15(43), **2013**:18,906.

174 | Wheeler, D. A., et al. Nanostructured hematite: synthesis, characterization, charge carrier dynamics, and photoelectrochemical properties. *Energy and Environmental Science*, 5(5), **2012**:6682.

175 | Adinaveen, T., Judith Vijaya, J., and John Kennedy, L. Studies on the Structural, Morphological, Optical, and Magnetic Properties of α-Fe_2O_3 Nanostructures by a Simple One-Step Low Temperature Reflux Condensing Method. *Journal of Superconductivity and Novel Magnetism*, 27(7), **2014**:1721–1727.

176 | Braun, A., et al. Direct Observation of Two Electron Holes in a Hematite Photoanode during Photoelectrochemical Water Splitting. *The Journal of Physical Chemistry C*, 116(32), **2012**:16,870–16,875.

177 | Iordanova, N., Dupuis, M., and Rosso, K. M. Charge transport in metal oxides: A theoretical study of hematite α-Fe2O3. *The Journal of Chemical Physics*, 122(14), **2005**:144,305.

178 | Liao, P. and Carter, E. A. Hole transport in pure and doped hematite. *Journal of Applied Physics*, 112(1), **2012**:013,701.

179 | Adelstein, N., et al. Density functional theory based calculation of small-polaron mobility in hematite. *Physical Review B*, 89(24), **2014**:245,115.

180 | Blumberger, J. and McKenna, K. P. Constrained density functional theory applied to electron tunnelling between defects in MgO. *Physical Chemistry Chemical Physics*, 15(6), **2013**:2184.

181 | Peng, H. and Lany, S. Semiconducting transition-metal oxides based on d^5 cations: Theory for MnO and Fe_2O_3. *Phys. Rev. B*, 85, **2012**:201,202.

182 | Grossman, J. C., et al. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. *The Journal of Chemical Physics*, 120(1), **2004**:300.

183 | Santra, B., et al. On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters. II. The water hexamer and van der Waals interactions. *The Journal of Chemical Physics*, 129(19), **2008**:194,111.

184 | Soper, A. K. Joint structure refinement of x-ray and neutron diffraction data on disordered materials: application to liquid water. *Journal of Physics: Condensed Matter*, 19(33), **2007**:335,206.

185 | Todorova, T., et al. Molecular Dynamics Simulation of Liquid Water: Hybrid Density Functionals. *The Journal of Physical Chemistry B*, 110(8), **2006**:3685–3691.

186 | Kuo, I.-F. W., et al. Liquid Water from First Principles: Investigation of Different Sampling Approaches. *The Journal of Physical Chemistry B*, 108(34), **2004**:12,990–12,998.

187 | Grossman, J. C., et al. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. *The Journal of Chemical Physics*, 120(1), **2004**:300.

188 | Goerigk, L. and Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Physical Chemistry Chemical Physics*, 13(14), **2011**:6670.

189 | Moellmann, J. and Grimme, S. DFT-D3 Study of Some Molecular Crystals. *The Journal of Physical Chemistry C*, 118(14), **2014**:7615–7621.

190 | Soper, A. K., Bruni, F., and Ricci, M. A. Site–site pair correlation functions of water from 25 to 400 °C: Revised analysis of new and old diffraction data. *The Journal of Chemical Physics*, 106(1), **1997**:247.

191 | Narten, A. H. Liquid Water: Atom Pair Correlation Functions from Neutron and X-Ray Diffraction. *The Journal of Chemical Physics*, 56(11), **1972**:5681.

192 | Ling, Y., et al. Sn-Doped Hematite Nanostructures for Photoelectrochemical Water Splitting. *Nano Letters*, 11(5), **2011**:2119–2125.

193 | Cheng, J., et al. Hole Localization and Thermochemistry of Oxidative Dehydrogenation of Aqueous Rutile TiO2(110). *ChemCatChem*, 4(5), **2012**:636–640.

