Quantum to Transport

Modeling Transport Properties of Aqueous Potassium Hydroxide by Machine Learning Molecular Force Fields from Quantum Mechanics

Master Thesis

Vladimir Jelle Lagerweij

Delft University of Technology

TU Delft

Quantum to Transport

Modeling Transport Properties of Aqueous Potassium Hydroxide by Machine Learning Molecular Force Fields from Quantum Mechanics

by

Vladimir Jelle Lagerweij

Student number:	4982878
Supervisors:	O. A. Moultos P. Dey P. Habibi
Thesis comitee:	O. A. Moultos T. J. H. Vlugt R. Taormina P. Habibi
Faculty:	Faculty of Mechanical, Maritime and Materials Engineering, Delft
Cover:	System of H_2O modeled using AIMD, visualized with the VMD software package[a].

[a] Humphrey, W.; Dalke, A.; Schulten, K "VMDK Visual Molecular Dynamics" . *Journal of Molecular Graphics* **1996**, *14*, 33-38.

TU Delft

Acknowledgments

This document represents the completion of my masters in Mechanical Engineering at the Delft University of Technology. I extend my sincere gratitude to my supervisors Dr. Othonas Moultos, Dr. Poulumi Dey, and Parsa Habibi. Dr. Othonas Moultos' meticulous feedback and Parsa Habibi's unwavering availability for my queries have been indispensable. Dr. Poulumi Dey's presence during progress meetings and her insights into density functional theory have significantly enriched this work. I also thank all members of my thesis comitee, Dr. Othonas Moultos, Prof. Dr. Thijs Vlugt, Dr. Riccardo Taormina and Parsa Habibi for the time and effort spend in reading this work in detail.

In addition, I am deeply thankful to Prof. Dr. Thijs Vlugt, Seyed Jamali, and Parsa Habibi for their support during my honors projects, particularly in the investigation of accurately predicting electric conductivities of aqueous NaCl and KCl mixtures. My gratitude also extends to Prof. Carlos Vega, Jose Abascal, and Samuel Blazquez of Universidad Complutense de Madrid for their collaboration, providing me with my first experience in writing a scientific paper, on which our research groups collaborated. This work has successfully been published in July.[a]

I am indebted to Delft University of Technology for being a nurturing environment that has fostered my growth over the past five years. During my stay, I befriended Bauke Schenkelaars, Koen Schouten, Timo Stienstra, and Willemijn Drenth, with whom I wrote my bachelor thesis and still often play board games. Another student I am thankful to have met is Aneek Chakraborty, with whom I cooperated in many projects during my masters. I enjoyed scientific discussions with José Gonzalez Morales as well, whom I had the honor supervise daily during his master thesis on the design of non-flow solid hydrogen storage for long-distance maritime applications.

My appreciation also goes to the DelftBlue supercomputer[b], which provided me with computational resources. Dennis Palagin of the DelftBlue support team assisted in compiling the software I needed for this work. I also feel thankful that this work will not be my final story at the TU Delft. Prof. Dr. Thijs Vlugt and Dr. Othonas Moultos have accepted me as a future PhD candidate, adding four more years to my stay in this institution.

Lastly, I would like to express my heartfelt gratitude to my family for their support, especially during the challenging times I faced when studying Applied Physics at the Eindhoven University of Technology. Their belief in me, along with my wife's steadfast encouragement, patience, and her meticulous editing of this thesis, have been pivotal in bringing this work to fruition. Merel's unwavering faith in my abilities, coupled with the countless hours she spent reviewing and refining the English in this thesis, have truly made this achievement a success.

Vladimir Jelle Lagerweij
Delft, August 2023

[a] Blazquez, S.; Abascal, J. L. F.; Lagerweij, J.; Habibi, P; Dey, P.; Vlugt, T. J. H.; Moultos, O. A.; Vega, C. "Computation of Electrical Conductivities of Aqueous Electrolyte Solutions: Two Surfaces, One Property". *Journal of Chemical Theory and Computation* **2023**.

[b] Delft High Performance Computing Center (DHPC) *DelftBlue Supercomputer (phase 1)*, 2022.

Abstract

In this work, the added value of machine learning (ML) molecular force fields (FF) for the community of molecular simulations is showcased by successfully calculating transport properties of aqueous potassium hydroxide (KOH (aq)). Classical FFs use relatively simple interatomic potentials to simulate the nano scale. These simulations can predict macroscopic properties, such as density, heat of evaporation, viscosity, and self-diffusivity of the modeled materials. However, these FFs struggle to model materials in which more complicated interactions are relevant for the macroscopic behavior. Examples of such interactions are three-body interactions and chemical reactions. Quantum scale simulation methods are able to compute properties of materials in which these challenging interactions occur, although these methods are limited in length and time scales that can be modeled with realistic computational costs. Transport properties, such as viscosity, self-diffusivity and electric conductivity need these larger length and time scales to be determined accurately. ML can be used for a multi scale approach, bridging the gap between the quantum and the nano scale by training coefficients of general interatomic potentials. This provides the possibility of reaching the time and length scales of traditional molecular simulations with the accuracy of quantum mechanic models. KOH (aq) is selected to highlight the prospects of these multi scale techniques, as the self-diffusion of OH^- in this electrolyte is dominated by proton transfer reactions, which has not been modeled successfully with classical FFs.

Results of structure properties produced with ab initio molecular dynamics (AIMD, at quantum scale) simulations are compared with machine learning molecular dynamics (MLMD, at multi scale) simulations. There are no significant differences in the calculated shortest typical atomic distances and coordination numbers for both KOH (aq) and pure water systems. The determined transport properties are in the same order of magnitude as experimental results, although the calculated viscosity is overestimated and the self-diffusion of H_2O and K^+ are underestimated. This is because the system is simulated at a higher than experimental density and hydrogen bonding is overestimated with the selected quantum mechanics model. The proton transfer reactions are captured in the MLMD simulations, calculating the enhanced self-diffusion of OH^- to be $(6 \pm 2) \times 10^{-9}\,m^2\,s^{-1}$, which matches experimental results at infinite dilution.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
AIMD	Ab initio molecular dynamics
(aq)	Species are dissolved in water
Brick-CFCMC	Brick Continuous Fractional Component MC software
D3	VdW correction for RPBE
DeePMD	Deep Potential Molecular Dynamics kit
DF	Density functional
DFT	Density functional theorem
EoS	Equation of State
FF	Force field
FP	First principles
GGA	Generalized gradient approximation functionals family
GROMACS	GROningen MAchine for Chemical Simulation MD software
HB	Hydrogen bonding
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator MD software
LJ	Lennard Jones potential
MC	Monte Carlo
MD	Molecular Dynamics
ML	Machine Learning
MLFF	Machine Learning Force Field
MLIP	Machine-Learning Interatomic Potentials package
MLMD	Machine Learning Molecular Dynamics
MSD	Mean Squared Displacement
NPT	Ensemble with fixed number of particles, pressure and, temperature
NVT	Ensemble with fixed number of particles, volume and, temperature
RDF	Radial Distribution Function
RPBE	Revised Perdew-Burke-Ernzerhof exchange correlation functional
rVV10	VdW correction for r2SCAN
r2SCAN	The second order version of the Strongly Constrained and Appropriately Normed semilocal density functional
VASP	Vienna Ab-Initio Simulation Package
VdW	Van der Waals

Symbols

Symbol	Definition	Unit
D_s	Self-diffusion coefficient	$\mathrm{m^2\,s^{-1}}$
E_kin	Kinetic energy	eV
E_pot	Potential energy	eV
E_tot	Total energy	eV
$g(r)$	The radial distribution function	-
h	Enthalpy	$\mathrm{J\,mol^{-1}}$
k_B	Boltzmann constant	$\mathrm{m^2\,kg\,s^{-2}\,K^{-1}}$
L	Simulation box length	m
n	Coordination number	-
q	Charge	C
r	Radius	m
T	Temperature	K
t	Time	s
S	Solubility	$\mathrm{mol\,kg^{-1}}$
V	Volume	$\mathrm{m^3}$
γ	Activity coefficient	-
$\Delta_\mathrm{e}h$	Enthalpy of evaporation	$\mathrm{J\,mol^{-1}}$
ΔG_hyd	Hydration free energy	$\mathrm{J\,mol^{-1}}$
ε	Dielectric constant	-
Θ	Angle	°
κ_T	Isothermal compressibility	$\mathrm{Pa^{-1}}$
ν	Viscosity	Pa s
ρ	Density	$\mathrm{kg\,m^{-3}}$
σ_e	Electric conductivity	$\mathrm{S\,m^{-1}}$
σ_NE	Electric conductivity from Nernst Einstein relation	$\mathrm{S\,m^{-1}}$
σ_s	Surface tension coefficient	$\mathrm{J\,m^{-2}}$

1

Introduction

Understanding the behavior of electrolytes is critical for human society. Electrolytes are compounds conducting electricity through ion migration instead of electron transfer.[1] Generally, electrolytes consist of salts or acids dissolved in a liquid solvent such as water. Seawater is an electrolyte by this definition,[2] and is of great importance for the understanding of the inner workings of life.[3] From a human health studies perspectives, electrolytes are used in production of medicine and influence the human body, for example, by formation of kidney stones.[4] Unfortunately, fresh water is not available everywhere on earth. Over 300 million people worldwide depend on desalination processes for their fresh water production.[5] Desalinating seawater is highly energy intensive.[6,7] Improving desalination methods can lead to increased accessibility of clean water for people, as prices go down. Similarly to clean water, food production also depends on electrolytes, which is demonstrated through their use in the production methods of fertilizers. Attempts are made to improve these processes and prepare it for a sustainable, circular economy by using electrochemistry and aqueous potassium hydroxide (KOH (aq)).[8] Electrolytes are used in industrial separation techniques as well, for example, KOH (aq) is used to recover vanadium from spent catalysts[9] or to remove heavy metal pollutants from surface water.[10] To achieve the Paris climate agreement goals,[11] energy storage solutions have to be improved further. Most battery designs use electrolytes, from traditional lead-acid[12] and nickel batteries[13] to newly emerging technologies like aqueous batteries[14–16] and redox flow batteries.[17,18] Besides batteries, another possible energy storage medium is hydrogen (H_2). Alkaline electrolysers used for H_2 production typically depend on KOH (aq) for electric conductivity.[19,20] To reduce the losses during electrolysis, the electric conductivity of the electrolyte should be maximised. This can be reached when the behavior of KOH (aq) in bulk and confined systems[21,22] is better understood.

The macroscopic properties of electrolytes play a pivotal role as key indicators, guiding the evaluation and comparison of various potential engineering designs within the aforementioned applications. Of great interests are thermodynamic properties, such as density (ρ),[23] enthalpy (h),[24] solubility (S),[25] surface tension (σ_s),[26] activity (γ)[27] and transport properties, such as viscosity (ν),[28] self-diffusivity (D_s),[29] and electric conductivity (σ_e).[30] Traditionally, the acquisition of these properties requires various experimental techniques, an expensive process, especially when considering the exhaustive array of species, concentrations, pressures, and temperatures. Analytical and numerical methodologies offer a pragmatic avenue to curtail these costs, through preliminary screening of promising candidates[31] or by accurately predicting all aforementioned properties.

Equation of state (EoS) modeling is such an approach, where the behavior of the material is expressed in analytical equations. These equations can be tested, improved and fitted by using less extensive experimental databases, examples include the Soave-Redlich-Kwong (SRK),[32] Peng-Robinson (PR),[33] cubic-plus-association (CPA)[34] and the statistical associating fluid theory (SAFT).[35] Employing these existing models presents an advantage due to their ease of use, rendering them highly compatible with engineering applications. Despite considerable research invested in these models, they encounter challenges when it comes to accurately predicting electrolyte behavior. This complication arises from the non-ideal nature of these mixtures. In EoS modeling, deviation from ideality can be expressed using

activity coefficients, where deviations from 1 express non-ideality. The activity coefficient of sodium chloride dissolved in water (NaCl (aq)) at a molality of 1 mol kg^{-1} at standard condition is 0.667.[36] On a molecular level, water has a structure which greatly depends on the dipole moments of the individual molecules. Dissolving ions into water adds strong coulombic interactions, disrupting this structure. Comparing the molecular structure of pure water with that of NaCl (aq) shows large changes in the surroundings of individually dissolved ions.[37,38] Well-performing models should therefore be able to predict changes in macroscopic properties based on insights in changes in the nano scale. A fundamental breakthrough in electrolyte modeling is the Debye-Hückel theory, which provided a tool to describe electrolytes at low ionic concentrations.[39] This tool can be extended into higher concentrations by adding Born's ion solvation term.[40] Modern EoS for electrolytes are specifically adjusted versions of the above introduced basic EoS which have additional terms, typically inspired by Born, Debye, and Hückel. Examples of these are e-PR,[41] e-CPA[42] and ePC-SAFT.[43] Continued research efforts are dedicated to refining these methods,[44,45] often by adding more terms to these models. As these models heavily rely on adjustable parameters, they are dependent on accurate experimental results. It also raises the question if certain models perform well based on understanding of physics or just regression of many variables.

A third method describing thermodynamic interactions is molecular simulations. As explained above, changes in the nano scale domain are the cause of the macroscopic properties. Therefore, a method that investigates this nano scale behavior and translate that to macroscopic properties is sequacious. This nano scale analysis not only makes it possible to compute macroscopic properties, but also provides the possibility to understand the underlying physics. In electrolyte systems, well-described molecular simulations make it possible to evaluate the structures of atoms directly. Hydrogen bonding (HB) can be recognized, as well as the hydration of dissolved species and, depending on the technique used, the motion of individual species. This nano scale perspective makes molecular simulations not just an alternative to EoS modeling, it provides EoS researchers with insights on how to implement physics better in their models[46,47] or how to fit their thermodynamic relation parameters.[48] Another major difference between EoS modeling and molecular simulations is in their availability. EoS are generally oriented at industry applications and therefore come with licensing fees, whereas fast and highly parallelized molecular simulation packages such as LAMMPS,[49–51] GROMACS,[52] and Brick-CFCMC[53,54] are aimed at research applications and, therefore, open-source for anyone willing to spend the time learning how to use these tools.

Two main methods of molecular simulations exist, Monte Carlo (MC) and Molecular Dynamic (MD) methods. MC simulations use energy potentials to accept or decline proposed changes to the system. This method is typically used for vapor liquid equilibria (VLE) and other time-independent properties. MD simulations use the gradient of the same energy potentials to compute intermolecular forces, which are then integrated in time. These energy potentials are typically referred to as force fields (FF), because of the relation between forces and interaction energies. Unlike MC, MD is suitable for time-dependent phenomena, such as transport properties and crystal growth rates. Most mixtures can be described using three classical interaction types; the (i) strong short-range repulsive Pauli potential guarantees that atoms will not overlap in space, a (ii) weaker mid-range attractive Van der Waals (VdW) potential, and the (iii) long-range coulombic interaction that model charged species and dipole moments.[55] The Pauli and VdW interactions are typically combined into the Lennard-Jones potential (LJ).[56]

Many molecular simulation FFs for aqueous electrolytes have been developed. These are generally created by adding specific ion interactions to well-known models for pure water. An often used FF is the three-site extended simple point charge (SPC/E)[57] water model combined with Joung-Cheatham (JC)[58] or Kirkwood-Buff (KB)[59] ions. JC uses a rigid water molecule with positive charges on the hydrogen atoms and a negative charge on the oxygen atom. The properties of this and further mentioned water models are summarized in table 1.1. FFs are generally obtained by fitting their parameters to predict specific properties, of which experimental results are available. SPC/E is parameterized to reproduce experimental density and the potential energy while the JC ion FF is fitted using hydration free energy, lattice energies and constants, whereas the KB FF is fitted using the Kirkwood-Buff integrals. These models use full-charge representation ions, for example Na^+ has a charge of $1\,e$ and SO_4^{2-} is modeled using a charge of $-2\,e$. The transferable intermolecular potential 4P (TIP4P/2005)[60] is a four-site model. It is rigid with charges on the hydrogen atoms, similar to the SPC/E model. However, the charge of the

oxygen atom is modeled at its own location (the fourth site) in plane with the oxygen and hydrogen atoms and on the bisector of the H-O-H angle. TIP4P/2005 is parameterized using temperature of maximum density, isothermal compressibility, diffusivity of water and other properties. TIP4P/2005 can be combined with Madrid-2019,[61,62] Madrid-Transport[63] as well as the DFF/OH^- FF for OH^- ions.[29] The Madrid-2019 FF used solution densities, radial distribution functions (RDF) as well as hydration numbers of the ions for fitting. The Madrid-Transport and DFF/OH^- FFs differs from Madrid-2019 as charges are scaled from 0.85 to 0.75 and transport properties, such as diffusion and viscosity are more rigorously included in the parameterization.

The aforementioned FFs do not describe all relevant characteristics of water well. This is because two important physics principles are not included in these models, of which the first is the polarizability of water.[64] It has been shown that non-polarizable FFs for water struggle to model the vapor pressure, dielectric constants, virial coefficients, and the critical properties of water accurately.[65] At the expense of more degrees of freedom, and therefore with 3 to 10 times higher computational cost,[66] polarizable FFs for water have been developed. They allow at least the charge at the oxygen atom in the water molecule to change its location compared to its own molecule, although more complex models implement more degrees of freedom. A well-known polarizable FF is the Simple four-site Water model with Drude Polarizability (SWM4-DP).[67] This water FF has a TIP4P-like structure, but an additional negatively charged Drude particle is added, connected with a spring to the now opposite charged oxygen atom. This FF is fitted to enthalpy of evaporation, density and self-diffusion coefficients under ambient conditions. An alternative is the Baranyai-Kiss 3 FF (BK3),[68] which models all charges as harmonic spring oscillators connected to their basic sites. Another noticeable difference is that BK3 uses Gaussian charge distributions instead of standard point charges. Baranyai and Kiss attempted to parameterize their FF using many properties in a systematic approach. They included density-temperature relations, dielectric constants in liquid and hexagonal ice phases, as well as viscosity and self-diffusion coefficients in liquid phase. Force fields for alkali and halite (AH) ions have been developed for both water models, the AH/SWM4-DP[69] and the AH/BK3.[70] Although both models differ, both of them are parameterized using the hydration free energy of the ions within different hydrate structures. Comparing the performance of AH/BK3 and AH/SWM4-DP shows that AH/SWM4-DP has unsatisfactory performance for higher salt concentrations, at which crystallization occurs.[71] In the same communication, it is shown that the AH/BK3 FF predicts the activity coefficients, mean molar chemical potential and osmotic pressure of NaCl(aq) well for a wide range of concentrations. In other research, it is found that AH/BK3 predicts transport properties better than unscaled charge models, such as SPC/E with JC or KB, and similar or marginally better compared to scaled charge models, like TIP4P/2005 with Madrid-2019 or Madrid-Transport.[72] Models such as AH/SWM4-DP and AH/BK3 have many variables that are fitted, making them more dependent on the properties and conditions for which they are fitted. This can be demonstrated, as the AH/BK3 FF predictions for $\Delta_e h$ and ν deviate over 10 % for higher temperatures.[66,73,74]

The second relevant physics principle that is not included in classical FFs is that of three-body interactions. Literature shows that three-body interactions are important in liquid phase water.[75–77] Around 80 % of the total interaction energy is from two-body interactions and 20 % of the total energy is from three-body interactions. Higher order, many-body interactions contribute up to only 1 % and can generally be neglected.[76,77] Therefore, explicit three-body interactions are needed for detailed descriptions of water behaviour and its structures.[78,79] As the local structure of water undergoes a significant change when ions are added, especially the HB structure, it can be hypothesized that these three-body interactions should be included when simulating aqueous electrolytes. There are non-polarizable FFs, such as the explicit three-body potentials (E3B1, E3B2, E3B3) proposed by Skinners group[80–82] and polarizable FFs, such as the hydrogen-bonding polarizable (HBP)[66] water model with explicit three-body terms. Although a specific E3B2/MP-S ion FF exists,[83] and it is proposed that the initial Madrid FF, created for TIP4P/2005, should work with the E3B3 water FF due to the similarities in the water models,[84] little progress has been made in modeling electrolytes with three-body potentials using the traditional parameterization approach. Similar to EoS modeling, more complex FFs for water have to fit many parameters, making it difficult to distinguish between improved performance due to accurate descriptions of atomic interactions or more freedom in regression. The traditional parameterization approach of fitting FFs to macroscopic properties is also somewhat contradicting, as the strong suit of molecular simulations is the insights they provide in nano scale behavior.

Table 1.1: Properties of FFs for pure water and the ion models they can be combined with. Insights in to which properties are used to parameterize the FFs are provided, as well as which salt model is compatible with the indicated water models. HB indicates if the FF has specific three-body hydrogen bonding terms. For reference, the indicated properties are the; lattice constant (LC), lattice energy (LE), hydration number (n), potential energy (E_{pot}), self-diffusion coefficient (D_{s}), radial distribution function ($g(r)$), enthalpy of evaporation ($\Delta_{\text{e}}h$), hydration free energy (ΔG_{hyd}), dielectric constant (ε), isothermal compressibility (κ_T), viscosity (ν) and density (ρ).

Water FF		Sites	Polar-izible	Charge	HB	Fitted to	Ion FF		Salts	Charge scale	Fitted to
SPC/E	[57]	3	no	point	no	E_{pot}, ρ	JC	[58]	Li, Na, K, Rb, Cs, F, Cl, Br, I	1	LC, LE, ΔG_{hyd}
							KB	[59]	Na, Cl	1	KB-integrals
TIP4P/2005	[60]	4	no	point	no	ε, κ_T, ρ	Madrid-2019	[61, 62]	Li, Na, K, Rb, Cs, Mg, Ca, F, Cl, Br, I, SO_4	0.85	n, $g(r)$, ρ
							Madrid-T	[63]	Na, K, Cl	0.75	$g(r)$, ν, ρ
							DFF/OH^-	[29]	Na, K, Cl	0.75	$g(r)$, ν, ρ
SWM4-DP	[67]	4	yes	point	no	D_{s}, $\Delta_{\text{e}}h$, ρ	AH	[69]	Li, Na, K, Rb, Cs, F, Cl, Br, I	1	ΔG_{hyd}, ε
BK3	[68]	4	yes	Gauss-ian	no	D_{s}, ε, ν, ρ	AH	[70]	Li, Na, K, Rb, Cs, F, Cl, Br, I	1	ΔG_{hyd}, ρ
E3B	[80–82]	4	no	point	yes	D_{s}, $g(r)$, ρ	MP-S	[83]	Li, Na, K, Rb, Cs, F, Cl, Br, I	0.75	ε
HBP	[66]	4	yes	Gauss-ian	yes	$g(r)$, ε, ρ	-				

Instead of seeking information for nano scale interactions from macro scale properties, it is possible to retrieve information from the lower quantum scale. It is generally accepted that the true description of atomic interaction is in quantum mechanics, -since it has full atomic information- which are often referred to as first principle (FP) calculations. Simulation techniques for quantum mechanic interactions are the steady state density functional theorem (DFT),[85] and its transient extension ab initio molecular dynamics (AIMD).[86] Alternatives to density functionals in AIMD exist, but aim to provide the same length scale information to compute the particle trajectories.[87] These methods solve the electronic structure of the entire system directly. This consequently contains all dipole moments and full many-body interactions. DFT, AIMD and related methods are also used to compute reaction characteristics, such as activation energy barriers on catalysts[88] or absorption properties of 2D materials.[89] Ideally, these methods should solve all interactions fully, however, they depend on the quality of selected density functionals (DF). An example of this is how the popular generalized gradient approximation (GGA) Perdew-Burke-Ernzerhof functional (PBE)[90] greatly underestimates the dispersive VdW interactions.[91] Specific (non-)local VdW DFs are proposed to improve on these issues[92,93] and some modern metaGGA DFs like the Strongly Constrained and Appropriately Normed semilocal density functional (SCAN)[94] intrinsically include short and medium range VdW interactions and have additional long range VdW extensions.[95] When describing aqueous water, AIMD has shown to accurately predict the structure of pure water.[96–98] Self-diffusion coefficients of pure water are typically underestimated,[99,100] even if they are corrected for finite size effects. This is caused by DFs that overestimate hydrogen bonding, making the system over-structured, as well as lack of statistics, due to system sizes and time scales. AIMD is able to calculate structure properties of electrolytes, successfully showing ion-water interactions and hydration structures around ions.[101,102] As AIMD has shown its capability in computing chemical reactions, it is also used to simulate proton transfer.[103,104] In aqueous electrolytes containing OH^-, like NaOH (aq) and KOH (aq), OH^- molecules will react with H_2O, greatly increasing the effective self-diffusion coefficient of OH^-.[29,103,104]

Although AIMD seems the ideal tool to compute electrolyte properties, it has its limitations. As AIMD is designed to solve quantum scale behavior, it is computationally expensive, limiting its system sizes up to hundreds of atoms and the simulation time scale to tens of ps. Transport properties of electrolytes need hundreds of ps simulation time scales and thousands of atoms, which is a typical time scale and system size of MD simulations. This creates a multi scale problem, where DFT and AIMD hold a better description of the interactions between atoms, but MD is the only method to resolve the appropriate time and length scales. Machine learning (ML) is a promising technique to solve multi scale problems,[105,106] since it can be used as a regression tool to transfer the information from the smaller scale to the larger scale. Example cases in other fields show how ML can provide nanofluidics simulations with MD derived material properties[107] or how finite elements modeling can use microscopic information for non-linear material behavior.[108] Currently, attempts are made to create computationally cheap classical FFs based on AIMD. For example, by training the parameters in a neural network (NN).[109,110] This makes it possible to train the FF parameters directly on the atomic interactions compared to fitting the FFs to experimental macroscopic data. Besides fitting coefficients for traditional FFs, more complicated interactions can be added as well. The ReaxFF water FF fits coefficients for other empirically determined interaction types as well, including energy terms for hydrogen bonds, over- and undercoordination, and lone pairs.[111–114] This model is able to capture reactive events in water, such as the aforementioned proton transfer, and produced self-diffusion coefficients of OH^- close to experimental values.[111,112] Using empirical atomic interactions can introduce errors, as the real interaction might not be expressed correctly. Alternatively, ML can be used on general two- and three-body interaction functions that then directly accelerate AIMD simulations into the time and length scales of molecular dynamics.[74,115–118]

Some software tools have ready-to-use ML implementations for such general two- and three-body potential. For this work, three of such software tools are of importance; the Vienna Ab-Initio Simulation Package (VASP),[119–124] the Machine-Learning Interatomic Potentials package (MLIP),[125] and the Deep Potential Molecular Dynamics kit (DeePMD).[126] VASP is a proprietary tool that performs DFT simulations, as well as AIMD. From version 6.3.0, VASP has implemented an ML algorithm, visualized in figure 1.1, which speeds up AIMD on-the-fly by using a Gaussian or polynomial kernel function approach. During the simulation, VASP estimates the error of the already trained FF. If the error in a certain configuration exceeds a set limit, it will perform a traditional DFT calculation and adjust the MLFF directly. The kernel function method selected can be adjusted with relatively little computational expense, unlike an NN, enabling an adaptable approach. The VASP ML has shown great potential by predicting phase transitions of hybrid perovskites,[124] computing the melting temperatures of Al, Si, Ge, Sn and MgO,[123] and chemical potentials of dissolved Li, Si and F ions in water.[127] This last study is especially relevant, as it shows the ability of computing properties of electrolytes which are outside of reach for traditional AIMD scales. Although the produced structures match that of FP calculations, it must be noted that these differed significantly from experimental results as they used a revised PBE density functional instead of a more accurate metaGGA or hybrid density functional.

MLIP uses a different machine learning approach,[125] with support for both linear and concurrent ML approaches, which are illustrated in figure 1.1. For the concurrent method, it uses DFT output of preselected configurations to train an MLFF based on moment tensor potentials (MTPs). This MLFF is then exported to perform MD simulations in LAMMPS. The LAMMPS implementation of MLIP also estimates how far away a configuration is from the DFT data every time step. If it estimates that the configuration is within the data, interpolation, or slightly outside the data, minor extrapolation, it will continue the simulation normally. If a little more extrapolation is estimated, it will store the specific configuration separately and continue. Even more extrapolation results in canceling the LAMMPS simulation. The separately stored configurations are then inserted in VASP for DFT calculations and used to expand the dataset on which the MLFF is trained. The LAMMPS simulation is then restarted from scratch. This process repeats itself until the simulation in LAMMPS is completed. This approach assures that the MLFF does not undergo any change during the MD simulation, unlike the on-the-fly ML in VASP. MLIP has been used to investigate solids, such as 2D nanoporous carbon nitrides[128] and BC_2N monolayers,[129] in time and length scales not achievable with traditional AIMD methods.

The third software tool, DeePMD,[126,130] uses a deep neural network to train its FF. Like MLIP,[125] DeePMD supports linear and concurrent training. Both the VASP and Quantum ESPRESSO[131] simulation packages are used to provide DFT data for DeePMD. DeePMD allows for training using multiple system sizes, atom species and concentrations, and needs a separate set of testing data. This can be used to evaluate the created FF after training. Afterwards, the FF can be used for MD in both LAMMPS and GROMACS. The DeePMD-kit has seen a rise in popularity, besides modeling solid or melting systems[132–136] such as VASP en MLIP, it has also been used to model water.[137,138] DeePMD also simulated more challenging systems, reactions such as the decomposition of urea in water[139] or alkali carbonate-hydroxide electrolytes.[140] Furthermore, the influence of ions on the structure of water is further investigated, showing that locally, the solvation of ions does not have the same effect as applying pressure,[141] although this was assumed to be the case.[142,143] Of special significance are attempts to introduce long-range electrostatic interactions into DeePMD as well.[144] Additionally, Malosso et al. [145] illustrated how viscosity can be derived using DeePMD for MLFFs.

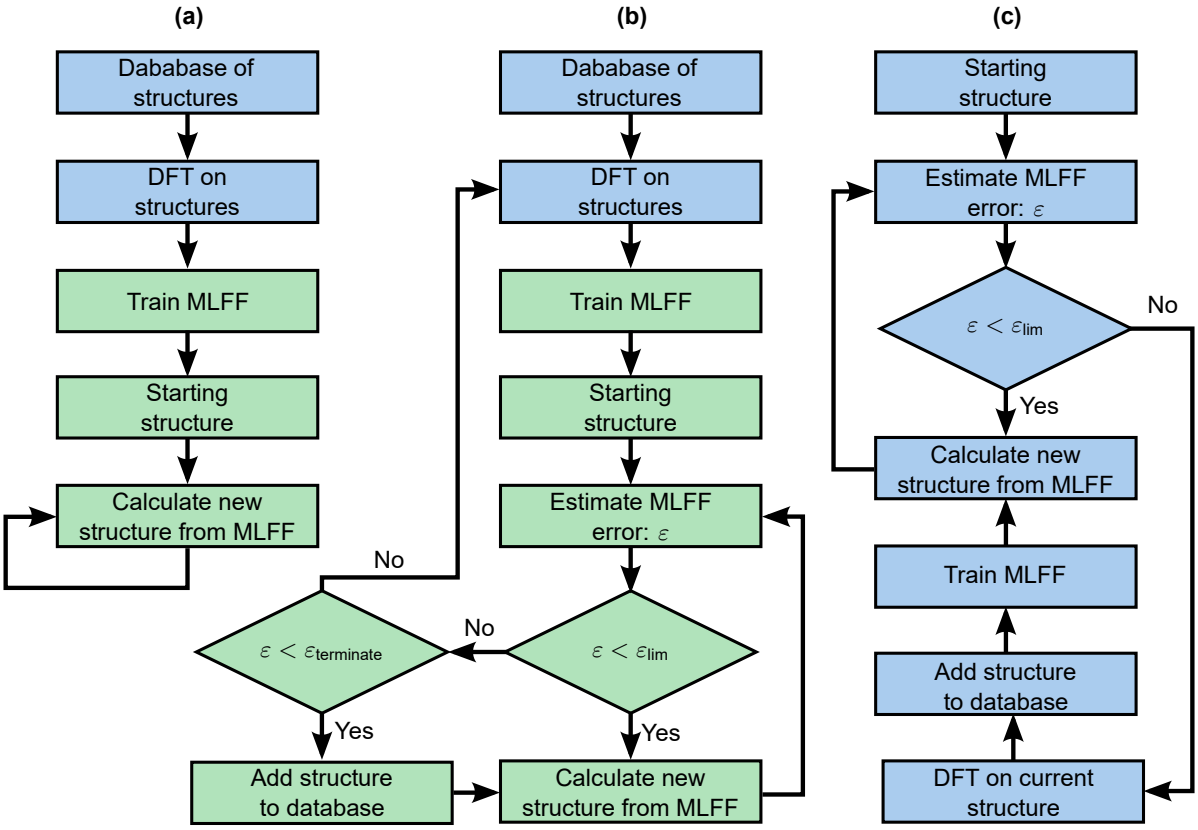

Figure 1.1: Three common algorithms for MLFF. (a) indicates the most simple linear method, (b) a concurrent algorithm, and (c) the on-the-fly algorithm. DeePMD[126,130] and MLIP[125] are both introduced using linear algorithms. However, they are improved to also support the concurrent algorithm (b). ML in VASP[123,124] uses the on-the-fly algorithm (c), and is selected for this study. The calculation steps in blue are executed in quantum mechanics software packages and the green steps are executed in traditional molecular simulation software. ε_{lim} and $\varepsilon_{terminate}$ are error thresholds, while ε is the estimated error of the MLFF on the current structure. The molecular trajectory calculations start from the starting structure and will be completed after the intended number of new structures are calculated.

It has been shown how current molecular simulation techniques of electrolytes are limited by their FFs. This is most apparent when complicated molecular interactions are modeled. Examples of these are three-body interactions and chemical reactions. Quantum scale simulation techniques, such as AIMD, can model these type of interactions accurately. However, these simulations are limited in time and length scales. To accurately compute transport properties, the time and length scales of traditional MD is needed. ML is an emerging technique[123,125,126] to automate the production of FFs. It provides a multi scale approach by transferring quantum scale properties to the nano scale. In this work, we showcase this multi scale approach by successfully calculating the transport properties of KOH (aq)

using the on-the-fly ML implemented in VASP.[119–124] To our knowledge, this has not been achieved with classical FFs. KOH (aq) is an excellent electrolyte to showcase the effectiveness of MLFFs, as both hydrogen bonding and proton transfer reactions influence the transport properties. The applied methods are explained in chapter 2. There the applied quantum scale modeling, ML approach and nano scale methods are all discussed. Both the validation and main results are discussed in chapter 3. Chapter 4 concludes this work and provides an outlook for future research opportunities and relevant improvements needed to get most out of ML for molecular simulations.

2

Methods

In this chapter, the theoretical background, as well as the simulation settings are explained. Separate considerations have been taken for all involved length scales, which are described in this chapter in order of scale. On the quantum scale, the inter-atomic forces are computed by using AIMD, as described in section 2.1. Machine Learning, elaborated on in section 2.2, is then used to connect the quantum scale to the nano scale. The methods used in the nano scale are shown in section 2.3. How the macroscopic system properties are extracted out of the performed simulations is described in section 2.4. An overview of the three system and simulation settings is shown in table 2.1. The simulations are executed in VASP 6.4.1,[119–124] using gamma-point only calculations, and post-processed using Python 3 code. The simulations are performed using the Delft-Blue[146] and Snellius super-computers.

2.1. Quantum Mechanical Calculations

The smallest length scale investigated in this work is the quantum scale. Atomic interactions are modeled by solving the electron distribution. In this work, the Kohn-Sham density functional theory (KS-DFT)[85,147] is used. This method for solving quantum mechanics uses density functionals (DF) to describe the (self)interactions of atomic species. In a way, DFs serve an equivalent purpose in KS-DFT as FFs in molecular simulations. Similarly, an appropriate DF has to be selected for the specific case. This section also explains how medium and long range dispersion interactions are added to the selected DF, as well as how the computational effort is minimized and grid independence is ensured.

2.1.1. Functional Selection

DFs range from purely local descriptions to more modern fully non-local descriptions. The Jacob's ladder[148] classifies five generations of DFs. The first step of this ladder is the Local Density Approximation (LDA), followed by the general gradient approximation (GGA), which is improved to the meta-generalized gradient approximation (metaGGA) method. The highest two steps are the hybrid and fully non-local DFs. The more complex DF generations trade higher accuracy with increased computational effort. Specific DFs are fitted to meet (semi)empirical limits, such as the local Lieb-Oxford bound.[149] Meeting all of these is impossible, as there are more limits than LDA and (meta)GGA methods can satisfy analytically. Therefore, the creators of DFs for these methods choose which limits to satisfy and which not.

In this work, the revised Perdew-Burke-Emzerhof (RPBE)[90,150] GGA and the second order version of the strongly constrained and appropriately normed (r2SCAN)[94,151] metaGGA DFs are selected. RPBE is a well-used functional. Thus, it is used to compare basic properties with existing research data. Compared to RPBE, r2SCAN is a recently proposed DF. The SCAN functional uses fourth order differentials to assess non-local properties based on local data.[94] This makes SCAN computationally expensive, even for metaGGAs. This is reduced to second order differentials in r2SCAN, reducing the computational cost significantly with none to little loss in accuracy.[151] Using both DFs add the possibility to assess the accuracy differences between the traditional GGA and modern metaGGA DFs besides the machine learning of FFs, which represents thirty years of scientific progress.

2.1.2. Medium and Long Range Dispersion Interactions

Although (meta)GGA DFs provide accurate descriptions of short range, as well as long range coulombic interactions, they typically underestimate medium and long range dispersion forces.[91,152–155] For these so-called Van der Waals (VdW) bonds, there are specific corrections available for most DFs. For the GGA DFs, the solution is mostly sought in a molecular simulation FF way, where VdW interaction energies are mostly expressed as $E_{\text{VdW}} \backsim r^{-6}$, r being the distance between two particles. Grimme and coworkers developed a purely empirical energy term, with the same scaling, which is added to the DF energy.[156] Most challenging for this is the medium range overlap regime, as most DFs show some VdW interactions, however, too little and only for shorter ranges. This can lead to double-counted medium-ranged interactions. A damping function is used to reduce added energies for shorter ranges and needs to be adjusted for the specific DF. Multiple generations of this DFT-D approach exist; the DFT-D1[156] method, DFT-D2,[157] with only refitted parameters, DFT-D3,[158,159] where fundamental improvements are proposed, and DFT-D4.[160,161] For this work, the RPBE DF with DFT-D3 (RPBE-D3) is selected, as this was the most modern correction which is implemented in VASP 6.4.1.[119–123]

Instead of using an empirical scaling based on atomic positions, VdW interactions can also be calculated directly from electronic effects and replace parts of the traditional DFs. These non-local VdW DFs have increased computational costs, but naturally include medium and even long-range VdW interaction energies. The VV10[162] and rVV10[163–165] VdW DFs are designed flexibly, so that they can be combined with other (meta)GGA DFs to model all relevant (covalent, ionic, metallic and VdW) bonds. These adjustments are also made for SCAN[95] and r2SCAN.[166] It was shown that rVV10 (r2)SCAN greatly improves the bonding between molecules,[167,168] and that rVV10 SCAN suffers from double counting errors, therefore significantly overestimates water density.[166] The rVV10 r2SCAN functional, as is used in this study, has been reported as a more sensible VdW-corrected DF.[168]

2.1.3. Minimizing Computational Effort in Solving DFT Calculations

Modeling in this scale is computationally expensive, so there needs to be a between calculating these interactions accurately and being efficient with the computational resources. This balance depends on the species and the DFs implemented, a convergence study is done to find the minimal values for the number of k-points, the energy cutoff for the plane wave basis set, and the Gaussian smearing width.

To check for convergence, an initial configuration of 64 H_2O molecules at 1000 kg m^{-3} is created first using PACKMOL 20.3.1.[169] An AIMD simulation of 5 ps at a temperature of 325 K is performed with 1 k-point, a Gaussian smearing width of 0.1 eV and an energy cutoff of 1000 eV. The output of this simulation represents a typical configuration. For this configuration, the electron distribution is solved using different numbers of k-points, energy cutoffs and smearing widths. The deviation in energy between certain settings and a very strictly-set reference solution is expressed in meV per atom. As this work intents to show the chemical reactions in KOH (aq) and how MLFFs can accelerate AIMD into MD time scales, there is no need in having overly strict criteria for these settings, making a deviation of 15 meV per atom in the system acceptable. It was found that the smearing width did not influence computation time and accuracy much, and it is set to 0.3 eV. Figures 2.1 and 2.2 indicate the results of investigating the behavior as a function of the number of k-points and the energy cutoff. The deviation from the reference simulation clearly reduces when the energy cutoff is increased, as shown in figure 2.1. Figure 2.2 visualizes clearly how this is traded for larger calculation time per time step. Additionally, it is clear that increasing the number of k-points does not improve the deviations much, while greatly increasing computation time. These figures illustrate that a mesh with 1 k-point and an energy cutoff of 550 eV will lead to a deviation just under 15 meV per atom.

2.2. Machine Learning

The goal of this work is twofold, investigating the effects that proton transfer has on self-diffusion and electric conductivity, as well as to illustrate how machine learning (ML) can accelerate AIMD calculations into time scales in which transport properties can be computed. Simulations with AIMD and MLMD with otherwise identical settings have been performed. Gaussian kernel functions for ML are implemented in VASP[123,124] which, on-the-fly, create a two and three-body FF. The background theory and how it differs from AIMD is explained, followed by the specific setup and settings selected for this work.