194 | Vura-Weis, J., et al. Femtosecond $M_{\{2,3\}}$-Edge Spectroscopy of Transition-Metal Oxides: Photoinduced Oxidation State Change in α-Fe_2O_3. *The Journal of Physical Chemistry Letters*, 4(21), **2013**:3667–3671.

195 | Katz, J. E., et al. Electron Small Polarons and Their Mobility in Iron (Oxyhydr)oxide Nanoparticles. *Science*, 337(6099), **2012**:1200–1203.

196 | Fujii, T., et al. In situ RHEED and XPS studies of epitaxial thin α-Fe2O3(0001) films on sapphire. *Surface Science*, 366(3), **1996**:579–586.

197 | Andreussi, O., Dabo, I., and Marzari, N. Revised self-consistent continuum solvation in electronic-structure calculations. *The Journal of Chemical Physics*, 136(6), **2012**:064,102.

198 | Fernandez-Serra, M. V. and Artacho, E. Network equilibration and first-principles liquid water. *The Journal of Chemical Physics*, 121(22), **2004**:11,136.

199 | Jorgensen, W. L., et al. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79, **1983**:926–935.

200 | Wales, D. J. and Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *The Journal of Physical Chemistry A*, 101(28), **1997**:5111–5116.

201 | O'Shea, S. J., Welland, M. E., and Rayment, T. Solvation forces near a graphite surface measured with an atomic force microscope. *Applied Physics Letters*, 60(19), **1992**:2356–2358.

202 | Hirunsit, P. and Balbuena, P. B. Effects of Confinement on Water Structure and Dynamics: A Molecular Simulation Study. *The Journal of Physical Chemistry C*, 111(4), **2007**:1709–1715.

203 | Curtiss, L. A., et al. Nonadditivity of ab initio pair potentials for molecular dynamics of multivalent transition metal ions in water. *The Journal of Chemical Physics*, 86(4), **1987**:2319.

204 | Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics*, 81, **1984**:511–519.

205 | Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31, **1985**:1695–1697.

206 | von Rudorff, G. F., et al. Hematite(001)-liquid water interface from hybrid density functional-based molecular dynamics. *Journal of Physics: Condensed Matter*, 28(39), **2016**:394,001.

207 | Chatman, S., et al. Effect of surface site interactions on potentiometric titration of hematite (α-Fe2O3) crystal faces. *Journal of Colloid and Interface Science*, 391, **2013**:125–134.

208 | Haghighat, S. and Dawlaty, J. M. Continuous Representation of the Proton and Electron Kinetic Parameters in the pH–Potential Space for Water Oxidation on Hematite. *The Journal of Physical Chemistry C*, **2015**:150313092955,002.

209 | Kallay, N., Dojnović, Z., and Čop, A. Surface potential at the hematite-water interface. *Journal of Colloid and Interface Science*, 286(2), **2005**:610–614.

210 | Tardy, Y. and Nahon, D. Geochemistry of laterites, stability of Al-goethite, Al-hematite, and Fe (super 3+) -kaolinite in bauxites and ferricretes; an approach to the mechanism of concretion formation. *American Journal of Science*, 285(10), **1985**:865–903.

211 | Hutter, J., et al. CP2K: Atomistic Simulations of Condensed Matter Systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(1), **2014**:15–25.

212 | Tanwar, K. S., et al. Fe(II) adsorption on hematite (0001). *Geochimica et Cosmochimica Acta*, 73(15), **2009**:4346–4365.

213 | Robinson, I. K. Crystal truncation rods and surface roughness. *Physical Review B*, 33(6), **1986**:3830–3836.

214 | Sanson, A., Mathon, O., and Pascarelli, S. Local vibrational dynamics of hematite (α-Fe2O3) studied by extended x-ray absorption fine structure and molecular dynamics. *The Journal of Chemical Physics*, 140(22), **2014**:224,504.

215 | Skelton, A. A., et al. Simulations of the Quartz(101-1)/Water Interface: A Comparison of Classical Force Fields, Ab Initio Molecular Dynamics, and X-ray Reflectivity Experiments. *The Journal of Physical Chemistry C*, 115(5), **2011**:2076–2088.