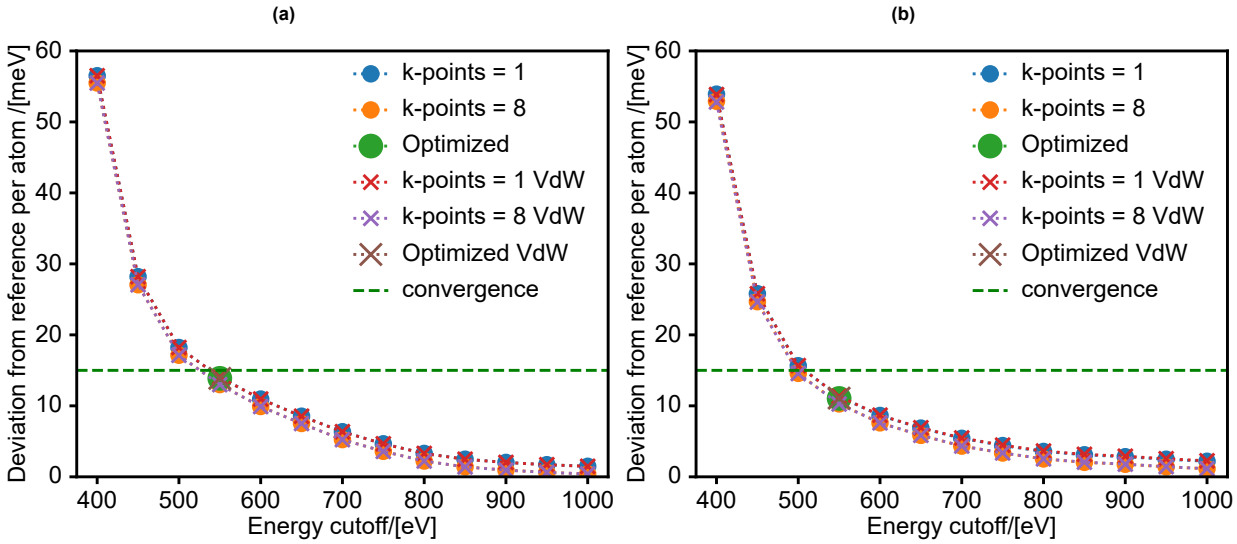

Figure 2.1: Minimizing the computational efforts for RPBE(-D3)[150,158,159] and (rVV10) r2SCAN[151,163,164,166] DFs are shown in (a) and (b), respectively. The absolute deviation of the computed total system energy from a reference calculation is shown for both DFs with and without their VdW correction. The system consist of 64 H_2O molecules at 1000 kg m^{-3} in a, with AIMD prepared, configuration. Different energy cutoff values and number of k-points are set with a smearing width of 0.3 eV. The reference is simulated using an energy cutoff of 1800 eV, 729 k-points, and a smearing width of 0.01 eV. The set maximum deviation of 15 meV is indicated as a horizontal, dashed green line. The final, optimized settings are indicated separately in both figures.

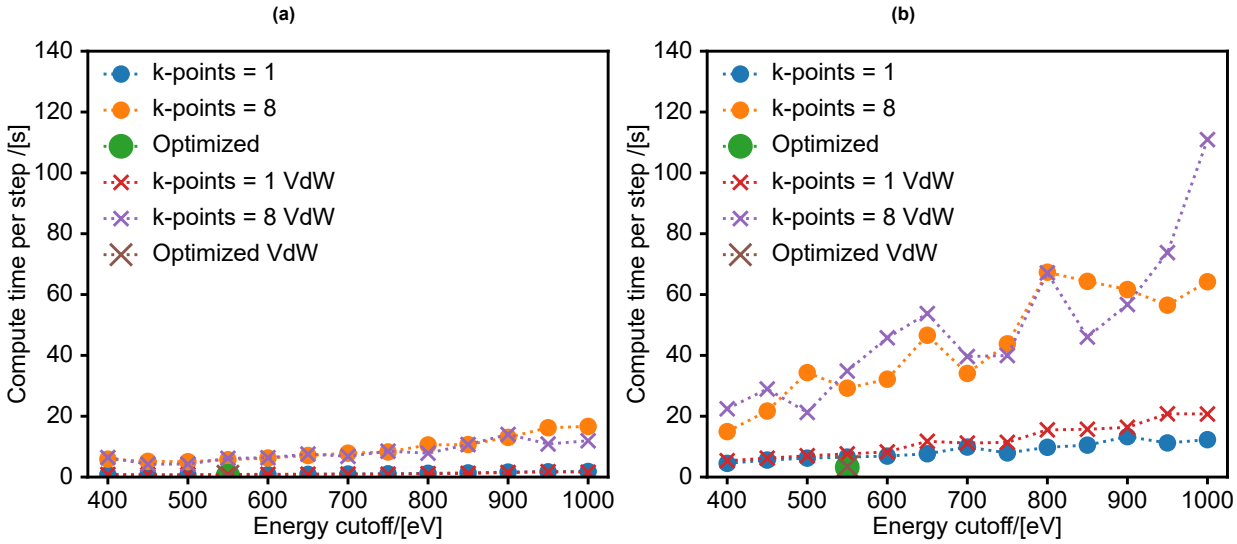

Figure 2.2: The effects of the number of k-points and the energy cutoff on the computation time per self-consistency iteration of RPBE(-D3)[150,158,159] and (rVV10) r2SCAN[151,163,164,166] DFs are shown in (a) and (b), respectively. The system consist of 64 H_2O molecules at 1000 kg m^{-3} in a, with AIMD prepared, configuration. The RPBE-D3 DF is clearly cheaper to compute than the rVV10 r2SCAN DF. As the rVV10 r2SCAN DF typically needs more self-consistency iterations to finalize the DFT calculation, computation time differences are even larger than shown. The final, optimized settings are indicated separately in both figures.

2.2.1. Comparing the AIMD and MLMD Algorithms

Some explanation on how AIMD works is needed before going into the details of MLMD. In AIMD, a DFT self-consistency loop is executed every timestep. The algorithms of AIMD and DFT are shown in figure 2.3a and figure 2.3b, respectively. The self-consistency loop returns the system energies and the electron density, which is used to compute the forces on all atoms. Executing this DFT self-consistency loop is computationally expensive. Time integration with a Verlet integration scheme[170] is used to compute the new atom positions, after which this process is repeated.

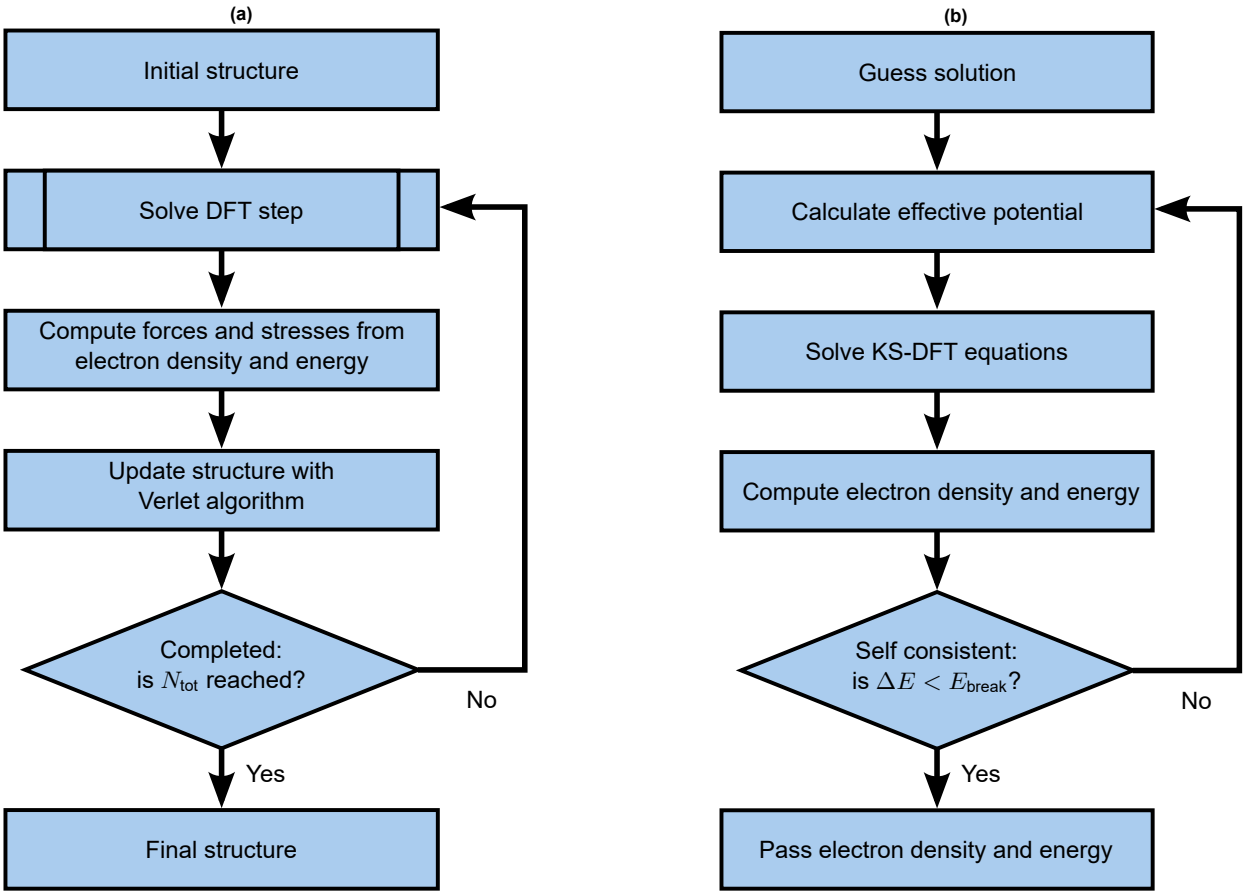

Figure 2.3: (a) Algorithms for AIMD and (b) DFT self-consistency calculations. Algorithm (b) is nested inside the algorithm (a) as a subroutine in the 'Solve DFT step'. In this subroutine, the solution of the last iteration is used as improved starting point. The AIMD loop (a) is to take N_{tot} time steps, ΔE and E_{break} in (b) refer to the energy difference between the current and last iteration and the convergence limit, respectively.

The ML algorithm in VASP aims to replace the expensive DFT self-consistency loop with a molecular FF.[123,124,127,171] For this, three things are needed: (i) a general framework for MLFFs with free parameters which can be tweaked. It is important that this general FF can (ii) predict the inter atomic forces, system energies and stresses (pressure), as this needs to be reported even when DFT calculations are skipped. The FF has to be trainable, by tweaking these parameters based on a structure database. This is done by performing DFT calculations instead of using the FF when an estimated error of the MLFF exceeds a certain limit. Therefore, the method requires an (iii) error estimation technique as well. The overall algorithm is best explained in detail by Jinnouchi et al. [123].

VASP identifies four different separate definitions which are of importance: a (i) structure database, consisting of many solved DFT (ii) structures. A structure is purely described by all atomic positions in the system. The database holds the total energy, inter atomic forces, the system stresses, and the Bravais lattice of all the solved DFT structures as well. The structure which is not solved with DFT is henceforth referred to as the current structure. Each structure, whether it is in the database or the current state, is described in (iii) local configurations, which are expressed in linear combinations of the (iv) descriptors. The general framework describes a local configuration by using radial and angular descriptor functions indicated with $\boldsymbol{X}^{(2)}$ and $\boldsymbol{X}^{(3)}$, respectively. A visual representation of these descriptors is shown in figure 2.4. The descriptors consist of a set of orthogonal basis functions of which a linear combination can express the radial and angular atomic distribution functions $\rho^{(2)}(r)$ and $\rho^{(3)}(r,s,\theta)$. The distribution functions have a smoothed cutoff and the particle locations of the neighbors are not represented as hard Dirac delta functions, but are damped with a Gaussian distribution, the so-called smooth overlap of atomic positions (SOAP) approach.[123,171,172]

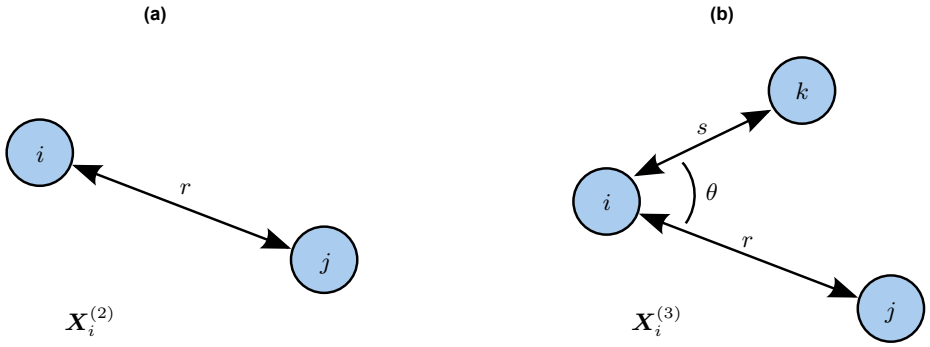

Figure 2.4: The (a) radial ($\boldsymbol{X}_i^{(2)}$) and (b) angular descriptors ($\boldsymbol{X}_i^{(3)}$) of a specific atom i used for MLFFs in VASP.[123] These descriptors are vectors containing the weights of an orthogonal set of basis functions. This basis set consist of Legendre polynomials. Together, these describe the local configuration around atom i.

The MLFF aims to create a potential energy surface depending on the distribution functions expressed in terms of the descriptors. To assess the total potential energy (U) of a current structure with N_A atoms by using an already trained MLFF with a reference structure consisting of N_B configurations, the following equation has to be solved:

$$U = \sum_{i=1}^{N_A} U_i = \sum_{i=1}^{N_A} F[\rho_i^{(2)}(r),\ \rho_i^{(3)}(r)] \tag{2.1}$$

$$= \sum_{i=1}^{N_A} \sum_{i_B=1}^{N_B} w_{i_B} \left[\beta \left(\boldsymbol{X}_i^{(2)} \cdot \boldsymbol{X}_{i_B}^{(2)} \right) + (1-\beta) \left(\hat{\boldsymbol{X}}_i^{(3)} \cdot \hat{\boldsymbol{X}}_{i_B}^{(3)} \right)^{\zeta^{(3)}} \right]. \tag{2.2}$$

Here, the energy contribution of a local configuration around atom i is expressed in terms of local two- and three-body configurations. The N_B local reference configurations are expressed in terms of the descriptors $\boldsymbol{X}_{i_B}^{(2)}$ and $\boldsymbol{X}_{i_B}^{(3)}$, with the three-body descriptors normalized to $\hat{\boldsymbol{X}}_{i_B}^{(3)}$. The variable β is a weighting factor between the two- and thee-body interaction and $\zeta^{(3)}$ controls sharpness. The potential energy belonging to the reference configurations i_B is w_{i_B}. The part inside the square brackets calculates the similarity of the configuration i and the specific reference i_B by using the dot product between a local configuration of the current structure $\boldsymbol{X}_i$ and a reference $\boldsymbol{X}_{i_B}$. Retrieving the energy of any current structure as shown in equation (2.2) only depends on having the correct values for all w_{i_B}, collected together in vector $\boldsymbol{w} \in \mathbb{R}^{N_B}$. Per structure of N_A atoms, $m = 1 + 3N_A + 6$ values have to be retrieved for the total potential energy, the inter-atomic forces per atom, and system stresses, respectively. As by their definition, inter-atomic forces and system-stresses depend on the potential energy surface. Therefore, a linear combination can be devised which expresses all these properties in $\boldsymbol{w}$. To retrieve these properties, the following vector equation is derived:

$$\boldsymbol{y} = \boldsymbol{\phi}\boldsymbol{w}, \tag{2.3}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{\phi} \in \mathbb{R}^{m \times N_B}$ are the structures output properties and the design matrix which describes the linear combination, respectively. Now, there is both a general framework for MLFFs which is tweakable by changing the N_B reference configurations and the energy weights in $\boldsymbol{w}$, as well as a method for retrieving the pressure and inter-atomic forces. What is left is a method of retrieving the values in $\boldsymbol{w}$ as function of a structure database and a method for estimating the errors, which is needed for the algorithm in figure 2.5. Suppose there are N_α structures in the database with each N_A atoms. Instead of equation (2.3), which is an expression for the properties of a single structure, a supervector-matrix equation can be derived with similar goals which expresses a total database of structures:

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{w}, \tag{2.4}$$

where $\boldsymbol{Y}$ and $\boldsymbol{\Phi}$ are a supervector and supermatrix in $\mathbb{R}^{N_\alpha \times m}$ and $\mathbb{R}^{N_\alpha \times m \times N_B}$, respectively, and $\boldsymbol{w}$ is still in $\mathbb{R}^{N_B}$. $\boldsymbol{Y}$ is the structure database and consist of N_α solved DFT structures and their computed properties in the format of $\boldsymbol{y}$. $\boldsymbol{\Phi}$ is a known standard matrix derived from the fundamental relations between the potential energy and inter-atomic forces and the system stresses. Solving equation (2.4) for $\boldsymbol{w}$ retrieves the free parameters of the MLFF. No direct solution for this exists, as the problem is overconstrained and small errors are expected in $\boldsymbol{Y}$ as the DFT self-consistency calculations are only solved up to certain accuracy. These errors are assumed to be Gaussian noise. Retrieving an error estimate of the MLFF is done by using these errors and the overconstrained property of the problem, resulting in a mean $\bar{\boldsymbol{w}}$ and an estimated error $\boldsymbol{\Sigma}$. This $\bar{\boldsymbol{w}}$ and $\boldsymbol{\Sigma}$ are used to assess a current structure and calculate the values (as in equation (2.3)) and errors for the energy, forces and system stresses. A total error value is calculated $\varepsilon_{\text{config}}$, which is used for the decision-making, shown in figure 2.5. The value for $\varepsilon_{\text{limit}}$ can be either fixed or adapt depending on the settings. Much older configurations can be intentionally forgotten or retained before retraining with a new configuration. For a more accurate and stable algorithm, the method calculates two different parameters in the error estimating scheme, which are optimized using the generalized maximum likelihood method[123,173] and the system is sparsified to avoid overcompleteness by using a similarity kernel and removing overly similar local configurations.

2.2.2. Machine Learning Settings for this Work

VASP provides the user with the option to tweak the ML method. The training algorithm is set quite strict. As for this work, the aim is not only to show how ML can train a simple FF, but also to show that MLFFs can capture more complicated interactions, such asmodel reaction events. This means that there can be some oversampling of the result, and some unneeded DFT self-consistency calculations. However, this increases the likelihood that the reaction events are also shown in the ML setting. For this, the strict `ML_ICRITERIA = 2` setting is used,[174] which adjusts the $\varepsilon_{\text{limit}}$ based on the past Bayesian errors with a moving average. This criterion is reduced during the simulation, which assures that long MLMD runs keep performing DFT self-consistency loops during special occasions, such as reactions. After every DFT self-consistency calculation, $\varepsilon_{\text{limit}}$ is relaxed temporarily, as not to include too similar structures into the structure database. The reduction speed of this criterion is a VASP setting as well. The appropriate setting for this has been found to be `ML_CX = 0.1`. This is determined by comparing RDFs of the pure water and KOH validation AIMD and MLMD simulations, of which the set properties are indicated in table 2.1. These results are shown in section 3.1 and section 3.4, respectively. The radial descriptor is set to use 12 basis functions and has a cutoff of half the box size. The angular descriptor has 8 basis functions using the same cutoff. The full input files with additional comments for both the AIMD and the MLMD runs are available on GitHub.[a]

2.3. Molecular Dynamics Simulations with VASP

In VASP, the standard time integration is done using the Verlet algorithm.[170] This algorithm is energy conserving and is used to model systems of a constant number of particles, volume, and energy (NVE). In practice, performing simulations with a constant number of particles, pressure and temperature (NPT) or a constant number of particles, volume, and temperature (NVT) are used to retrieve the correct density or simulate the properties at a given temperature, respectively. Setting up the NPT simulations to run stable with the Langevin thermostat[55,175] has proven to be difficult. Alternatively, NVT pre-production runs at varying densities are performed, indicated as KOH density in table 2.1. To not use too much computation time, these simulations are ran using MLMD. The production runs, in which the results are produced, are performed using NVT with the Nosé-Hoover thermostat.[176–178] Unlike the Langevin thermostat, the heat bath of a certain mass is included in the system.[55,175–178] The Nosé-Hoover system is deterministic, whereas the Langevin thermostat is stochastic. Nosé-Hoover also provides more natural temperature fluctuations. There is no Nosé-Hoover NPT ensemble implemented in VASP, although this does exist in other MD software, such as LAMMPS[49–51] or GROMACS.[52] The time integrator used for both thermostats in VASP is an adjusted Verlet algorithm.[175] The appropriate mass of the Nosé-Hoover heat bath is `SMASS = 5`, which is determined by comparing RDFs of pure water and RPBE-D3 at 1000 kg m^{-3} with those produced by Forster-Tonigold and Groß [179]. The results of this are presented in section 3.1.

[a] The hyperlink to the post-processing code on GitHub: `https://github.com/JelleLagerweij/Quantum_to_Transport`

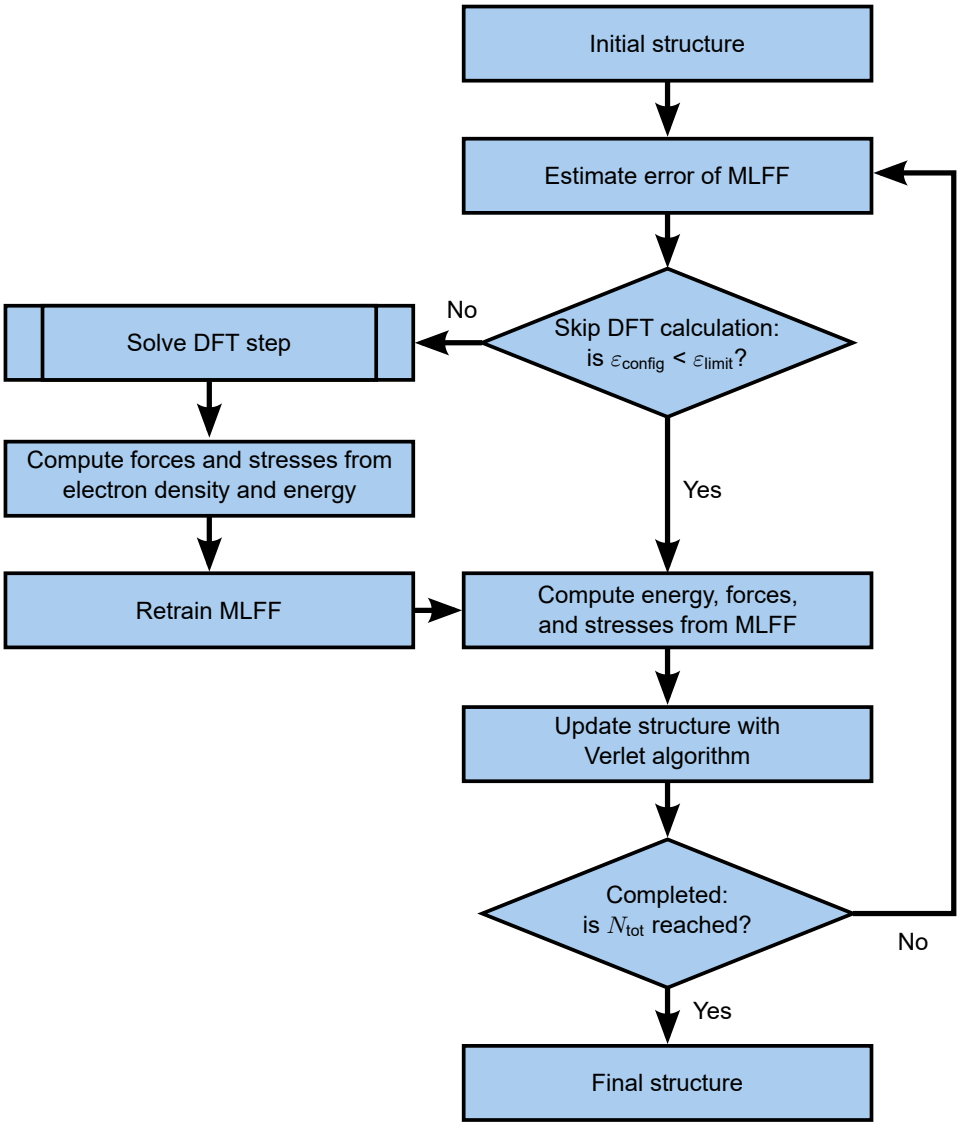

Figure 2.5: The MLMD algorithm integrated in VASP.[119–124] How the energy, forces, and stresses are calculated for a certain configuration is expressed in equations (2.2) and (2.3). The subroutine 'Solve DFT step' is shown in figure 2.3b, the algorithm aims to avoid this path if possible, as these DFT self-consistency calculations are computationally expensive. The MLMD loop is to take N_{tot} time steps, ε_{config} and ε_{limit} refer to the estimated error of the MLFF for the current structure and set tolerance for this error, respectively. If the error estimation exceeds the set tolerance, it will perform the DFT self-consistency calculation and retrain the MLFF.

2.4. Computing System Properties

Multiple system properties are computed from the trajectories, which are categorized into two groups. The first category of properties, presented in section 2.4.1, is used to verify the settings of both the AIMD and MLMD simulations. The second category are results which provide new insights. These can now be retrieved using recent gains in computing power and efficiency with MLFFs.

2.4.1. Calculating Energies and Structure Properties

To check the selected AIMD and MLMD settings in VASP, energies and structure properties are calculated. This is done using shorter simulations, henceforth referred to as KOH validation simulations, of which relevant properties are listed in table 2.1. The radial distribution function (RDF), the system energies (E_{tot}, E_{kin}, and E_{pot}), pressure (P), temperature (T), and density are investigated. For the MLMD runs, estimated errors in inter atomic forces and the stress tensor components are added as well. Most of these results are retrieved directly from the VASP output files. VASP returns the following energy tags

with their values in eV; E, F, $E0$, EK, SP, and SK.[120,180] Here, E is the total energy, including the kinetic, SK, and potential, SP, energy of the Nosé-Hoover heat bath. The kinetic energy of the atomic system is EK.[180] The total energy without the heat bath, E_{tot}, the kinetic energy E_{kin} and potential energy, E_{pot} of the atomic system can be derived from the VASP outputs with the following equations:[180]

$$E_{\text{tot}} = E - SP - SK \tag{2.5}$$
$$E_{\text{kin}} = EK \tag{2.6}$$
$$E_{\text{pot}} = E_{\text{tot}} - E_{\text{kin}} = E - SP - SK - EK. \tag{2.7}$$

As the properties A that are sampled every time step, (P, T, E_{tot}, E_{kin}, and E_{pot}) are auto-correlated, the estimated error of the mean value ε_A per simulation run is determined using;

$$\varepsilon_A = \sqrt{\frac{2N_{\text{cor}}}{N_{\text{samp}}}\left(\langle A^2\rangle - \langle A\rangle^2\right)}, \tag{2.8}$$

where N_{samp} is the total number of samples and N_{cor} is the correlation time, found from fitting the auto-correlation function to $e^{-t/N_{\text{cor}}}$. The error propagation of properties calculated from these variables are addressed using the uncertainties python package,[181] version 3.1.7.

The RDFs need to be computed separately, as VASP cannot make a distinction between the oxygen atoms in OH^- or H_2O molecules. The RDFs are computed from the post-processing script that tracks the OH^- particles, of which the algorithm is presented in figure 2.7. The first shell coordination number of species i to the oxygen atoms of the water molecules around it ($n_{i\text{O}}$) is computed from the RDFs using:

$$n_{i\text{O}} = \frac{4\pi N_{\text{O}}}{V}\int_{r=0}^{r_{\text{shell}}} r^2 g_{i\text{O}}(r)dr, \tag{2.9}$$

where $g_{i\text{O}}(r)$ is the radial distribution function of species i to the oxygen in the water molecules, N_{O} is the number of oxygen atoms in water, and V the volume of the simulation box. The value of r_{shell} depends on the species i and represents the radius of the first solvation shell.

2.4.2. Reaction Rates and Transport Properties

Properties that rely on longer simulation times, such as shear viscosity, self-diffusion, electric conductivity, and to a lesser extent reaction rates, are derived using long MLMD simulations. The system properties of these KOH production runs are listed in table 2.1. The shear viscosity, ν, is calculated using Green-Kubo relations.[55,175] This depends on the stress tensor which can be directly retrieved from VASP. To improve statistics, five instead of three shear viscosity values are calculated per simulation,[73,182,183] with $P_{\alpha\beta}$ being the off-diagonal stress tensor components, P_{xy}, P_{xz}, and P_{yz}, as well as the rotational invariant combinations of the diagonal terms, $(P_{xx} + P_{yy})/2$ and $(P_{yy} + P_{zz})/2$. All five individual $\nu_{\alpha\beta}$, calculated as:

$$\nu_{\alpha\beta} = \frac{V}{k_{\text{B}}T}\int_0^t \langle P_{\alpha\beta}(t_0)P_{\alpha\beta}(t_0+t)\rangle\, dt, \tag{2.10}$$

are combined to compute a mean and standard deviation per simulation, where the the autocorrelations are calculated using a timeshift t and a time origin t_0. A cubic spline is used to interpolate the autocorrelation function. This is needed because the viscosity is very sensitive to the details of the first peak. The temperature T is the average simulation temperature and V is the system volume. The Boltzmann constant is indicated with k_{B}.

The reaction rates, self-diffusion coefficients, and the electrical conductivity cannot be retrieved directly from the VASP outputs. This is because VASP makes no distinction between oxygen atoms in H_2O and oxygen atoms in OH^- molecules. As VASP can track single particles by their particle ID, it is possible to track an oxygen of a known OH^- molecule until it reacts. After this, the particle ID of this oxygen

will refer to the oxygen in a H_2O molecule. A visualization of such reactions is illustrated in figure 2.6. A post-processing script is constructed and this is available as open source.[b] This reads the particle trajectories from VASP AIMD or MLMD runs and computes which particle ID belongs to the oxygen atoms of the OH^- molecules. It computes the number of hydrogen atoms which are bonded with oxygen atoms to differentiate the oxygen atoms in OH^- from those in H_2O. The algorithm is shown in figure 2.7.

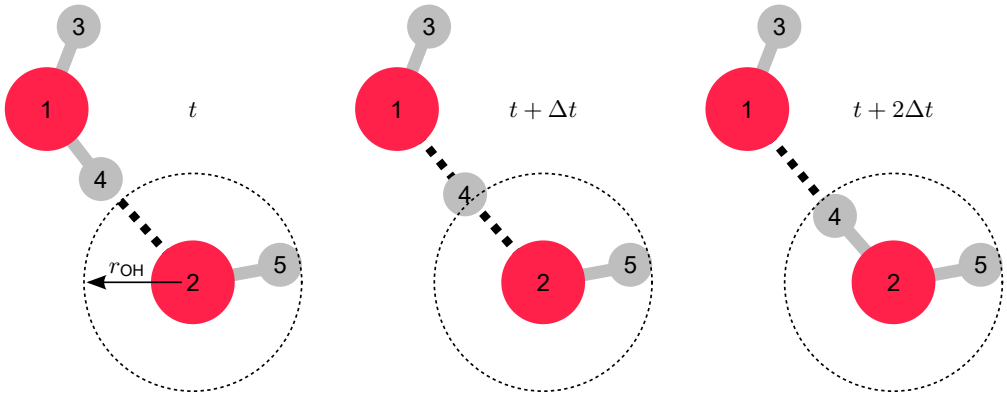

Figure 2.6: The reaction event changes which atoms are in the OH^- molecule. The time propagation of the reaction is shown from left to right (t till $t + 2\Delta t$). To determine the trajectory of the OH^- molecule, a post-processing code identifies the particle ID of the oxygen in the OH^-. The algorithm for this is presented in figure 2.7. It computes the number of hydrogen atoms within r_{OH} from the oxygen in the known OH^-, in this case atom ID 2 in (a). When it encounters 2 hydrogen atoms within this range, as indicated in (c), it searches for the new atom ID of the oxygen in the OH^- molecule, finding atom ID 1.

The post-processing code tracks the atom IDs and the unwrapped positions of oxygen atoms in OH^- molecules. The reaction rate and the self-diffusion coefficient of the OH^- can be calculated with these. The self-diffusion coefficient of species i is calculated using the Einstein-Helfand[55,175,184] relations, by tracking the mean square displacement (MSD) of the particle positions. The molecular self-diffusion coefficients, D_i^{MD} are corrected for finite-size effects:[185–187]

$$D_i^{\text{MD}} = \lim_{t\to\infty} \frac{1}{6N_i t} \left\langle \sum_{j=1}^{N_i} \left(\boldsymbol{r}_{j,i}(t_0 + t) - \boldsymbol{r}_{j,i}(t_0)\right)^2 \right\rangle \tag{2.11}$$

$$D_i^{\infty} = D_i^{\text{MD}} + D_i^{\text{corr}} = D^{\text{MD}} + \frac{k_{\text{B}} T \xi}{6\pi\nu L}. \tag{2.12}$$

where D_i^{∞} is the self-diffusion coefficient at the thermodynamic limit, N_i the number of particles, and $\boldsymbol{r}_{j,i}$ indicates the atomic position of atom j. The finite-size correction uses the mean shear viscosity ν, derived in equation (2.10), which does not show finite-size effects.[185,186,188,189] The box size is indicated as L and the dimensionless constant for cubic lattices $\xi = 2.837\,298$.[185–189] The finite-size correction is the same for all species in the mixture, as it only depends on the system properties.

The electric conductivity, σ_{NE}, is computed using the Nernst-Einstein relation.[190,191] The self-diffusion coefficients in the thermodynamic limit of the ions and their charge q are combined into:

$$\sigma_{\text{NE}} = \frac{N_{\text{ions}} q^2}{V k_{\text{B}} T} \left(D_+^{\infty} + D_-^{\infty}\right), \tag{2.13}$$

where N_{ions} is the number of ion pairs. The Nernst-Einstein relation is derived for infinite dilute electrolyte mixtures, as ion-ion correlations are neglected. Direct methods for retrieving electrical conductivity with Green-Kubo[192] or Einstein-Helfand[193,194] are more accurate, as they include these correlations.[194] Simulations with larger systems and time scales of hundreds of ns are needed for this, which is not achievable for these systems at this point.

[b]The hyperlink to the post-processing code on GitHub: `https://github.com/JelleLagerweij/Quantum_to_Transport`

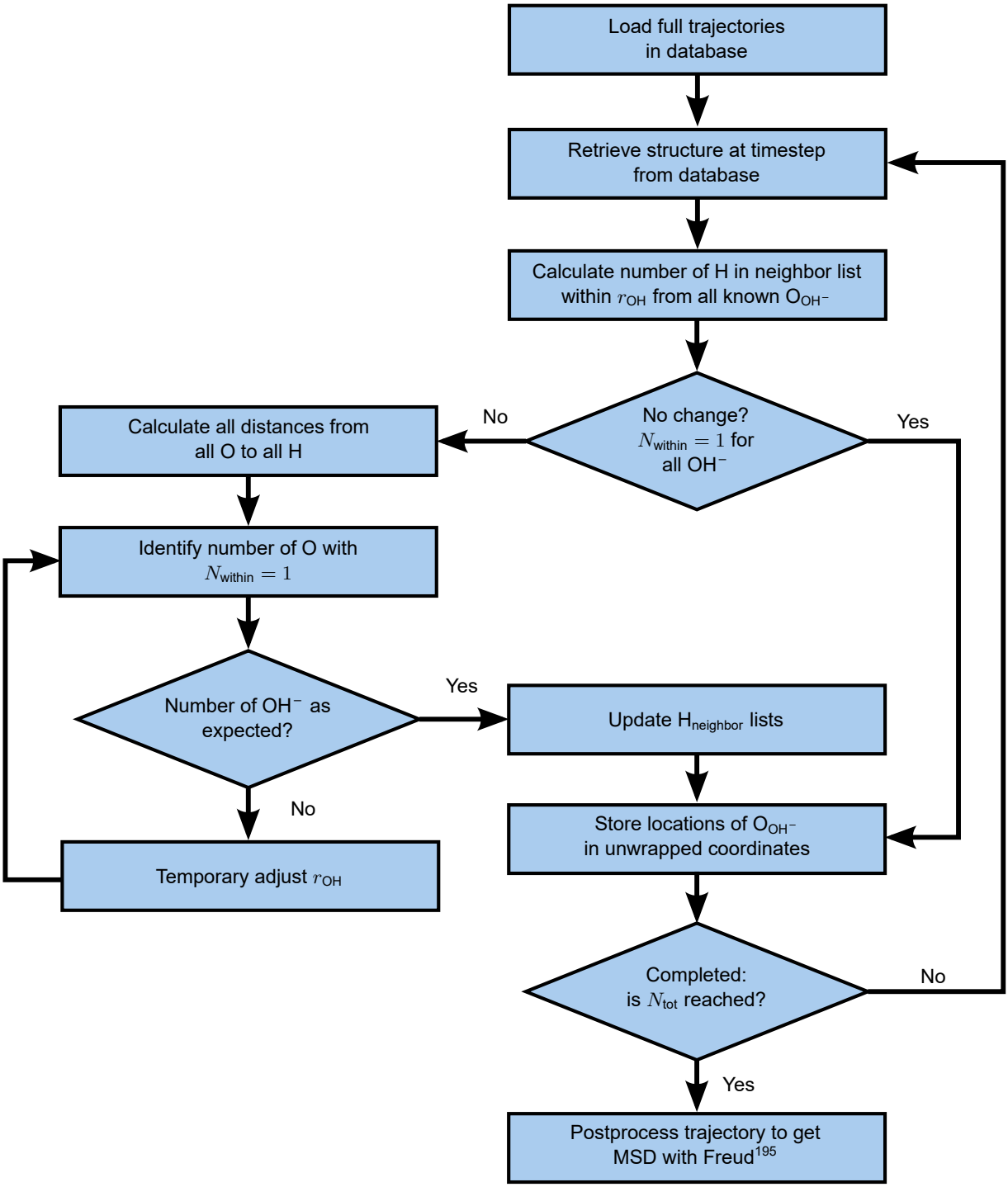

Figure 2.7: The algorithm to track the OH^- molecule locations. The algorithm is initiated by providing the atom IDs of the oxygen atoms of the OH^- (O_{OH^-}) molecules and their neighbor lists with the $N_{H_{neighbor}}$ closest hydrogen atoms and a VASP trajectory file. Implementing neighbor lists with 10 hydrogen atoms per OH^- speeds up the post-processing by 2, with a system of 1 KOH and 55 water molecules. This difference is increased for larger systems sizes. The post-processing script consists of Python 3 code and uses the HDF5[196] database output of VASP to efficiently handle the trajectory file and Freud[195] to calculate multiple windowed mean square displacements from the selected OH^- trajectories. The algorithm identifies the O_{OH^-} by counting the number of hydrogen atoms within set distance r_{OH}, which is shown in figure 2.6. This is set slightly above the typical bond length of water. This distance adjusts automatically if the expected number of O_{OH^-} is not found, as AIMD and the MLMD have flexible bond lengths.

Table 2.1: An overview of the four different systems that are simulated. The pure water and the KOH validation simulations are used to validate the settings and the KOH production simulations to retrieve the reaction rates and the transport properties. Validation simulations on pure water systems, section 3.1, show a better structure for RPBE-D3, with less computational costs. The simulations regarding KOH (aq) are therefore performed only using the RPBE-D3 DF. The pure water and KOH validation simulations of 10 ps are performed with AIMD and MLMD, while using otherwise identical settings. The KOH density and KOH production simulations are performed using MLMD only.

	unit	pure water	KOH density	KOH validation	KOH production
H_2O		64	55	55	55
K^+		0	1	1	1
OH^-		0	1	1	1
Molality KOH	$mol\,kg^{-1}$	0	1	1	1
Density	$kg\,m^{-3}$	1000	1100 - 1500	1335	1335
energy cutoff	eV	550	550	550	550
Smearing width	eV	0.3	0.3	0.3	0.3
k-points		1	1	1	1
Temperature	K	325	325	325	325
Thermostat		Nosé-Hoover	Nosé-Hoover	Nosé-Hoover	Nosé-Hoover
Heat bath mass		5	5	5	5
Time step size	fs	1	1	1	1
Time	ps	5	20	10	100

3

Results and Discussion

The pure water system is used to validate the DFT self-consistency calculations, as well as the ML. The results of this are presented in section 3.1. To adequately model the aqueous KOH with MLMD, the equilibrium density is determined, which is shown in section 3.2. The final simulations are 10 and 100 ps MLMD calculations, which are compared to 10 ps AIMD simulations. The computational efficiency gains by using MLFF for these systems are presented in section 3.3, after which the produced structure and energy properties are shown in section 3.4. This section also discusses the effect of ML on these properties. Section 3.5 shows the transport properties that are derived from the 100 ps MLMD calculations, where the effective diffusion coefficient of the proton transfer OH^- is presented.

3.1. Validating the Pure Water System

The structure of the simulated pure water system is used to investigate both the settings for the DFT self-consistency calculations and those of the ML. The structures calculated with AIMD are compared to experimental[197] and AIMD reference calculations.[179] Additionally, the performance of the new rVV10 r2SCAN DF is compared with the traditional RPBE-D3 DF. Structure properties are summarized in table 3.1 and visualized in figures 3.1 to 3.3. In the oxygen-oxygen RDFs (figure 3.1), a small shift in the first peak position is visible between RPBE-D3 and its reference structure[179] and comparing rVV10 r2SCAN with experimental work.[197] The reference[179] structure of the hydrogen-hydrogen RDFs (figure 3.2), shows a higher first peak than is calculated in this work, as well as what is found experimentally.[197] This peak represents the distance between the two hydrogen atoms within a water molecule, suggesting that the bond angle in the water molecule is stiffer in the RPBE-D3 reference than predicted experimentally. The RPBE-D3 results in this work matches the experimental work better. The hydrogen-oxygen RDFs (figure 3.3) show good agreement with both the RPBE-D3 reference[179] and experimental[197] results. The bond length of the water molecule (r_{OH}), shown in table 3.1, is significantly shorter than determined experimentally[198] and in RPBE-D3 reference work.[199]

Table 3.1: The produced structural properties of pure water. All relevant simulation settings are shown in table 2.1. The bond length (r_{OH}) and angle (θ_{HOH}) of the water molecule are derived from the first peaks of the $g_{HH}(r)$ and $g_{HO}(r)$ functions. The coordination number n_{OO} is calculated with equation (2.9), using $g_{OO}(r)$. The r_{shell} is set to 3.4 Å. The shown confidence intervals are the calculated error of the mean with a 1σ width.