216 | Rossmeisl, J., et al. Calculated Phase Diagrams for the Electrochemical Oxidation and Reduction of Water over Pt(111). *The Journal of Physical Chemistry B*, 110(43), **2006**:21,833–21,839.

217 | Fiorentini, V. and Methfessel, M. Extracting convergent surface energies from slab calculations. *Journal of Physics: Condensed Matter*, 8(36), **1996**:6525–6529.

218 | Chen, C. T. and Cahan, B. D. Visible and ultraviolet optical properties of single-crystal and polycrystalline hematite measured by spectroscopic ellipsometry. *Journal of the Optical Society of America*, 71(8), **1981**:932.

219 | Young, K. F. and Frederikse, H. P. R. Compilation of the Static Dielectric Constant of Inorganic Solids. *Journal of Physical and Chemical Reference Data*, 2(2), **1973**:313.

220 | Nguyen, T. T., Grosberg, A. Y., and Shklovskii, B. I. Screening of a charged particle by multivalent counterions in salty water: Strong charge inversion. *The Journal of Chemical Physics*, 113(3), **2000**:1110.

221 | LeFormal, F., et al. Passivating surface states on water splitting hematite photoanodes with alumina overlayers. *Chem. Sci.*, 2(4), **2011**:737–743.

222 | Klahr, B., et al. Water Oxidation at Hematite Photoelectrodes: The Role of Surface States. *Journal of the American Chemical Society*, 134(9), **2012**:4294–4302.

223 | Silvi, B. and Savin, A. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature*, 371(6499), **1994**:683–686.

224 | Becke, A. D. and Edgecombe, K. E. A simple measure of electron localization in atomic and molecular systems. *The Journal of Chemical Physics*, 92(9), **1990**:5397.

225 | Savin, A., et al. Electron Localization in Solid-State Structures of the Elements: the Diamond Structure. *Angewandte Chemie International Edition in English*, 2, **1992**:187–188.

226 | Wang, J., et al. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25(2), **2006**:247–260.

227 | Kashtanov, S., et al. Local structures of liquid water studied by x-ray emission spectroscopy. *Physical Review B*, 69(2), **2004**:024,201.

228 | Cheng, J. and Sprik, M. Aligning electronic energy levels at theTiO2/H2Ointerface. *Physical Review B*, 82(8), **2010**:081,406.

229 | Ishii, H., et al. Kelvin probe study of band bending at organic semiconductor/metal interfaces: examination of Fermi level alignment. *physica status solidi (a)*, 201(6), **2004**:1075–1094.

230 | Kathmann, S. M., Kuo, I.-F. W., and Mundy, C. J. Electronic Effects on the Surface Potential at the Vapor−Liquid Interface of Water. *Journal of the American Chemical Society*, 130(49), **2008**:16,556–16,561.

231 | Otani, M. and Sugino, O. First-principles calculations of charged surfaces and interfaces: A plane-wave nonrepeated slab approach. *Physical Review B*, 73(11), **2006**:115,407.

232 | Steiner, T. The Hydrogen Bond in the Solid State. *Angewandte Chemie International Edition*, 41(1), **2002**:48–76.

233 | Luzar, A. and Chandler, D. Hydrogen-bond kinetics in liquid water. *Nature*, 379(6560), **1996**:55–57.

234 | Ohno, K., et al. 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. *Physical Chemistry Chemical Physics*, 7(16), **2005**:3005.

235 | Lee, T. D. and Yang, C. N. Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model. *Physical Review*, 87(3), **1952**:410–419.

236 | Kudo, T. and Katsura, S. A Method of Determining the Orderings of the Ising Model with Several Neighbor Interactions under the Magnetic Field and Applications to Hexagonal Lattices. *Progress of Theoretical Physics*, 56(2), **1976**:435–449.

237 | Wolfram, T. and Ellialtioglu, S. Applications of Group Theory to Atoms, Molecules, and Solids. Cambridge University Press, **2013**.