		$r_{OH}/[Å]$	$\theta_{HOH}/[°]$	n_{OO}
RPBE-D3	AIMD	0.936 ± 0.002	104.8 ± 0.6	4.78 ± 0.08
	MLMD	0.936 ± 0.001	102.9 ± 0.4	4.92 ± 0.04
	Reference	0.985 ± 0.001[199]	105 ± 2[199]	4.76[179]
rVV10 r2SCAN	AIMD	0.935 ± 0.006	105.7 ± 0.5	4.57 ± 0.06
	MLMD	0.934 ± 0.002	105.2 ± 0.5	4.6 ± 0.1
Experimental		0.9572 ± 0.0003[198]	104.52 ± 0.05[198]	4.6 ± 0.1[197,200]

Differences between the RPBE-D3 and rVV10 r2SCAN structures show how rVV10 r2SCAN typically produces higher, more accentuated peaks than found experimentally[197] or with RPBE-D3. This can best be seen in the second peak in figure 3.2, the first and third peak in figure 3.3. r_{OH} is similar for RPBE-D3 and rVV10 r2SCAN, while the angle in the water molecule (θ_{HOH}) seems to be slightly overestimated with rVV10 r2SCAN. The coordination number of water (n_{OO}) of rVV10 r2SCAN matches experimental results[197,200] better than RPBE-D3. Overall, the structure properties of RPBE-D3 are preferred over those calculated with rVV10 r2SCAN.

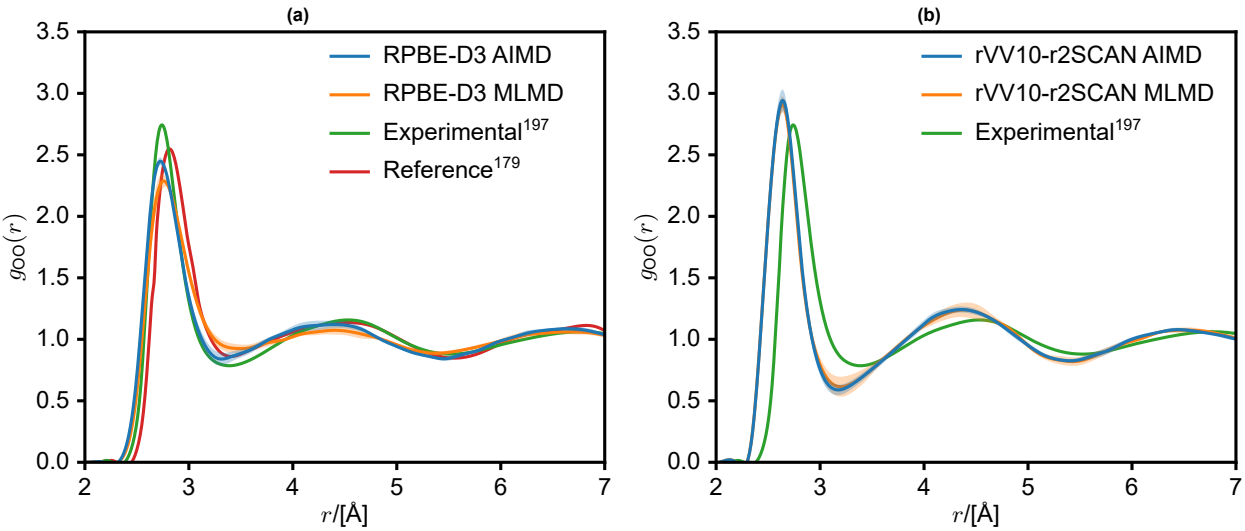

Figure 3.1: The oxygen-oxygen RDFs of pure water obtained with the (a) RPBE-D3 and (b) rVV10 r2SCAN DFs. The RDFs of AIMD and MLMD are shown together with their standard deviations. Cubic splines are used to refine the raw data. All relevant simulation settings are shown in table 2.1. The experimental work by Soper [197] and the reference structure of RPBE-D3 by Forster-Tonigold and Groß [179] are shown as well.

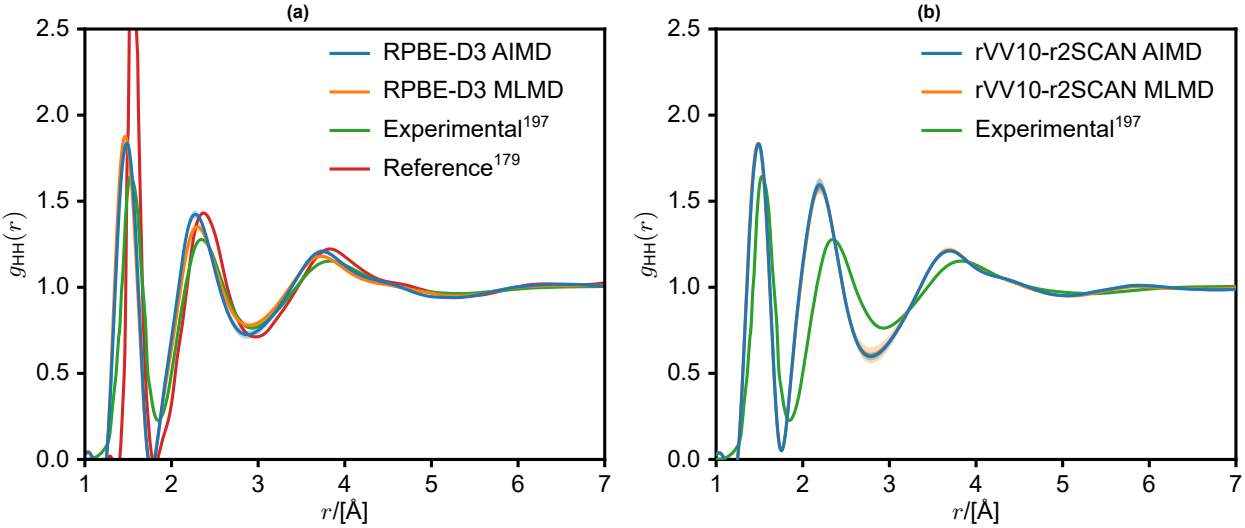

Figure 3.2: The hydrogen-hydrogen RDFs of pure water obtained with the (a) RPBE-D3 and (b) rVV10 r2SCAN DFs. The RDFs of AIMD and MLMD are shown together with their standard deviations. Cubic splines are used to refine the raw data. All relevant simulation settings are shown in table 2.1. The experimental work by Soper [197] and the reference structure of RPBE-D3 by Forster-Tonigold and Groß [179] are shown as well.

The performance of MLFFs in the pure water system is assessed by comparing the structures of the AIMD calculation with those of the MLMD simulation. MLMD is expected to perform well, as there is only one molecule type (H_2O) with only two atomic species. Therefore, the system should be quite efficient in collecting enough structures and local configurations to skip many DFT self-consistency calculations.

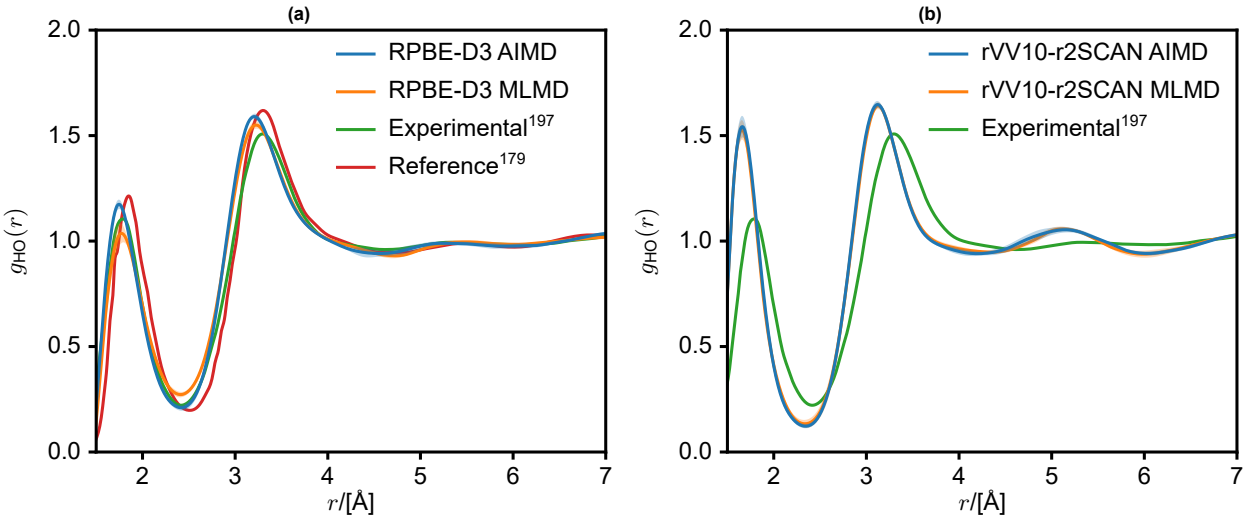

Figure 3.3: The hydrogen-oxygen RDFs of pure water obtained with the (a) RPBE-D3 and (b) rVV10 r2SCAN DFs. The RDFs of AIMD and MLMD are shown together with their standard deviations. Cubic splines are used to refine the raw data. All relevant simulation settings are shown in table 2.1. The experimental work by Soper [197] and the reference structure of RPBE-D3 by Forster-Tonigold and Groß [179] are shown as well.

Upon visual inspection, all the AIMD and MLMD RDFs match well for both RPBE-D3 and rVV10 r2SCAN. The largest differences can be seen in the first hydration shell of the water molecules with the RPBE-D3 method. This is in the $g_{OO}(r)$ and in n_{OO}. Additionally, there is some difference in θ_{HOH} of RPBE-D3, indicated in table 3.1. The rVV10 r2SCAN DF shows the best result with ML. Figure 3.1 shows that there is more variance between the individual MLMD simulations with the rVV10 r2SCAN DF for larger distances (r) compared to the AIMD simulations with the same DF. This is indicated with the orange hue around lines. Note that the mean results still match closely.

3.2. Retrieving the Correct Density of KOH (aq)

Determining the equilibrium density of KOH (aq) at 325 K and 1 bar is done by performing 20 ps MLMD simulations and selecting the density which is above 0 bar. It has been shown that DFs without VdW corrections underestimate[201,202] and those with VdW corrections typically overestimate the density of water.[99,201,202] The experimental density of KOH (aq) at a molality of 1 mol of salt per kg of water is 1048 kg m^{-3}.[203] The calculated pressures and the estimated error of the mean are shown in figure 3.4. In this work, it is computed that the equilibrium-density is between 1330 kg m^{-3} and 1335 kg m^{-3}. This is a larger overestimation than expected, which could be caused by overestimating hydrogen-bonding. The simulations are performed with gamma-point only calculations and a single k-point. The energy cutoff for the plane wave basis set and the number of k-points is investigated using a convergence study on the system energy, as explained in section 2.1.3. The pressure might be more sensitive for the number of k-points than the energy, overestimating long range attraction between atoms. Retrieving correct densities with AIMD is challenging, as small systems have large pressure fluctuations. Because of this, many studies simulate systems of fixed, experimental densities. Works by Tuckerman,[100,103,104] Parrinello[154] and others[100,179,199,204,205] are performed at fixed densities and do not report any pressure results. In other works, negative pressures are reported.[99,206] There are studies performed using NPT or other methods ensuring positive pressures. These are mostly works which presents density or density effects as their main results.[201,202] Negative pressures might be suitable for short AIMD studies, however, the aim of this work is to compute transport properties, using long MLMD simulations. These negative pressures can lead to phase separation and energy drifts. Therefore, 1335 kg m^{-3} is selected as density for the KOH validation and KOH production simulations, see table 2.1. Additional simulations at experimental density[203] (1048 kg m^{-3}) are performed as well. These are available in appendix A.

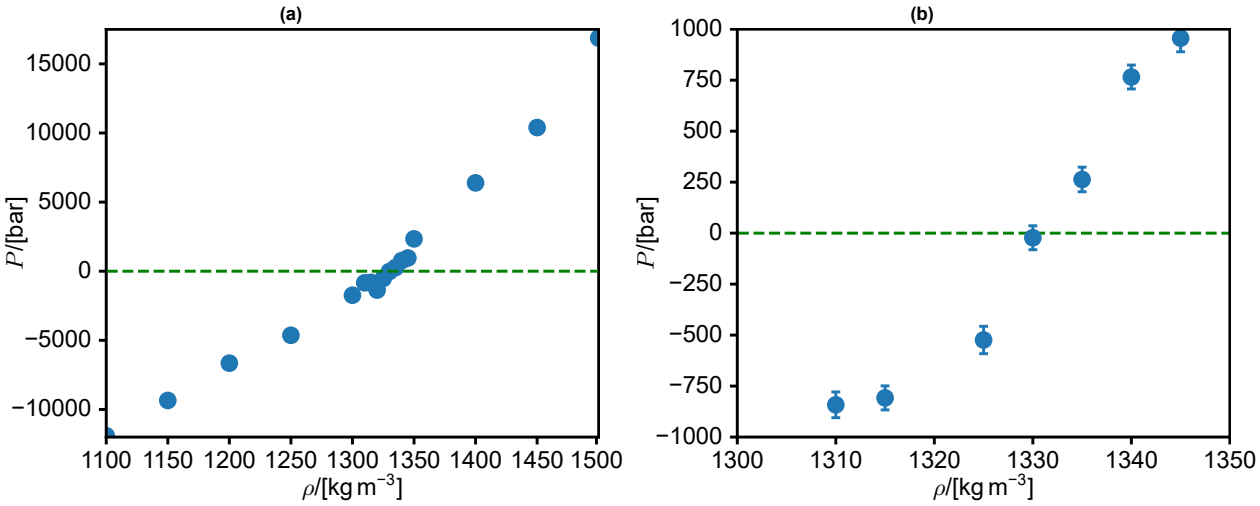

Figure 3.4: The average system pressure (P) and its standard deviation of KOH modeled with RPBE-D3 as function of its density (ρ). The entire investigated density range (a) and a zoomed in section around $P = 0$ bar (b) are shown. The dashed green line represents 0 bar, the minimal acceptable average system pressure.

3.3. Machine Learning Efficiency and Error Estimation

The efficiency gains of MLFFs are significant. Table 3.2 shows an overview of the gains in computation time. The time needed for MLMD calculations is around 3.6 to 11 times less than that of comparable AIMD calculations. In the pure water systems, the gains are the largest when using the rVV10 r2SCAN DF. This is because only the DFT self-consistency calculations are more expensive with the rVV10 r2SCAN DF, while training and using the MLFF costs the same computation time for both DFs. MLFFs with the KOH systems showed significantly less gain in efficiency than the pure water system. This has three causes: (i) the KOH system has an additional atom type for which a FF needs to be trained, (ii) the cutoff distance of the MLFF is set higher than that of the pure water system for both the radial and angular descriptors, and (iii) the radial and angular descriptors are expanded to use more basis functions. This makes training the MLFF more expensive. Additionally, more time is spent calculating the atomic forces from an already trained FF. Using these strict settings is done, because the main purpose of this work is to show that MLFFs can simulate reactions such as the proton transfer as well. Some gain in efficiency is needed to expand the time scales of AIMD to be able to predict transport properties accurately. However, these time scales are reached and this work does not focus on optimizing the computational efficiency with MLMD in VASP.

Table 3.2: The efficiency gains of MLMD compared to AIMD runs. This table indicates the number of DFT self-consistency calculations (N_{DFT}) in absolute number of steps and the percentage compared to AIMD calculations, as well as the total computational time used (t_{comp}) and the relative gain. All indicated simulations are preceded by a 5 ps initiation simulation. This initiation period is also used to pre-train the MLFF, needing around 160 N_{DFT}. The number of DFT calculations and the time spent on those are not included in the data of this table. The calculations are performed on full 48 core nodes of two intel XEON E5-6248R processors.[146] The computation time can only be used as indication, as it depends on the hardware.

	method	N_{DFT}/[-]	%	t_{comp}/[h]	times faster
Water RPBE-D3	AIMD 5 ps	5000		15	
	MLMD 5 ps	240	5	2.5	7.3
Water rVV10 r2SCAN	AIMD 5 ps	5000		70	
	MLMD 5 ps	246	5	6.4	11
KOH RPBE-D3	AIMD 10 ps	1000		20	
	MLMD 10 ps	287	2.9	4.6	4.3
	MLMD 100 ps	1879	1.9	55	3.6

The output of VASP 6.4[119–123] provides the user with the estimated errors used for the on-the-fly learning of the FF. As explained in section 2.2.2, the error threshold in the decision scheme, shown as ($\varepsilon_{\text{limit}}$) in figure 2.5, is set to decrease during the running of the simulations. It is increased again temporarily after every DFT self-consistency calculation, as not to include similar structures in the structure database. Figure 3.5 shows how the estimated error of the force interacts with the threshold that is set. Note that when the estimated force error spikes, a DFT self-consistency calculation is run to retrain the FF, after which the threshold is increased as well. This threshold is then lowered again gradually until the estimated maximum error in the force (max($\varepsilon_{\text{Force}}$)) exceeds the threshold.

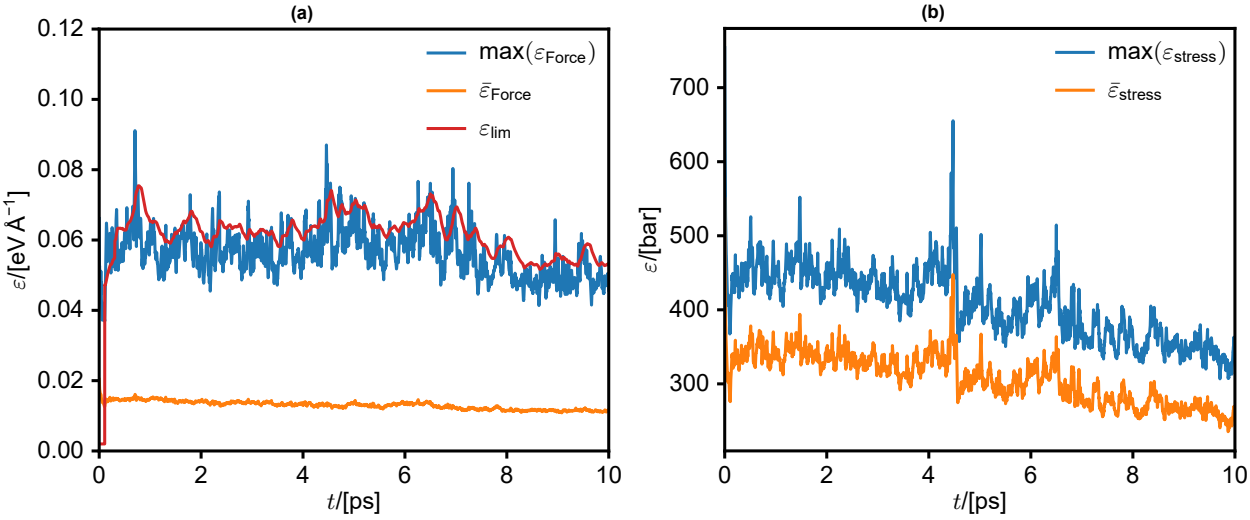

Figure 3.5: The estimated maximum and mean Bayesian error of (a) the forces and (b) system stress tensor are indicated as function of time (t). These results are derived from 10 ps MLMD simulations of aqueous KOH. The error threshold (ε_{lim}) is illustrated in (a). All relevant simulation settings are shown in table 2.1 and the simulations are performed using the RPBE-D3 DF.

3.4. Structure and Energies of KOH (aq)

The structure and energies of the KOH systems are simulated using AIMD and MLMD. The AIMD calculations are limited by calculation time. For an accurate comparison, 10 ps AIMD and MLMD simulations are set up with further identical settings. Additionally, 100 ps MLMD simulations are performed, which are compared to the shorter AIMD and MLMD simulations. The 100 ps simulations are also used to calculate the transport properties, which are further discussed in section 3.5. The structure properties of the investigated systems are summarized in table 3.3 and the energies are reported in table 3.4. The RDFs, indicating the hydration of the water and ions, are shown in figures 3.6 and 3.7. The structure around the K^+ is as expected. The first peak distance ($d_{K^+O}/[\text{Å}]$) in table 3.3 matches experimental work. This is also visible in the RDFs shown in figure 3.6a. Other works report that the structure around dissolved K^+ cation does not depend significantly on the anion species,[207] and that this distance should be close to 2.8 Å for a large concentration range. The coordination number (n_{K^+O}) is slightly higher than determined in other works.[207] This could be caused by overestimation of the equilibrium density and the hydrogen bonding between the surrounding water molecules.

Table 3.3: The produced structural results of the KOH simulations, comparing AIMD with the 10 ps and 100 ps MLMD simulations. d_{iO} indicates the shortest typical distance to water molecules from the center of molecule type i. The oxygen atoms are used as the center of H_2O and OH^- molecules. The coordination numbers are calculated using equation (2.9) with an r_{shell} of 4, 3.5, and 3.3 Å for n_{OO}, n_{K^+O} and n_{OH^-O}, respectively. All relevant simulation settings are shown in table 2.1 and the simulations are performed using the RPBE-D3 DF.

	$d_{OO}/[\text{Å}]$	n_{OO}	$d_{K^+O}/[\text{Å}]$	n_{K^+O}	$d_{OH^-O}/[\text{Å}]$	n_{OH^-O}
AIMD	2.784 ± 0.003	13.379 ± 0.007	2.827 ± 0.005	8.9 ± 0.2	2.572 ± 0.004	6.5 ± 0.1
MLMD 10 ps	2.758 ± 0.006	13.332 ± 0.006	2.82 ± 0.02	8.84 ± 0.09	2.59 ± 0.01	6.36 ± 0.04
MLMD 100 ps	2.758 ± 0.004	13.386 ± 0.007	2.818 ± 0.004	8.73 ± 0.03	2.601 ± 0.006	6.33 ± 0.06
Experimental	2.8[208]	-	2.79 ± 0.08[207]	6-8[207]	2.67 ± 0.07[209]	3-5[209]

Table 3.4: The produced average energies (E_{tot}, E_{kin}, E_{pot}), pressure (P), temperature (T), and average time between reactions per dissolved OH^- molecule (t_{react}) of 10 ps AIMD and MLMD, and 100 ps MLMD simulation. All relevant simulation settings are shown in table 2.1 and are performed with the RPBE-D3 DF.

	E_{tot}/[eV]	E_{kin}/[eV]	E_{pot}/[eV]	T/[K]	P/[bar]	t_{react}/[ps]
AIMD	-813.42 ± 0.04	6.82 ± 0.01	-820.25 ± 0.09	316.0 ± 0.6	531 ± 36	0.7 ± 0.1
MLMD 10 ps	-813.50 ± 0.04	6.82 ± 0.02	-820.32 ± 0.04	315.9 ± 0.8	1700 ± 40	0.32 ± 0.05
MLMD 100 ps	-813.47 ± 0.01	6.810 ± 0.004	-820.28 ± 0.01	315.49 ± 0.06	760 ± 60	0.39 ± 0.02

The structures around the OH^- and the H_2O molecules (determined from their respective oxygen atoms) is very different from those of pure water, as well as those of salts without OH^-. The main cause of this is the proton transfer reactions. After a reaction, the new OH^- molecule is in a local environment of that of an H_2O molecule. After this, two events can happen: (i) after a few time steps, the reaction occurs again and the molecules go back to their old environment. This often occurs, as the configuration between the molecules just after reacting is most suitable for the reaction to take place. Or, (ii) the molecules move away from each other and start to recreate their most preferable local structure. These reactions typically occur every ps, leading to a constant restructuring of the hydration shells. As the RDFs are sampled using long time averaging, they might not show the preferred local environment of OH^- and H_2O, but they could show this effect of being placed in a preferred configuration and the restructuring that occurs afterwards. Experimental work on highly concentrated hydroxide solutions[210–212] shows similar behavior. Some of their work, which is indicated in figure 3.7, shows a more drastic change in the $g_{OO}(r)$. These works are performed at higher concentrations than those in this work. However, they indicate a similar trend. The first peak is lowered and widened and the second peak of the RDF (typically found at 4.5 Å) is entirely merged into the first hydration shell. This could indicate that the tetrahedral short-range coordination in water is absent in this system. Note that the experimental work does not differentiate between the oxygen atoms in the H_2O and those in OH^-, as their neutron diffraction method can only track atomic species, and the H_2O and OH^- both consist of oxygen and hydrogen.

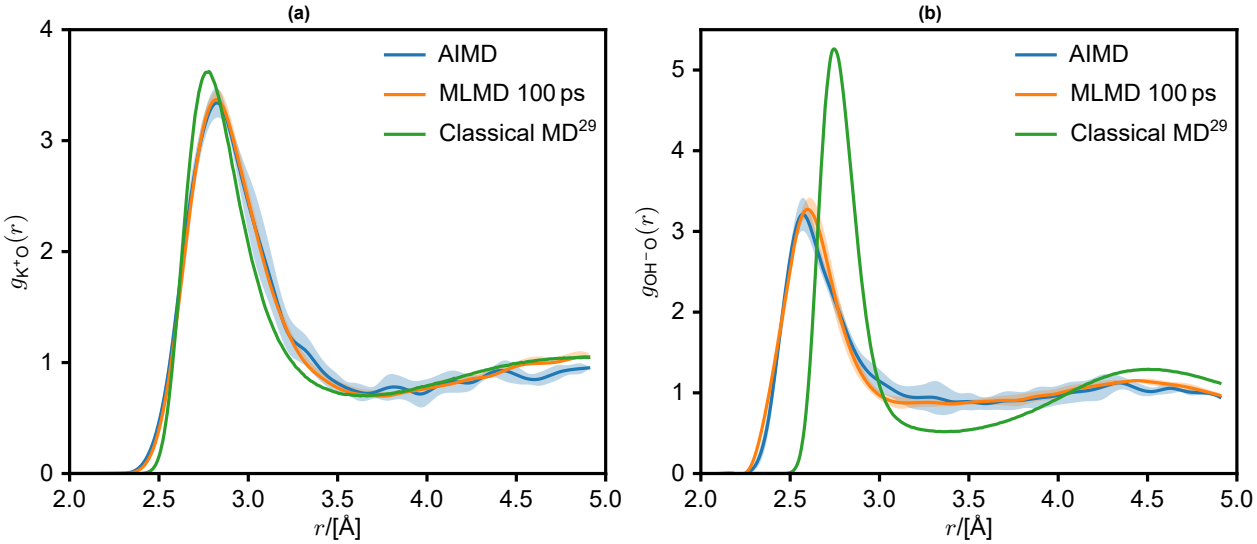

Figure 3.6: The RDFs of the (a) K^+ cation and (b) OH^- anion to their surrounding water molecules. The oxygen atoms of OH^- and H_2O molecules are used as the center of the molecule. Cubic splines are used to refine the raw data. All relevant simulation settings are shown in table 2.1. The large indicated uncertainty interval of the AIMD results in both subfigures is due to short simulation times. The AIMD RDFs are sampled over just 10 ps and the system only has a single K^+ and OH^-. The MLMD 10 ps RDFs are not shown as they behave similarly to the MLMD 100 ps RDFs with larger uncertainties. The Classical MD results[29] shown in (a) cannot be found in the original publication and are based on raw data that the author provided on request.

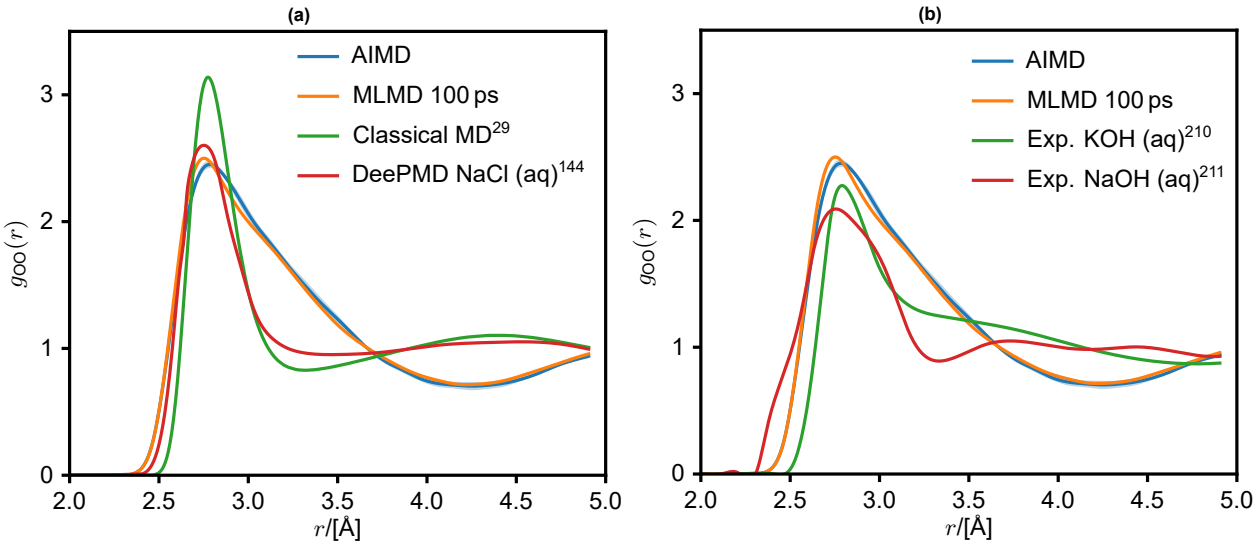

Figure 3.7: The RDFs of the water molecules to their surrounding water molecules. (a) compares the produced RDFs with classical FFs, whereas (b) compares the results with experimental work of dissolved hydroxides. The oxygen atoms of the H_2O molecules are used as the center. Cubic splines are used to refine the raw data. All relevant simulation settings are shown in table 2.1. The MLMD 10 ps RDFs are not shown as they behave similarly to the MLMD 100 ps RDFs. The Classical MD results[29] shown in (a) cannot be found in the original publication and are based on raw data that the author provided on request. The experimental reference work by Imberti et al.[210] (indicated as Exp. NaOH (aq)) and Bruni et al.[211] shown in (b) are at 4.5 and 10.5 mol kg^{-1} water, respectively. The simulations in this work are performed using 1 mol kg^{-1} water. Both experimental works used neutron diffraction methods, and Imberti et al.[210] used empirical potential structure refinement to retrieve RDFs from this.

Comparing AIMD and MLMD shows no relevant differences in the produced structures. This is indicated in table 3.3 and figures 3.6 and 3.7. Between the 10 ps and 100 ps MLMD simulations, mostly an improvement in error estimation can be seen. Comparing the energies and other state properties in table 3.4 shows well-matching energy and temperature results between all three simulation settings. No energy drifts occurred during the simulations, of which a visualization is available in appendix B. Significant differences are seen in both the average pressure and reaction rate. The calculated pressures are sensitive to slight changes in the potential energy, which might introduce some errors with MLMD. Comparing the reaction time t_{react} shows that the AIMD simulations predict a longer time between reactions than MLMD. This suggests that MLMD overestimates the reaction rate. This is noteworthy, as other works on DeePMD report that structures in the middle of reactions had to be specifically added into the training data set. Otherwise, reactions would not be captured.[139,140] The on-the-fly ML algorithm implemented in VASP is more suitable to train these reactive systems, as it will perform DFT self-consistency calculations, if the error threshold is exceeded during simulations. Besides that, the reaction happens once or twice every ps, therefore, the 10 ps AIMD and MLMD simulation only find few reactions and lack statistical significance.

A rough estimate of the influence of the proton transfer to the self-diffusion of OH^- (D_{react}) can be made by treating the reaction rate as a random walk in 3D:

$$D_{\text{react}} \backsim \frac{\Delta r^2}{6\Delta t}, \tag{3.1}$$

where Δr is the step size of the random walk. In this reaction, this is the distance between the OH^- and H_2O just before the reaction. Δr is estimated to be just over twice the bond length of water, 2 Å. Δt is the time between the steps of the random walk. This method overestimates the influence of the proton transfer on self-diffusion if Δt is taken as the calculated reaction time (t_{react}). The walk is biased to go back to where it came from, which is explained as event type (i) earlier in this section. Δt is chosen to be $2t_{\text{react}}$, which corrects for this bias if it is assumed that half of the reactions follow event type (i). The estimated increase in self-diffusion due to proton transfer is in the range of (4 - 10) $\times$ 10^{-9} m^2 s^{-1}. Although this is a rough estimate, it is in the same order of magnitude as experimental methods have calculated for OH^-.[213]

3.5. Transport Properties of KOH (aq)

The transport properties are calculated using the 100 ps MLMD. The final results are summarized in table 3.5. The shear viscosity is calculated using the last 75 ps of the KOH production simulations, as the estimated error in the stress is high in at least the first 10 ps, which is visualized in figure 3.5. The calculated shear viscosity is significantly higher than expected from experimental results.[203] AIMD simulations of pure water and MLFFs based on those indicate higher viscosity as well.[145] This is explained by the overestimated hydrogen bonding and a melting temperature shift caused by that.[145] Other work with a simplified AIMD scheme using the RPBE-D3 DF, found a viscosity close to experimental results.[204] They used experimental densities when simulating the system and did not report the system pressures and energies. This difference can be explained by the deviating density, as higher viscosities are expected at higher densities. For this work, transport properties are also calculated at experimental density. These are available in appendix A. Although these match experimentally found values better than the results at 1335 kg m^{-3} and no energy drifts are observed, the average system pressure is $(-12.95 \pm 0.07) \times 10^3$ bar, which is why these results are not presented as the main results of this work.

Muñoz-Santiburcio [204] also calculated self-diffusion of H_2O and OH^- at infinite dilution. They report a self-diffusion coefficient of H_2O ($D^{\infty}_{H_2O}$), which is close to the experimental result of 2.45×10^{-9} m^2 s^{-1}.[204,214] In this work, the calculated self-diffusion of H_2O is lower than found experimentally, which follows logically from the overestimation of the viscosity. The calculated $D^{\infty}_{K^+}$ is similarly lower, however, $D^{\infty}_{OH^-}$ is dominated by the proton-transfer events. The calculated self-diffusion of the anion is $(6 \pm 2) \times 10^{-9}$ m^2 s^{-1}, much closer to experimental work and matching the 4.7×10^{-9} m^2 s^{-1} determined by Muñoz-Santiburcio [204]. This indicates that although the system is less mobile than it should be, the proton transfer is modeled successfully. The self-diffusion coefficients are derived from the MSDs of the specific species. It must be noted that although the MSD of H_2O and OH^- are above L^2 (where L is the simulation box size), the MSD of K^+ was below this. Determining the self-diffusion from MSDs smaller than L^2 can introduce additional errors. The MSDs of the species are visualized in appendix B.3.

Table 3.5: The produced transport properties calculated using the 100 ps MLMD simulation, showing the viscosity (ν) and self-diffusivity of H_2O, K^+, and OH^- ($D^{\infty}_{H_2O}$, $D^{\infty}_{K^+}$, and $D^{\infty}_{OH^-}$) in the thermodynamic limit. The electric conductivity σ is calculated using the Nernst-Einstein shown in equation (2.13). All relevant simulation settings are shown in table 2.1 and are performed with the RPBE-D3 DF.

	ν/[mPa s]	D^{corr}/[$\times$ 10^{-9} m^2 s^{-1}]	$D^{\infty}_{H_2O}$/[$\times$ 10^{-9} m^2 s^{-1}]
MLMD 100 ps	8.64 ± 0.07	0.0695 ± 0.0005	1.16 ± 0.08
Experimental	1.005[203]	-	2.45[214]
	$D^{\infty}_{K^+}$/[$\times$ 10^{-9} m^2 s^{-1}]	$D^{\infty}_{OH^-}$/[$\times$ 10^{-9} m^2 s^{-1}]	σ_{NE}/[S m^{-1}]
MLMD 100 ps	0.6 ± 0.1	6 ± 2	29 ± 7
Experimental	1.96[215]	5.27[215]	26[213]

4

Conclusions and Outlook

Aqueous hydroxide salts are studied extensively, as understanding their behavior is relevant for industrial applications[9,10] and energy storage solutions, such as batteries and electrolyzers.[20–22] These studies range from experimental works[210–213] to computational models.[29,100,103,104] Accurately describing the behavior of KOH with computer models is challenging, as the OH^- and H_2O exchange hydrogen atoms. This is a proton transfer reaction and cannot be captured well with current MD methods. Simulations using first principle methods, such as AIMD, can capture these reactions. However, these methods have higher computational expenses, which results in limited length and time scales that can be modeled. Longer length and time scales are needed to determine transport properties, such as self-diffusion of the OH^-. This work shows that MLFFs can significantly reduce the computational costs of quantum mechanics simulations. Additionally, it proves that on-the-fly MLMD, implemented in VASP 6.4,[119–124] can produce similar structures and energies as ab initio molecular dynamics. The results show that not only simple molecular interactions such as pure water can be described this way, but also complicated mechanisms such as proton transfer reactions are captured. This work also shows that MLMD predicts that these reactions influence the structure the same as in AIMD. This has been found experimentally[210–212] and with the complex ReaxFF molecular FF,[111–114] but cannot be captured with traditional force fields.[29] Additionally, the self-diffusion coefficient of OH^- is computed to be $(6 \pm 2) \times 10^{-9}\,m^2\,s^{-1}$, at a molality of 1 mol of KOH per kg of water. This is close to the experimental result of $5.27 \times 10^{-9}\,m^2\,s^{-1}$, at infinite dilution. The reported gains in computation time are modest, since simulations are completed between 3.6 and 11 times faster. This is caused by the ML settings selected, which is set to ensure the capture of the proton transfer events. The radial and angular descriptors consist of more basis functions and a longer cutoff distance than is set by default. Additionally, the error threshold implemented is set to reduce over time, and is only temporarily relaxed after new AIMD calculations are performed. RPBE-D3, the selected DF, is relatively cheap to converge. Most gains in computation efficiency can be reached by using modern, computationally more expensive DFs. This is because the computational costs of calculating the forces on atoms by using a trained MLFF does not depend on the DF. The DFT self-consistency calculations will be more expensive. However, these are only a small percentage of the time steps.

Other results here show some limitations to MLMD calculations of aqueous KOH. The most noteworthy example of this is the calculated equilibrium density, which is calculated to be $1335\,kg\,m^{-3}$, significantly higher than the experimental value $1048\,kg\,m^{-3}$.[203] This difference is especially noteworthy as the system is in the liquid state. This has multiple causes. The RPBE-D3 DF overestimates hydrogen bonding, which results in an overestimated freezing point[145] and density for water systems. Besides that, only NVT simulations at different densities are executed and the lowest density with a pressure above 0 bar has been selected. Alternatively, NPT calculations could have been performed. However, for small simulation box sizes, the pressure fluctuates significantly. This increases the difficulty in setting up NPT calculations correctly. This higher density, and overestimation of hydrogen bonding also explains the deviation in viscosity, which is overestimated (7.4 ± 0.1 instead of 1.005 mPa s). This is because the system is over-bonded and less mobile than it should be. The underestimation of the self-diffusion of K^+ and H_2O also logically follows from this. The self-diffusion of the OH^- ions is dominated by the

proton transfer and suffers less from this effect. Improvements to this can be made by setting up NPT calculations with larger box-sizes and by selecting more accurate DFs.

Future work into the local environment of when the proton transfer reactions occur might provide new insights. In 1995, Tuckerman et al.[103] found a local configuration to be stable when the OH^- was hydrated in a $(H_9O_5)^-$ group and a less stable if it was hydrated in $(H_7O_4)^-$. This $(H_7O_4)^-$ configuration often results in multiple proton transfer event back and forth. MLMD with modern, more accurate, DFs might provide better insights in which local structures lead to these reactions. Combining this with energy barriers which can be calculated with nudged elastic band simulations might provide enough information to create an accurate MD-MC scheme. Trajectories could then be calculated with a classical FF and when a local configuration is found to be appropriate, something equivalent to an MC step using the energy barrier could be executed to simulate the reaction. Although MLMD is significantly faster than AIMD, classical FFs cost orders of magnitude less computational effort. With such a FF, it would be possible to predict the electric conductivity of aqueous hydroxide salts including ion-ion correlations.[194] This needs to be in the order of 100 ns, which is not yet achievable with current MLMD techniques with enough detail to capture reactions.

MLFFs provides new opportunities to the community of molecular simulations. MLFFs are able to explicitly describe all two- and three-body interactions and implicitly express many-body interactions, as it provides an additional tool to molecular simulation researchers. Optimizing the coefficients of all separate bond-types needed to describe molecular systems takes time and skill. MLFFs provide methods which can automate this process, based on quantum mechanics. This also introduces new challenges, as DFs are traditionally created for solid-state physics or catalysis research communities. There, the interaction distances are small, thus the DFs are typically suitable to describe metallic systems and chemical reactions. New DFs that are more suitable for liquids and gasses should be further developed. Additionally, molecular simulations with an engineering aim are often parameterized to experimental data which can be relevant for their specific use case. Basing the entire description of atomic interactions on quantum mechanics might be correct from a physics point of view, however, this is not always practical. Creating a combined approach, using macroscopic experimental results together with quantum mechanics simulations when creating MLFFs, would expand the use case of MLFFs.

Some improvements to the current ML implementation in VASP[123,124] are needed as well. For example, VASP is not created for traditional molecular simulation use. This can be recognized in three issues. The (i) parallelization of VASP is optimized for DFT self-consistency calculations and not for computing MD time steps from a MLFF. This is a small issue as MLMD is still significantly faster than AIMD and the system sizes are typically still small. For most computational gains, it is logical to use the on-the-fly ML on a relatively small system, after which a pure MD simulation on a larger system is executed using this pre-trained FF. However, (ii) changing the size of the box has significant consequences, as VASP cuts of all interactions at the cutoff distance. Adding long-range Coulombic interactions is not possible. The (iii) available post-processing codes for molecular simulation properties are also not readily available for VASP. All three issues can be solved together by adding a tool to export a pre-trained MLFF to established molecular simulation software, such as LAMMPS[49–51] or GROMACS.[52] These software are optimized for MD calculations and post-processing codes are available as well. This export tool could then add electrostatics as a classical interaction with support for Particle–Particle–Particle–Mesh or Ewald summation to r1educe computational effort. A method to remove the short range electrostatics from the MLFF should also be developed, to not double count these interactions.