238 | von Rudorff, G. F., et al. Fast Interconversion of Hydrogen Bonding at the Hematite (001)–Liquid Water Interface. *Journal of Physical Chemistry Letters*, **2016**:1155–1160.

239 | Stillinger, F. H. Theory and Molecular Models for Water. *Advances in Chemical Physics*, 31, **1975**:1–101.

240 | Weinhold, F. and Klein, R. A. What is a hydrogen bond? Mutually consistent theoretical and experimental criteria for characterizing H-bonding interactions. *Molecular Physics*, 110(9-10), **2012**:565–579.

241 | Gasparotto, P. and Ceriotti, M. Recognizing molecular patterns by machine learning: An agnostic structural definition of the hydrogen bond. *The Journal of Chemical Physics*, 141(17), **2014**:174,110.

242 | Eaves, J. D., et al. Hydrogen bonds in liquid water are broken only fleetingly. *Proceedings of the National Academy of Sciences*, 102(37), **2005**:13,019–13,022.

243 | Zichi, D. A. and Rossky, P. J. Solvent molecular dynamics in regions of hydrophobic hydration. *The Journal of Chemical Physics*, 84(5), **1986**:2814.

244 | Henderson, M. The interaction of water with solid surfaces: fundamental aspects revisited. *Surface Science Reports*, 46(1-8), **2002**:1–308.

245 | He, Y., et al. Local ordering and electronic signatures of submonolayer water on anatase TiO2(101). *Nature Materials*, 8(7), **2009**:585–589.

246 | Ceccarelli, C., Jeffrey, G., and Taylor, R. A survey of O-H$\cdots$O hydrogen bond geometries determined by neutron diffraction. *Journal of Molecular Structure*, 70, **1981**:255–271.

247 | Markovitch, O. and Agmon, N. Structure and Energetics of the Hydronium Hydration Shells. *The Journal of Physical Chemistry A*, 111(12), **2007**:2253–2256.

248 | Fornili, S., Migliore, M., and Palazzo, M. Hydration of the hydronium ion. *Chemical Physics Letters*, 125(4), **1986**:419–424.

249 | Heuft, J. M. and Meijer, E. J. A density functional theory based study of the microscopic structure and dynamics of aqueous HCl solutions. *Physical Chemistry Chemical Physics*, 8(26), **2006**:3116.

250 | Tilocca, A. and Selloni, A. Structure and Reactivity of Water Layers on Defect-Free and Defective Anatase TiO2(101) Surfaces. *The Journal of Physical Chemistry B*, 108(15), **2004**:4743–4751.

251 | Nelson, B. P., et al. Control of Surface and ζ Potentials on Nanoporous TiO2Films by Potential-Determining and Specifically Adsorbed Ions. *Langmuir*, 16(15), **2000**:6094–6101.

252 | Markovitch, O., et al. Special Pair Dance and Partner Selection: Elementary Steps in Proton Transport in Liquid Water. *The Journal of Physical Chemistry B*, 112(31), **2008**:9456–9466.

253 | Brown, G. E., et al. Metal Oxide Surfaces and Their Interactions with Aqueous Solutions and Microbial Organisms. *Chemical Reviews*, 99(1), **1999**:77–174.

254 | Ceriotti, M., et al. Nuclear quantum effects and hydrogen bond fluctuations in water. *Proceedings of the National Academy of Sciences*, 110(39), **2013**:15,591–15,596.

255 | Agmon, N. The Grotthuss mechanism. *Chemical Physics Letters*, 244(5-6), **1995**:456–462.

256 | Kumar, N., et al. Faster proton transfer dynamics of water on SnO2 compared to TiO2. *The Journal of Chemical Physics*, 134(4), **2011**:044,706.

257 | Liu, P., et al. Reaction of water vapor with α-Al2O3(0001) and α-Fe2O3(0001) surfaces: synchrotron X-ray photoemission studies and thermodynamic calculations. *Surface Science*, 417(1), **1998**:53–65.