References

(1) Robinson, R. A.; Stokes, R. H., *Electrolyte Solutions the Measurement and Interpretation of Conductance, Chemical Potential, and Diffusion in Solutions of Simple Electrolytes.* 2d ed., rev; Butterworths: London, 1970; 571 pp.

(2) Cuartero, M. “Electrochemical Sensors for In-Situ Measurement of Ions in Seawater”. *Sensors and Actuators B: Chemical* **2021**, *334*, 129635.

(3) Stockar, U. von, *Biothermodynamics: The Role of Thermodynamics in Biochemical Engineering*; PPUR Presses polytechniques: 2013; 632 pp.

(4) Ibis, F.; Nuhu, M.; Penha, F.; Yu, T.; Van Der Heijden, A.; Kramer, H.; Eral, H. “Role of Hyaluronic Acid on the Nucleation Kinetics of Calcium Oxalate Hydrates in Artificial Urine Quantified with Droplet Microfluidics”. *Crystal Growth and Design* **2022**, *22*, 3834–3844.

(5) Ibrahim, Y.; Ismail, R. A.; Ogungbenro, A.; Pankratz, T.; Banat, F.; Arafat, H. A. “The Sociopolitical Factors Impacting the Adoption and Proliferation of Desalination: A Critical Review”. *Desalination* **2021**, *498*, 114798.

(6) Zarzo, D.; Prats, D. “Desalination and Energy Consumption. What Can We Expect in the near Future?” *Desalination* **2018**, *427*, 1–9.

(7) Darre, N. C.; Toor, G. S. “Desalination of Water: A Review”. *Current Pollution Reports* **2018**, *4*, 104–111.

(8) J. Medvedev, J.; Tobolovskaya, Y.; V. Medvedeva, X.; W. Tatarchuk, S.; Li, F.; Klinkova, A. “Pathways of Ammonia Electrooxidation on Nickel Hydroxide Anodes and an Alternative Route towards Recycled Fertilizers”. *Green Chemistry* **2022**, *24*, 1578–1589.

(9) Mazurek, K.; Białowicz, K.; Trypuć, M. “Extraction of Vanadium Compounds from the Used Vanadium Catalyst with the Potassium Hydroxide Solution”. *Polish Journal of Chemical Technology* **2010**, *12*, 23–28.

(10) Ofomaja, A. E.; Naidoo, E. B.; Modise, S. J. “Biosorption of Copper(II) and Lead(II) onto Potassium Hydroxide Treated Pine Cone Powder”. *Journal of Environmental Management* **2010**, *91*, 1674–1685.

(11) United Nations Environment Programme “Paris Agreement”. **2015**.

(12) Fritz, H. P.; Besenhard, J. O.; Paxinos, A. S. “High Power Density Lead Acid Cell Operating at 70 °C”. *Zeitschrift für Naturforschung B* **1978**, *33*, 1245–1253.

(13) Shukla, A. K.; Venugopalan, S.; Hariprakash, B. “Nickel-Based Rechargeable Batteries”. *Journal of Power Sources* **2001**, *100*, 125–148.

(14) Song, W.; Ji, X.; Zhu, Y.; Zhu, H.; Li, F.; Chen, J.; Lu, F.; Yao, Y.; Banks, Craig. E. “Aqueous Sodium-Ion Battery Using a Na3V2(PO4)3 Electrode”. *ChemElectroChem* **2014**, *1*, 871–876.

(15) Zhang, Q.; Yang, Z.; Ji, H.; Zeng, X.; Tang, Y.; Sun, D.; Wang, H. “Issues and Rational Design of Aqueous Electrolyte for Zn-ion Batteries”. *SusMat* **2021**, *1*, 432–447.

(16) Shen, Y.; Liu, B.; Liu, X.; Liu, J.; Ding, J.; Zhong, C.; Hu, W. “Water-in-Salt Electrolyte for Safe and High-Energy Aqueous Battery”. *Energy Storage Materials* **2021**, *34*, 461–474.

(17) Kwabi, D. G.; Ji, Y.; Aziz, M. J. “Electrolyte Lifetime in Aqueous Organic Redox Flow Batteries: A Critical Review”. *Chemical Reviews* **2020**, *120*, 6467–6489.

(18) Hugo, Y. A.; Kout, W.; Dalessi, G.; Forner-Cuenca, A.; Borneman, Z.; Nijmeijer, K. “Techno-Economic Analysis of a Kilo-Watt Scale Hydrogen-Bromine Flow Battery System for Sustainable Energy Storage”. *Processes* **2020**, *8*, 1492.

(19) Chatenet, M.; G. Pollet, B.; R. Dekel, D.; Dionigi, F.; Deseure, J.; Millet, P.; D. Braatz, R.; Z. Bazant, M.; Eikerling, M.; Staffell, I.; Balcombe, P.; Shao-Horn, Y.; Schäfer, H. “Water Electrolysis: From

Textbook Knowledge to the Latest Scientific Strategies and Industrial Developments". *Chemical Society Reviews* **2022**, *51*, 4583–4762.

(20) Schalenbach, M.; Zeradjanin, A. R.; Kasian, O.; Cherevko, S.; Mayrhofer, K. J. J. "A Perspective on Low-Temperature Water Electrolysis – Challenges in Alkaline and Acidic Technology". *International Journal of Electrochemical Science* **2018**, *13*, 1173–1226.

(21) Kraglund, M. R.; Aili, D.; Jankova, K.; Christensen, E.; Li, Q.; Jensen, J. O. "Zero-Gap Alkaline Water Electrolysis Using Ion-Solvating Polymer Electrolyte Membranes at Reduced KOH Concentrations". *Journal of The Electrochemical Society* **2016**, *163*, F3125.

(22) Solovey, V. V.; Shevchenko, A. A.; Zipunnikov, M. M.; Kotenko, A. L.; Khiem, N. T.; Tri, B. D.; Hai, T. T. "Development of High Pressure Membraneless Alkaline Electrolyzer". *International Journal of Hydrogen Energy* **2022**, *47*, 6975–6985.

(23) Rogers, P. S. Z.; Bradley, D. J.; Pitzer, K. S. "Densities of Aqueous Sodium Chloride Solutions from 75 to 200.Degree.C at 20 Bar". *Journal of Chemical & Engineering Data* **1982**, *27*, 47–50.

(24) Messikomer, E. E.; Wood, R. H. "The Enthalpy of Dilution of Aqueous Sodium Chloride at 298.15 to 373.15 K, Measured with a Flow Calorimeter". *The Journal of Chemical Thermodynamics* **1975**, *7*, 119–130.

(25) Langer, H.; Offermann, H. "On the Solubility of Sodium Chloride in Water". *Journal of Crystal Growth* **1982**, *60*, 389–392.

(26) Hosseini, M.; Fahimpour, J.; Ali, M.; Keshavarz, A.; Iglauer, S. "H2-brine Interfacial Tension as a Function of Salinity, Temperature, and Pressure; Implications for Hydrogen Geo-Storage". *Journal of Petroleum Science and Engineering* **2022**, *213*, 110441.

(27) Truesdell, A. H. "Activity Coefficients of Aqueous Sodium Chloride from 15° to 50°C Measured with a Glass Electrode". *Science* **1968**, *161*, 884–886.

(28) Ostroff, A. G.; Snowden, B. S.; Woessner, D. E. "Viscosities of Protonated and Deuterated Water Solutions of Alkali Metal Chlorides". *The Journal of Physical Chemistry* **1969**, *73*, 2784–2785.

(29) Habibi, P.; Rahbari, A.; Blazquez, S.; Vega, C.; Dey, P.; Vlugt, T. J. H.; Moultos, O. A. "A New Force Field for OH– for Computing Thermodynamic and Transport Properties of H2 and O2 in Aqueous NaOH and KOH Solutions". *The Journal of Physical Chemistry B* **2022**, *126*, 9376–9387.

(30) Bešter-Rogač, M.; Neueder, R.; Barthel, J. "Conductivity of Sodium Chloride in Water + 1,4-Dioxane Mixtures from 5 to 35°C II. Concentrated Solution". *Journal of Solution Chemistry* **2000**, *29*, 51–61.

(31) Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. "Virtual Screening Using Molecular Simulations". *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 1940–1951.

(32) Redlich, Otto.; Kwong, J. N. S. "On the Thermodynamics of Solutions. V. An Equation of State. Fugacities of Gaseous Solutions." *Chemical Reviews* **1949**, *44*, 233–244.

(33) Peng, D.-Y.; Robinson, D. B. "A New Two-Constant Equation of State". *Industrial & Engineering Chemistry Fundamentals* **1976**, *15*, 59–64.

(34) Kontogeorgis, G. M.; Voutsas, E. C.; Yakoumis, I. V.; Tassios, D. P. "An Equation of State for Associating Fluids". *Industrial & Engineering Chemistry Research* **1996**, *35*, 4310–4318.

(35) Chapman, W. G.; Gubbins, K. E.; Jackson, G.; Radosz, M. "SAFT: Equation-of-state Solution Model for Associating Fluids". *Fluid Phase Equilibria* **1989**, *52*, 31–38.

(36) Stokes, R. H.; Robinson, R. A. "Ionic Hydration and Activity in Electrolyte Solutions". *Journal of the American Chemical Society* **1948**, *70*, 1870–1878.

(37) Ohtaki, Hitoshi.; Radnai, Tamas. "Structure and Dynamics of Hydrated Ions". *Chemical Reviews* **1993**, *93*, 1157–1204.

(38) Marcus, Y. "Effect of Ions on the Structure of Water: Structure Making and Breaking". *Chemical Reviews* **2009**, *109*, 1346–1370.

(39) Debye, P.; Hückel, E. "Zur Theorie der Elektrolyte". *Physikalische Zeitschrift* **1923**, *24*, 185–206.

(40) Born, M. "Volumen und Hydratationswärme der Ionen". *Zeitschrift für Physik* **1920**, *1*, 45–48.

(41) Myers, J. A.; Sandler, S. I.; Wood, R. H. “An Equation of State for Electrolyte Solutions Covering Wide Ranges of Temperature, Pressure, and Composition”. *Industrial & Engineering Chemistry Research* **2002**, *41*, 3282–3297.

(42) Lin, Y.; Thomsen, K.; de Hemptinne, J.-c. “Multicomponent equations of state for electrolytes”. *AIChE Journal* **2007**, *53*, 989–1005.

(43) Held, C.; Cameretti, L. F.; Sadowski, G. “Modeling Aqueous Electrolyte Solutions: Part 1. Fully Dissociated Electrolytes”. *Fluid Phase Equilibria* **2008**, *270*, 87–96.

(44) Kontogeorgis, G. M.; Schlaikjer, A.; Olsen, M. D.; Maribo-Mogensen, B.; Thomsen, K.; von Solms, N.; Liang, X. “A Review of Electrolyte Equations of State with Emphasis on Those Based on Cubic and Cubic-Plus-Association (CPA) Models”. *International Journal of Thermophysics* **2022**, *43*, 54.

(45) Costa Reis, M. “Current Trends in Predictive Methods and Electrolyte Equations of State”. *ACS Omega* **2022**, *7*, 16847–16855.

(46) Nikolaidis, I. K.; Poursaeidesfahani, A.; Csaszar, Z.; Ramdin, M.; Vlugt, T. J. H.; Economou, I. G.; Moultos, O. A. “Modeling the Phase Equilibria of Asymmetric Hydrocarbon Mixtures Using Molecular Simulation and Equations of State”. *AIChE Journal* **2019**, *65*, 792–803.

(47) Rahbari, A.; Hens, R.; Nikolaidis, I. K.; Poursaeidesfahani, A.; Ramdin, M.; Economou, I. G.; Moultos, O. A.; Dubbeldam, D.; Vlugt, T. J. H. “Computation of Partial Molar Properties Using Continuous Fractional Component Monte Carlo”. *Molecular Physics* **2018**, *116*, 3331–3344.

(48) Tsimpanogiannis, I. N.; Maity, S.; Celebi, A. T.; Moultos, O. A. “Engineering Model for Predicting the Intradiffusion Coefficients of Hydrogen and Oxygen in Vapor, Liquid, and Supercritical Water Based on Molecular Dynamics Simulations”. *Journal of Chemical & Engineering Data* **2021**, *66*, 3226–3244.

(49) Plimpton, S. “Fast Parallel Algorithms for Short-Range Molecular Dynamics”. *Journal of Computational Physics* **1995**, *117*, 1–19.

(50) Plimpton, S.; Kohlmeyer, A.; Thompson, A.; Moore, S.; Berger, R. *LAMMPS Stable Release 29 September 2021*, Zenodo, 2021.

(51) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in ’t Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. “LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales”. *Computer Physics Communications* **2022**, *271*, 108171.

(52) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. “GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation”. *Computer Physics Communications* **1995**, *91*, 43–56.

(53) Hens, R.; Rahbari, A.; Caro-Ortiz, S.; Dawass, N.; Erdős, M.; Poursaeidesfahani, A.; Salehi, H. S.; Celebi, A. T.; Ramdin, M.; Moultos, O. A.; Dubbeldam, D.; Vlugt, T. J. H. “Brick-CFCMC: Open Source Software for Monte Carlo Simulations of Phase and Reaction Equilibria Using the Continuous Fractional Component Method”. *Journal of Chemical Information and Modeling* **2020**, *60*, 2678–2682.

(54) Polat, H. M.; Salehi, H. S.; Hens, R.; Wasik, D. O.; Rahbari, A.; de Meyer, F.; Houriez, C.; Coquelet, C.; Calero, S.; Dubbeldam, D.; Moultos, O. A.; Vlugt, T. J. H. “New Features of the Open Source Monte Carlo Software Brick-CFCMC: Thermodynamic Integration and Hybrid Trial Moves”. *Journal of Chemical Information and Modeling* **2021**, *61*, 3752–3757.

(55) Frenkel, D.; Smit, B., *Understanding Molecular Simulation: From Algorithms to Applications*; Elsevier: 2001; 661 pp.

(56) Lennard-Jones, J. E. “Cohesion”. *Proceedings of the Physical Society* **1931**, *43*, 461.

(57) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. “The Missing Term in Effective Pair Potentials”. *The Journal of Physical Chemistry* **1987**, *91*, 6269–6271.

(58) Joung, I. S.; Cheatham, T. E. I. “Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations”. *The Journal of Physical Chemistry B* **2008**, *112*, 9020–9041.

(59) Weerasinghe, S.; Smith, P. E. “A Kirkwood–Buff Derived Force Field for Sodium Chloride in Water”. *The Journal of Chemical Physics* **2003**, *119*, 11342–11349.

(60) Abascal, J. L. F.; Vega, C. “A General Purpose Model for the Condensed Phases of Water: TIP4P/2005”. *The Journal of Chemical Physics* **2005**, *123*, 234505.

(61) Zeron, I. M.; Abascal, J. L. F.; Vega, C. “A Force Field of Li+, Na+, K+, Mg2+, Ca2+, Cl-, and SO42- in Aqueous Solution Based on the TIP4P/2005 Water Model and Scaled Charges for the Ions”. *The Journal of Chemical Physics* **2019**, *151*, 134504.

(62) Blazquez, S.; Conde, M. M.; Abascal, J. L. F.; Vega, C. “The Madrid-2019 Force Field for Electrolytes in Water Using TIP4P/2005 and Scaled Charges: Extension to the Ions F-, Br-, I-, Rb+, and Cs+”. *The Journal of Chemical Physics* **2022**, *156*, 044505.

(63) Blazquez, S.; Conde, M. M.; Vega, C. “Scaled Charges for Ions: An Improvement but Not the Final Word for Modeling Electrolytes in Water”. *The Journal of Chemical Physics* **2023**, *158*, 054505.

(64) Warshel, A.; Kato, M.; Pisliakov, A. V. “Polarizable Force Fields: History, Test Cases, and Prospects”. *Journal of Chemical Theory and Computation* **2007**, *3*, 2034–2045.

(65) Vega, C.; F. Abascal, J. L. “Simulating Water with Rigid Non-Polarizable Models: A General Perspective”. *Physical Chemistry Chemical Physics* **2011**, *13*, 19663–19688.

(66) Jiang, H.; Moultos, O. A.; Economou, I. G.; Panagiotopoulos, A. Z. “Hydrogen-Bonding Polarizable Intermolecular Potential Model for Water”. *The Journal of Physical Chemistry B* **2016**, *120*, 12358–12370.

(67) Lamoureux, G.; MacKerell, A. D.; Roux, B. “A Simple Polarizable Model of Water Based on Classical Drude Oscillators”. *The Journal of Chemical Physics* **2003**, *119*, 5185–5197.

(68) Kiss, P. T.; Baranyai, A. “A Systematic Development of a Polarizable Potential of Water”. *The Journal of Chemical Physics* **2013**, *138*, 204507.

(69) Lamoureux, G.; Roux, B. “Absolute Hydration Free Energy Scale for Alkali and Halide Ions Established from Simulations with a Polarizable Force Field”. *The Journal of Physical Chemistry B* **2006**, *110*, 3308–3322.

(70) Kiss, P. T.; Baranyai, A. “A New Polarizable Force Field for Alkali and Halide Ions”. *The Journal of Chemical Physics* **2014**, *141*, 114501.

(71) Smith, W. R.; Moučka, F.; Nezbeda, I. “Osmotic Pressure of Aqueous Electrolyte Solutions via Molecular Simulations of Chemical Potentials: Application to NaCl”. *Fluid Phase Equilibria* **2016**, *407*, 76–83.

(72) Yue, S.; Panagiotopoulos, A. Z. “Dynamic Properties of Aqueous Electrolyte Solutions from Non-Polarisable, Polarisable, and Scaled-Charge Models”. *Molecular Physics* **2019**, *117*, 3538–3549.

(73) Jiang, H.; Mester, Z.; Moultos, O. A.; Economou, I. G.; Panagiotopoulos, A. Z. “Thermodynamic and Transport Properties of H2O + NaCl from Polarizable Force Fields”. *Journal of Chemical Theory and Computation* **2015**, *11*, 3802–3810.

(74) Panagiotopoulos, A. Z. “Simulations of Activities, Solubilities, Transport Properties, and Nucleation Rates for Aqueous Electrolyte Solutions”. *The Journal of Chemical Physics* **2020**, *153*, 010903.

(75) Riera, M.; Lambros, E.; T. Nguyen, T.; W. Götz, A.; Paesani, F. “Low-Order Many-Body Interactions Determine the Local Structure of Liquid Water”. *Chemical Science* **2019**, *10*, 8211–8218.

(76) Xantheas, S. S. “Ab Initio Studies of Cyclic Water Clusters (H2O)n, N=1–6. II. Analysis of Many-body Interactions”. *The Journal of Chemical Physics* **1994**, *100*, 7523–7534.

(77) Kim, K. S.; Dupuis, M.; Lie, G. C.; Clementi, E. “Revisiting Small Clusters of Water Molecules”. *Chemical Physics Letters* **1986**, *131*, 451–456.

(78) Medders, G. R.; Götz, A. W.; Morales, M. A.; Bajaj, P.; Paesani, F. “On the Representation of Many-Body Interactions in Water”. *The Journal of Chemical Physics* **2015**, *143*, 104102.

(79) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. “Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions”. *Chemical Reviews* **2016**, *116*, 7501–7528.

(80) Kumar, R.; Skinner, J. L. “Water Simulation Model with Explicit Three-Molecule Interactions”. *The Journal of Physical Chemistry B* **2008**, *112*, 8311–8318.

(81) Tainter, C. J.; Pieniazek, P. A.; Lin, Y.-S.; Skinner, J. L. “Robust Three-Body Water Simulation Model”. *The Journal of Chemical Physics* **2011**, *134*, 184501.

(82) Tainter, C. J.; Shi, L.; Skinner, J. L. “Reparametrized E3B (Explicit Three-Body) Water Model Using the TIP4P/2005 Model as a Reference”. *Journal of Chemical Theory and Computation* **2015**, *11*, 2268–2277.

(83) Kann, Z. R.; Skinner, J. L. “A Scaled-Ionic-Charge Simulation Model That Reproduces Enhanced and Suppressed Water Diffusion in Aqueous Salt Solutions”. *The Journal of Chemical Physics* **2014**, *141*, 104507.

(84) Benavides, A. L.; Portillo, M. A.; Chamorro, V. C.; Espinosa, J. R.; Abascal, J. L. F.; Vega, C. “A Potential Model for Sodium Chloride Solutions Based on the TIP4P/2005 Water Model”. *The Journal of Chemical Physics* **2017**, *147*, 104501.

(85) Kohn, W.; Sham, L. J. “Self-Consistent Equations Including Exchange and Correlation Effects”. *Physical Review* **1965**, *140*, A1133–A1138.

(86) Car, R.; Parrinello, M. “Unified Approach for Molecular Dynamics and Density-Functional Theory”. *Physical Review Letters* **1985**, *55*, 2471–2474.

(87) Hafner, J. “Ab-Initio Simulations of Materials Using VASP: Density-functional Theory and Beyond”. *Journal of Computational Chemistry* **2008**, *29*, 2044–2078.

(88) Hu, C.; Ting, S.-W.; Chan, K.-Y.; Huang, W. “Reaction Pathways Derived from DFT for Understanding Catalytic Decomposition of Formic Acid into Hydrogen on Noble Metals”. *International Journal of Hydrogen Energy* **2012**, *37*, 15956–15965.