258 | DelloStritto, M. J., Kubicki, J., and Sofo, J. O. Density functional theory simulation of hydrogen-bonding structure and vibrational densities of states at the quartz (1 0 1)-water interface and its relation to dissolution as a function of solution pH and ionic strength. *Journal of Physics: Condensed Matter*, 26(24), **2014**:244,101.

259 | Bergermayer, W., Schweiger, H., and Wimmer, E. Ab initio-thermodynamics of oxide surfaces:O2onFe2O3(0001). *Physical Review B*, 69(19), **2004**:195,409.

260 | McBriarty, M. E., et al. Dynamic Stabilization of Metal Oxide–Water Interfaces. *Journal of the American Chemical Society*, 139(7), **2017**:2581–2584.

261 | Bargar, J., et al. XAFS and Bond-Valence Determination of the Structures and Compositions of Surface Functional Groups and Pb(II) and Co(II) Sorption Products on Single-Crystal α-Al2O3. *Journal of Colloid and Interface Science*, 185(2), **1997**:473–492.

262 | Kendelewicz, T., et al. X-ray absorption and photoemission study of the adsorption of aqueous Cr(VI) on single crystal hematite and magnetite surfaces. *Surface Science*, 424(2-3), **1999**:219–231.

263 | Aboud, S., Wilcox, J., and Brown, G. E. Density functional theory investigation of the interaction of water with α−Al2O3 and α−Fe2O3 (1-102) surfaces: Implications for surface reactivity. *Physical Review B*, 83(12), **2011**:125,407.

264 | Boily, J.-F., et al. Thin Water Films at Multifaceted Hematite Particle Surfaces. *Langmuir*, 31(48), **2015**:13,127–13,137.

265 | Blöchl, P. E. Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges. *The Journal of Chemical Physics*, 103(17), **1995**:7422.

266 | Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theoretica Chimica Acta*, 44(2), **1977**:129–138.

267 | Stavitski, E. and de Groot, F. M. The CTM4XAS program for EELS and XAS spectral shape analysis of transition metal L edges. *Micron*, 41(7), **2010**:687–694.

268 | Lad, R. J. and Henrich, V. E. Photoemission study of the valence-band electronic structure in Fe_xO,Fe_3O_4, and α-Fe_2O_3 single crystals. *Physical Review B*, 39(18), **1989**:13,478–13,485.

269 | Bair, R. A. and Goddard, W. A. Theoretical studies of the ground and excited states of a model of the active site in oxidized and reduced rubredoxin. *Journal of the American Chemical Society*, 100(18), **1978**:5669–5676.

270 | Walker, L., Wertheim, G., and Jaccarino, V. Interpretation of the Fe{57} Isomer Shift. *Physical Review Letters*, 6(3), **1961**:98–101.

271 | Goerigk, L. and Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions – Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *Journal of Chemical Theory and Computation*, 6(1), **2010**:107–126.

272 | Jacquemin, D., et al. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *Journal of Chemical Theory and Computation*, 5(9), **2009**:2420–2435.

273 | Janesko, B. G., Henderson, T. M., and Scuseria, G. E. Screened hybrid density functionals for solid-state chemistry and physics. *Phys. Chem. Chem. Phys.*, 11(3), **2009**:443–454.

274 | Blumberger, J. and McKenna, K. P. Constrained density functional theory applied to electron tunnelling between defects in MgO. *Physical Chemistry Chemical Physics*, 15(6), **2013**:2184.

275 | Perdew, J. P. and Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10), **1981**:5048–5079.

276 | Wu, Q. and Van Voorhis, T. Extracting electron transfer coupling elements from constrained density functional theory. *The Journal of Chemical Physics*, 125(16), **2006**:164,105.

277 | Cohen, A. J., Mori-Sanchez, P., and Yang, W. Insights into Current Limitations of Density Functional Theory. *Science*, 321(5890), **2008**:792–794.

278 | Hong, J., et al. Spin-phonon coupling effects in transition-metal perovskites: A DFT + Uand hybrid-functional study. *Physical Review B*, 85(5), **2012**:054,417.

279 | Hu, Z. and Metiu, H. Choice of U for DFT+UCalculations for Titanium Oxides. *The Journal of Physical Chemistry C*, 115(13), **2011**:5841–5845.

280 | Kulik, H. J., et al. Density Functional Theory in Transition-Metal Chemistry: A Self-Consistent HubbardU-Approach. *Physical Review Letters*, 97(10), **2006**:103,001.

281 | Jain, A., et al. Formation enthalpies by mixing GGA and GGA+Ucalculations. *Physical Review B*, 84(4), **2011**:045,115.

282 | Pacchioni, G. First Principles Calculations on Oxide-Based Heterogeneous Catalysts and Photocatalysts: Problems and Advances. *Catalysis Letters*, 145(1), **2014**:80–94.

283 | Tsuneda, T. and Hirao, K. Self-interaction corrections in density functional theory. *The Journal of Chemical Physics*, 140(18), **2014**:18A513.

284 | Heyd, J., et al. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *The Journal of Chemical Physics*, 123(17), **2005**:174,101.

285 | Furche, F. and Perdew, J. P. The performance of semilocal and hybrid density functionals in 3d transition-metal chemistry. *The Journal of Chemical Physics*, 124(4), **2006**:044,103.

286 | Archer, T., et al. Exchange interactions and magnetic phases of transition metal oxides: Benchmarking advancedab initiomethods. *Physical Review B*, 84(11), **2011**:115,114.

287 | Franchini, C., et al. Density functional theory study of MnO by a hybrid functional approach. *Physical Review B*, 72(4), **2005**:045,132.

288 | VandeVondele, J., Borštnik, U., and Hutter, J. Linear Scaling Self-Consistent Field Calculations with Millions of Atoms in the Condensed Phase. *Journal of Chemical Theory and Computation*, 8(10), **2012**:3565–3573.

289 | Strout, D. L. and Scuseria, G. E. A quantitative study of the scaling properties of the Hartree–Fock method. *The Journal of Chemical Physics*, 102(21), **1995**:8448.

290 | Häser, M. and Ahlrichs, R. Improvements on the direct SCF method. *Journal of Computational Chemistry*, 10(1), **1989**:104–111.

291 | Kohn, W. Density functional theory for systems of very many atoms. *International Journal of Quantum Chemistry*, 56(4), **1995**:229–232.

292 | Yanai, T., Tew, D. P., and Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters*, 393(1-3), **2004**:51–57.

293 | Gerber, I. C. and Ángyán, J. G. Hybrid functional with separated range. *Chemical Physics Letters*, 415(1-3), **2005**:100–105.

294 | Izmaylov, A. F., Scuseria, G. E., and Frisch, M. J. Efficient evaluation of short-range Hartree-Fock exchange in large molecules and periodic systems. *The Journal of Chemical Physics*, 125(10), **2006**:104,103.

295 | Vydrov, O. A., et al. Importance of short-range versus long-range Hartree-Fock exchange for the performance of hybrid density functionals. *The Journal of Chemical Physics*, 125(7), **2006**:074,106.

296 | Kronik, L., et al. Excitation Gaps of Finite-Sized Systems from Optimally Tuned Range-Separated Hybrid Functionals. *Journal of Chemical Theory and Computation*, 8(5), **2012**:1515–1531.

297 | Halkier, A., et al. Basis-set convergence of the energy in molecular Hartree–Fock calculations. *Chemical Physics Letters*, 302(5-6), **1999**:437–446.

298 | Heyd, J. and Scuseria, G. E. Efficient hybrid density functional calculations in solids: Assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional. *The Journal of Chemical Physics*, 121(3), **2004**:1187.

299 | Rivero, P., et al. Description of magnetic interactions in strongly correlated solids via range-separated hybrid functionals. *Physical Review B*, 79(24), **2009**:245,129.

300 | von Rudorff, G. F., et al. Improving the Performance of Hybrid Functional-Based Molecular Dynamics Simulation through Screening of Hartree–Fock Exchange Forces. *Journal of Chemical Theory and Computation*, 13(5), **2017**:2178–2184.

301 | Calvo, F., et al. Phase changes in 38-atom Lennard-Jones clusters. II. A parallel tempering study of equilibrium and dynamic properties in the molecular dynamics and microcanonical ensembles. *The Journal of Chemical Physics*, 112(23), **2000**:10,350–10,357.

302 | Kolafa, J. Time-reversible always stable predictor-corrector method for molecular dynamics of polarizable molecules. *Journal of Computational Chemistry*, 25(3), **2004**:335–342.

303 | Niklasson, A. M. N., Tymczak, C. J., and Challacombe, M. Time-Reversible Born-Oppenheimer Molecular Dynamics. *Physical Review Letters*, 97(12), **2006**:123,001.

304 | Fattebert, J.-L. and Gygi, F. Linear-scaling first-principles molecular dynamics with plane-waves accuracy. *Physical Review B*, 73(11), **2006**:115,124.

305 | Ufimtsev, I. S. and Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *Journal of Chemical Theory and Computation*, 5(10), **2009**:2619–2628.

306 | Herbert, J. M. and Head-Gordon, M. Accelerated, energy-conserving Born–Oppenheimer molecular dynamics via Fock matrix extrapolation. *Physical Chemistry Chemical Physics*, 7(18), **2005**:3269.

307 | VandeVondele, J. and Sprik, M. A molecular dynamics study of the hydroxyl radical in solution applying self-interaction-corrected density functional methods. *Physical Chemistry Chemical Physics*, 7(7), **2005**:1363.

308 | Oshikiri, M., et al. Electronic structures of promising photocatalysts InMO4 (M=V, Nb, Ta) and BiVO4 for water decomposition in the visible wavelength region. *The Journal of Chemical Physics*, 117(15), **2002**:7313–7318.

309 | Byström, A. and Byström, A. M. The crystal structure of hollandite, the related manganese oxide minerals, and α-MnO2. *Acta Crystallographica*, 3(2), **1950**:146–154.

310 | Asbrink, S. and Waskowska, A. CuO: X-ray single-crystal structure determination at 196 K and room temperature. *Journal of Physics: Condensed Matter*, 3(42), **1991**:8173–8180.

311 | Liao, P. and Carter, E. A. Testing variations of the GW approximation on strongly correlated transition metal oxides: hematite (α-Fe2O3) as a benchmark. *Physical Chemistry Chemical Physics*, 13(33), **2011**:15,189.

312 | Liao, L., et al. Efficient solar water-splitting using a nanocrystalline CoO photocatalyst. *Nature Nanotechnology*, 9(1), **2013**:69–73.

313 | Sasaki, S., Fujino, K., and Takéuchi, Y. X-Ray Determination of Electron-Density Distributions in Oxides, MgO, MnO, CoO, and NiO, and Atomic Scattering Factors of their Constituent Atoms. *Proceedings of the Japan Academy. Ser. B: Physical and Biological Sciences*, 55(2), **1979**:43–48.

314 | Blake, R. L., et al. Refinement of the Hematite Structure. *American Mineralogist*, 51, **1966**:123–129.

315 | Vassilev, P., Louwerse, M. J., and Baerends, E. J. Ab initio molecular dynamics simulation of the OH radical in liquid water. *Chemical Physics Letters*, 398(1-3), **2004**:212–216.

316 | VandeVondele, J., et al. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Computer Physics Communications*, 167(2), **2005**:103–128.

317 | Kandalla, K., et al. Designing topology-aware collective communication algorithms for large scale InfiniBand clusters: Case studies with Scatter and Gather. *2010 IEEE International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW)*, **2010**:1–8.

318 | Guidon, M., Hutter, J., and VandeVondele, J. Robust Periodic Hartree−Fock Exchange for Large-Scale Simulations Using Gaussian Basis Sets. *Journal of Chemical Theory and Computation*, 5(11), **2009**:3010–3021.

319 | Guidon, M., et al. Ab initio molecular dynamics using hybrid density functionals. *The Journal of Chemical Physics*, 128(21), **2008**:214,104.

320 | Mangold, M., et al. Absolute pKa Values and Solvation Structure of Amino Acids from Density Functional Based Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation*, 7(6), **2011**:1951–1961.

321 | Sulpizi, M. and Sprik, M. Acidity constants from vertical energy gaps: density functional theory based molecular dynamics implementation. *Physical Chemistry Chemical Physics*, 10(34), **2008**:5238.

322 | Straatsma, T. P. and Berendsen, H. J. C. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *The Journal of Chemical Physics*, 89(9), **1988**:5876–5886.

323 | Hummer, G. Fast-growth thermodynamic integration: Error and efficiency analysis. *The Journal of Chemical Physics*, 114(17), **2001**:7330–7337.

324 | Shirts, M. R. and Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *The Journal of Chemical Physics*, 122(14), **2005**:144,107.

325 | Bruckner, S. and Boresch, S. Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method. *Journal of Computational Chemistry*, 32(7), **2011**:1303–1319.

326 | Jorge, M., et al. Effect of the Integration Method on the Accuracy and Computational Efficiency of Free Energy Calculations Using Thermodynamic Integration. *Journal of Chemical Theory and Computation*, 6(4), **2010**:1018–1027.

327 | Rodinger, T. and Pomès, R. Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Current Opinion in Structural Biology*, 15(2), **2005**:164–170.

328 | Fasnacht, M., Swendsen, R. H., and Rosenberg, J. M. Adaptive integration method for Monte Carlo simulations. *Physical Review E*, 69(5), **2004**:056,704.

329 | Kästner, J. and Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *The Journal of Chemical Physics*, 123(14), **2005**:144,104.

330 | Yang, W., Bitetti-Putzer, R., and Karplus, M. Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *The Journal of Chemical Physics*, 120(6), **2004**:2618–2628.

331 | Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6), **1988**:3098–3100.

332 | Bussi, G., Donadio, D., and Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1), **2007**:014,101.

333 | Goedecker, S., Teter, M., and Hutter, J. Separable dual-space Gaussian pseudopotentials. *Physical Review B*, 54(3), **1996**:1703–1710.

334 | Kubas, A., et al. Electronic couplings for molecular charge transfer: Benchmarking CDFT, FODFT, and FODFTB against high-level ab initio calculations. *The Journal of Chemical Physics*, 140(10), **2014**:104,105.

335 | Jones, J. E. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society of London. Series A*, 106, **1924**:463–477.

336 | Møller, C. and Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review*, 46, **1934**:618–622.

337 | d'Avezac, M., Calandra, M., and Mauri, F. Density functional theory description of hole-trapping in SiO2: A self-interaction-corrected approach. *Physical Review B*, 71(20), **2005**:205,210.

338 | Borštnik, U., et al. Sparse matrix multiplication: The distributed block-compressed sparse row library. *Parallel Computing*, 40(5-6), **2014**:47–58.

339 | Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theoretical Chemistry Accounts*, 114(1-3), **2005**:145–152.

340 | Frigo, M. and Johnson, S. The Design and Implementation of FFTW3. *Proceedings of the IEEE*, 93(2), **2005**:216–231.

341 | VandeVondele, J. and Hutter, J. An efficient orbital transformation method for electronic structure calculations. *The Journal of Chemical Physics*, 118(10), **2003**:4365.

342 | Hartwigsen, C., Goedecker, S., and Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Physical Review B*, 58(7), **1998**:3641–3662.

343 | Lippert, G., Parrinello, M., and Hutter, J. A hybrid Gaussian and plane wave density functional scheme. *Molecular Physics*, 92(3), **2010**:477–488.

344 | Schütt, O., et al. GPU-Accelerated Sparse Matrix-Matrix Multiplication for Linear Scaling Density Functional Theory. In *Electronic Structure Calculations on Graphics Processing Units*.

Wiley, **2016**, 173–190. URL http://dx.doi.org/10.1002/9781118670712.ch8.