(89) Habibi, P.; Saji, T. H. G.; Vlugt, T. J. H.; Moultos, O. A.; Dey, P. “Hydrogen Dissociation in Li-decorated Borophene and Borophene Hydride: An Ab-Initio Study”. *Applied Surface Science* **2022**, *603*, 154323.

(90) Perdew, J. P.; Burke, K.; Ernzerhof, M. “Generalized Gradient Approximation Made Simple”. *Physical Review Letters* **1996**, *77*, 3865–3868.

(91) Klimeš, J.; Michaelides, A. “Perspective: Advances and Challenges in Treating van Der Waals Dispersion Forces in Density Functional Theory”. *The Journal of Chemical Physics* **2012**, *137*, 120901.

(92) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. “Van Der Waals Density Functional for General Geometries”. *Physical Review Letters* **2004**, *92*, 246401.

(93) Román-Pérez, G.; Soler, J. M. “Efficient Implementation of a van Der Waals Density Functional: Application to Double-Wall Carbon Nanotubes”. *Physical Review Letters* **2009**, *103*, 096102.

(94) Sun, J.; Ruzsinszky, A.; Perdew, J. P. “Strongly Constrained and Appropriately Normed Semilocal Density Functional”. *Physical Review Letters* **2015**, *115*, 036402.

(95) Peng, H.; Yang, Z.-H.; Perdew, J. P.; Sun, J. “Versatile van Der Waals Density Functional Based on a Meta-Generalized Gradient Approximation”. *Physical Review X* **2016**, *6*, 041005.

(96) Del Bene, J.; Pople, J. A. “Intermolecular Energies of Small Water Polymers”. *Chemical Physics Letters* **1969**, *4*, 426–428.

(97) Santra, B.; DiStasio, R. A.; Martelli, F.; Car, R. “Local Structure Analysis in Ab Initio Liquid Water”. *Molecular Physics* **2015**, *113*, 2829–2841.

(98) Chen, M.; Ko, H.-Y.; Remsing, R. C.; Calegari Andrade, M. F.; Santra, B.; Sun, Z.; Selloni, A.; Car, R.; Klein, M. L.; Perdew, J. P.; Wu, X. “Ab Initio Theory and Modeling of Water”. *Proceedings of the National Academy of Sciences* **2017**, *114*, 10846–10851.

(99) Corsetti, F.; Artacho, E.; Soler, J. M.; Alexandre, S. S.; Fernández-Serra, M.-V. “Room Temperature Compressibility and Diffusivity of Liquid Water from First Principles”. *The Journal of Chemical Physics* **2013**, *139*, 194502.

(100) Lee, H.-S.; Tuckerman, M. E. “Dynamical Properties of Liquid Water from Ab Initio Molecular Dynamics Performed in the Complete Basis Set Limit”. *The Journal of Chemical Physics* **2007**, *126*, 164501.

(101) Lyubartsev, A. P.; Laasonen, K.; Laaksonen, A. “Hydration of Li+ Ion. An Ab Initio Molecular Dynamics Simulation”. *The Journal of Chemical Physics* **2001**, *114*, 3120–3126.

(102) Petersen, C. P.; Gordon, M. S. “Solvation of Sodium Chloride: An Effective Fragment Study of NaCl(H2O)n”. *The Journal of Physical Chemistry A* **1999**, *103*, 4162–4166.

(103) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. “Ab Initio Molecular Dynamics Simulation of the Solvation and Transport of H3O+ and OH- Ions in Water”. *The Journal of Physical Chemistry* **1995**, *99*, 5749–5752.

(104) Tuckerman, M. E.; Chandra, A.; Marx, D. “Structure and Dynamics of OH-(Aq)”. *Accounts of Chemical Research* **2006**, *39*, 151–158.

(105) Alber, M.; Buganza Tepole, A.; Cannon, W. R.; De, S.; Dura-Bernal, S.; Garikipati, K.; Karniadakis, G.; Lytton, W. W.; Perdikaris, P.; Petzold, L.; Kuhl, E. “Integrating Machine Learning and Multiscale Modeling—Perspectives, Challenges, and Opportunities in the Biological, Biomedical, and Behavioral Sciences”. *npj Digital Medicine* **2019**, *2*, 1–11.

(106) Peng, G. C. Y.; Alber, M.; Buganza Tepole, A.; Cannon, W. R.; De, S.; Dura-Bernal, S.; Garikipati, K.; Karniadakis, G.; Lytton, W. W.; Perdikaris, P.; Petzold, L.; Kuhl, E. “Multiscale Modeling Meets Machine Learning: What Can We Learn?” *Archives of Computational Methods in Engineering* **2021**, *28*, 1017–1037.

(107) Stephenson, D.; Kermode, J. R.; Lockerby, D. A. “Accelerating Multiscale Modelling of Fluids with On-the-Fly Gaussian Process Regression”. *Microfluidics and Nanofluidics* **2018**, *22*, 139.

(108) Rocha, I. B. C. M.; Kerfriden, P.; van der Meer, F. P. “On-the-Fly Construction of Surrogate Constitutive Models for Concurrent Multiscale Mechanical Analysis through Probabilistic Machine Learning”. *Journal of Computational Physics: X* **2021**, *9*, 100083.

(109) Wang, H.; Yang, W. “Force Field for Water Based on Neural Network”. *The Journal of Physical Chemistry Letters* **2018**, *9*, 3232–3240.

(110) Gomes-Filho, M. S.; Pereira, A. O.; Feliciano, G. T.; Pedroza, L. S.; Coutinho-Neto, M. D. “Extending the Applicability of Popular Force Fields for Describing Water/Metal Interfaces: Application to Water/Pd(111)”. *Physica Scripta* **2022**, *98*, 015009.

(111) Devanathan, R.; Venkatnathan, A.; Rousseau, R.; Dupuis, M.; Frigato, T.; Gu, W.; Helms, V. “Atomistic Simulation of Water Percolation and Proton Hopping in Nafion Fuel Cell Membrane”. *The Journal of Physical Chemistry B* **2010**, *114*, 13681–13690.

(112) Zhang, W.; van Duin, A. C. T. “Second-Generation ReaxFF Water Force Field: Improvements in the Description of Water Density and OH-Anion Diffusion”. *The Journal of Physical Chemistry B* **2017**, *121*, 6021–6032.

(113) Han, Y.; Jiang, D.; Zhang, J.; Li, W.; Gan, Z.; Gu, J. “Development, Applications and Challenges of ReaxFF Reactive Force Field in Molecular Simulations”. *Frontiers of Chemical Science and Engineering* **2016**, *10*, 16–38.

(114) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; van Duin, A. C. T. “The ReaxFF Reactive Force-Field: Development, Applications and Future Directions”. *npj Computational Materials* **2016**, *2*, 1–14.

(115) Morawietz, T.; Singraber, A.; Dellago, C.; Behler, J. “How van Der Waals Interactions Determine the Unique Properties of Water”. *Proceedings of the National Academy of Sciences* **2016**, *113*, 8368–8373.

(116) Behler, J. “Perspective: Machine Learning Potentials for Atomistic Simulations”. *The Journal of Chemical Physics* **2016**, *145*, 170901.

(117) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. “Quantum-Chemical Insights from Deep Tensor Neural Networks”. *Nature Communications* **2017**, *8*, 13890.

(118) Panagiotopoulos, A. Z.; Yue, S. “Dynamics of Aqueous Electrolyte Solutions: Challenges for Simulations”. *The Journal of Physical Chemistry B* **2023**, *127*, 430–437.

(119) Kresse, G.; Hafner, J. “Ab Initio Molecular Dynamics for Liquid Metals”. *Physical Review B* **1993**, *47*, 558–561.

(120) Kresse, G.; Furthmüller, J. “Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set”. *Computational Materials Science* **1996**, *6*, 15–50.

(121) Kresse, G.; Furthmüller, J. “Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set”. *Physical Review B* **1996**, *54*, 11169–11186.

(122) Kresse, G.; Joubert, D. “From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method”. *Physical Review B* **1999**, *59*, 1758–1775.

(123) Jinnouchi, R.; Karsai, F.; Kresse, G. “On-the-Fly Machine Learning Force Field Generation: Application to Melting Points”. *Physical Review B* **2019**, *100*, 014105.

(124) Jinnouchi, R.; Lahnsteiner, J.; Karsai, F.; Kresse, G.; Bokdam, M. “Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference”. *Physical Review Letters* **2019**, *122*, 225701.

(125) Novikov, I. S.; Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. “The MLIP Package: Moment Tensor Potentials with MPI and Active Learning”. *Machine Learning: Science and Technology* **2020**, *2*, 025002.

(126) Wang, H.; Zhang, L.; Han, J.; E, W. “DeePMD-kit: A Deep Learning Package for Many-Body Potential Energy Representation and Molecular Dynamics”. *Computer Physics Communications* **2018**, *228*, 178–184.

(127) Jinnouchi, R.; Karsai, F.; Kresse, G. “Making Free-Energy Calculations Routine: Combining First Principles with Machine Learning”. *Physical Review B* **2020**, *101*, 060201.

(128) Mortazavi, B.; Shojaei, F.; Shapeev, A. V.; Zhuang, X. “A Combined First-Principles and Machine-Learning Investigation on the Stability, Electronic, Optical, and Mechanical Properties of Novel C6N7-based Nanoporous Carbon Nitrides”. *Carbon* **2022**, *194*, 230–239.

(129) Mortazavi, B.; Novikov, I. S.; Shapeev, A. V. “A Machine-Learning-Based Investigation on the Mechanical/Failure Response and Thermal Conductivity of Semiconducting BC2N Monolayers”. *Carbon* **2022**, *188*, 431–441.

(130) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; E, W. “DP-GEN: A Concurrent Learning Platform for the Generation of Reliable Deep Learning Based Potential Energy Models”. *Computer Physics Communications* **2020**, *253*, 107206.

(131) Giannozzi, P. et al. “Advanced Capabilities for Materials Modelling with Quantum ESPRESSO”. *Journal of Physics: Condensed Matter* **2017**, *29*, 465901.

(132) Pitike, K. C.; Setyawan, W. “Accurate Fe–He Machine Learning Potential for Studying He Effects in BCC-Fe”. *Journal of Nuclear Materials* **2023**, *574*, 154183.

(133) Wu, Y.; Liu, D.; Chu, W.; Wang, B.; Vasenko, A. S.; Prezhdo, O. V. “Fluctuations at Metal Halide Perovskite Grain Boundaries Create Transient Trap States: Machine Learning Assisted Ab Initio Analysis”. *ACS Applied Materials & Interfaces* **2022**, *14*, 55753–55761.

(134) Li, P.; Ding, F. “Origin of the Herringbone Reconstruction of Au(111) Surface at the Atomic Scale”. *Science Advances* **2022**, *8*, eabq2900.

(135) He, R.; Wu, H.; Lu, Y.; Zhong, Z. “Origin of Negative Thermal Expansion and Pressure-Induced Amorphization in Zirconium Tungstate from a Machine-Learning Potential”. *Physical Review B* **2022**, *106*, 174101.

(136) Liu, X.; Peng, R.; Sun, Z.; Liu, J. “Moiré Phonons in Magic-Angle Twisted Bilayer Graphene”. *Nano Letters* **2022**, *22*, 7791–7797.

(137) Zhang, C.; Tang, F.; Chen, M.; Xu, J.; Zhang, L.; Qiu, D. Y.; Perdew, J. P.; Klein, M. L.; Wu, X. “Modeling Liquid Water by Climbing up Jacob’s Ladder in Density Functional Theory Facilitated by Using Deep Neural Network Potentials”. *The Journal of Physical Chemistry B* **2021**, *125*, 11444–11456.

(138) Zhao, W.; Qiu, H.; Guo, W. “A Deep Neural Network Potential for Water Confined in Graphene Nanocapillaries”. *The Journal of Physical Chemistry C* **2022**, *126*, 10546–10553.

(139) Yang, M.; Bonati, L.; Polino, D.; Parrinello, M. “Using Metadynamics to Build Neural Network Potentials for Reactive Events: The Case of Urea Decomposition in Water”. *Catalysis Today* **2022**, *387*, 143–149.

(140) Mondal, A.; Kussainova, D.; Yue, S.; Panagiotopoulos, A. Z. “Modeling Chemical Reactions in Alkali Carbonate–Hydroxide Electrolytes with Deep Learning Potentials”. *Journal of Chemical Theory and Computation* **2022**, *19*, 4584–4595.

(141) Zhang, C.; Yue, S.; Panagiotopoulos, A. Z.; Klein, M. L.; Wu, X. “Dissolving Salt Is Not Equivalent to Applying a Pressure on Water”. *Nature Communications* **2022**, *13*, 822.

(142) Leberman, R.; Soper, A. K. “Effect of High Salt Concentrations on Water Structure”. *Nature* **1995**, *378*, 364–366.

(143) Mancinelli, R.; Botti, A.; Bruni, F.; Ricci, M. A.; Soper, A. K. “Hydration of Sodium, Potassium, and Chloride Ions in Solution and the Concept of Structure Maker/Breaker”. *The Journal of Physical Chemistry B* **2007**, *111*, 13570–13577.

(144) Zhang, L.; Wang, H.; Muniz, M. C.; Panagiotopoulos, A. Z.; Car, R.; E, W. “A Deep Potential Model with Long-Range Electrostatic Interactions”. *The Journal of Chemical Physics* **2022**, *156*, 124107.

(145) Malosso, C.; Zhang, L.; Car, R.; Baroni, S.; Tisi, D. “Viscosity in Water from First-Principles and Deep-Neural-Network Simulations”. *npj Computational Materials* **2022**, *8*, 1–10.

(146) Delft High Performance Computing Centre (DHPC) *DelftBlue Supercomputer (Phase 1)*, 2022.

(147) Parr, R. G.; Yang, W., *Density-Functional Theory of Atoms and Molecules*; Oxford University Press, USA: 1994; 344 pp.

(148) Perdew, J. P.; Schmidt, K. “Jacob’s Ladder of Density Functional Approximations for the Exchange-Correlation Energy”. *AIP Conference Proceedings* **2001**, *577*, 1–20.

(149) Lieb, E. H.; Oxford, S. “Improved Lower Bound on the Indirect Coulomb Energy”. *International Journal of Quantum Chemistry* **1981**, *19*, 427–439.

(150) Hammer, B.; Hansen, L. B.; Nørskov, J. K. “Improved Adsorption Energetics within Density-Functional Theory Using Revised Perdew-Burke-Ernzerhof Functionals”. *Physical Review B* **1999**, *59*, 7413–7421.

(151) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. “Accurate and Numerically Efficient r2SCAN Meta-Generalized Gradient Approximation”. *The Journal of Physical Chemistry Letters* **2020**, *11*, 8208–8215.

(152) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. “Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories”. *Chemical Reviews* **2010**, *110*, 5023–5063.

(153) Tsuzuki, S.; Lüthi, H. P. “Interaction Energies of van Der Waals and Hydrogen Bonded Systems Calculated Using Density Functional Theory: Assessing the PW91 Model”. *The Journal of Chemical Physics* **2001**, *114*, 3949–3957.

(154) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. “Dispersion Corrections to Density Functionals for Water Aromatic Interactions”. *The Journal of Chemical Physics* **2004**, *120*, 2693–2699.

(155) Allen, M. J.; Tozer, D. J. “Helium Dimer Dispersion Forces and Correlation Potentials in Density Functional Theory”. *The Journal of Chemical Physics* **2002**, *117*, 11113–11120.

(156) Grimme, S. “Accurate description of van der Waals complexes by density functional theory including empirical corrections”. *Journal of Computational Chemistry* **2004**, *25*, 1463–1473.

(157) Grimme, S. “Semiempirical GGA-type density functional constructed with a long-range dispersion correction”. *Journal of Computational Chemistry* **2006**, *27*, 1787–1799.

(158) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. “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* **2010**, *132*, 154104.

(159) Grimme, S.; Ehrlich, S.; Goerigk, L. “Effect of the Damping Function in Dispersion Corrected Density Functional Theory”. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

(160) Caldeweyher, E.; Bannwarth, C.; Grimme, S. “Extension of the D3 Dispersion Coefficient Model”. *The Journal of Chemical Physics* **2017**, *147*, 034112.

(161) Gossel, G. H.; Lacombe, L.; Maitra, N. T. “On the Numerical Solution of the Exact Factorization Equations”. *The Journal of Chemical Physics* **2019**, *150*, 154112.

(162) Vydrov, O. A.; Van Voorhis, T. “Nonlocal van Der Waals Density Functional: The Simpler the Better”. *The Journal of Chemical Physics* **2010**, *133*, 244103.

(163) Klimeš, J.; Bowler, D. R.; Michaelides, A. “Van Der Waals Density Functionals Applied to Solids”. *Physical Review B* **2011**, *83*, 195131.

(164) Klimeš, J.; Bowler, D. R.; Michaelides, A. “Chemical Accuracy for the van Der Waals Density Functional”. *Journal of Physics: Condensed Matter* **2009**, *22*, 022201.

(165) Sabatini, R.; Gorni, T.; de Gironcoli, S. “Nonlocal van Der Waals Density Functional Made Simple and Efficient”. *Physical Review B* **2013**, *87*, 041108.

(166) Ning, J.; Kothakonda, M.; Furness, J. W.; Kaplan, A. D.; Ehlert, S.; Brandenburg, J. G.; Perdew, J. P.; Sun, J. “Workhorse Minimally Empirical Dispersion-Corrected Density Functional with Tests for Weakly Bound Systems: r2SCAN+rVV10”. *Physical Review B* **2022**, *106*, 075422.

(167) Wiktor, J.; Ambrosio, F.; Pasquarello, A. “Note: Assessment of the SCAN+rVV10 Functional for the Structure of Liquid Water”. *The Journal of Chemical Physics* **2017**, *147*, 216101.

(168) Kothakonda, M.; Kaplan, A. D.; Isaacs, E. B.; Bartel, C. J.; Furness, J. W.; Ning, J.; Wolverton, C.; Perdew, J. P.; Sun, J. “Testing the r2SCAN Density Functional for the Thermodynamic Stability of Solids with and without a van Der Waals Correction”. *ACS Materials Au* **2023**, *3*, 102–111.

(169) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. “PACKMOL: A package for building initial configurations for molecular dynamics simulations”. *Journal of Computational Chemistry* **2009**, *30*, 2157–2164.

(170) Verlet, L. “Computer ”Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”. *Physical Review* **1967**, *159*, 98–103.

(171) Jinnouchi, R.; Miwa, K.; Karsai, F.; Kresse, G.; Asahi, R. “On-the-Fly Active Learning of Interatomic Potentials for Large-Scale Atomistic Simulations”. *The Journal of Physical Chemistry Letters* **2020**, *11*, 6946–6955.

(172) Bartók, A. P.; Kondor, R.; Csányi, G. “On Representing Chemical Environments”. *Physical Review B* **2013**, *87*, 184115.

(173) MacKay, D. J. C. “Bayesian Interpolation”. *Neural Computation* **1992**, *4*, 415–447.

(174) Vaspwiki *ML_ICRITERIA* Vaspwiki, `https://www.vasp.at/wiki/index.php/ML_ICRITERIA` (accessed 07/05/2023).

(175) Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*; Oxford University Press: 2017; 641 pp.

(176) Hoover, W. G. “Canonical Dynamics: Equilibrium Phase-Space Distributions”. *Physical Review A* **1985**, *31*, 1695–1697.

(177) Nosé, S. “A Unified Formulation of the Constant Temperature Molecular Dynamics Methods”. *The Journal of Chemical Physics* **1984**, *81*, 511–519.

(178) Nosé, S. “Constant Temperature Molecular Dynamics Methods”. *Progress of Theoretical Physics Supplement* **1991**, *103*, 1–46.

(179) Forster-Tonigold, K.; Groß, A. “Dispersion Corrected RPBE Studies of Liquid Water”. *The Journal of Chemical Physics* **2014**, *141*, 064501.

(180) Vaspwiki *Vasp Tutorials: Melting Silicon* Part 1: Melting silicon, https://www.vasp.at/wiki/index.php/ML_CX (accessed 07/05/2023).

(181) Lebigot, E. O. *Welcome to the Uncertainties Package — Uncertainties Python Package 3.0.1 Documentation* Uncertainties, https://pythonhosted.org/uncertainties/ (accessed 07/26/2023).

(182) Daivis, P. J.; Evans, D. J. “Comparison of Constant Pressure and Constant Volume Nonequilibrium Simulations of Sheared Model Decane”. *The Journal of Chemical Physics* **1994**, *100*, 541–547.

(183) Evans, D. J.; Morriss, G. P., *Statistical Mechanics of Nonequilibrium Liquids*; ANU Press: 1990.

(184) Helfand, E. “Transport Coefficients from Dissipation in a Canonical Ensemble”. *Physical Review* **1960**, *119*, 1–9.

(185) Yeh, I.-C.; Hummer, G. “System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions”. *The Journal of Physical Chemistry B* **2004**, *108*, 15873–15879.

(186) Celebi, A. T.; Jamali, S. H.; Bardow, A.; Vlugt, T. J. H.; Moultos, O. A. “Finite-Size Effects of Diffusion Coefficients Computed from Molecular Dynamics: A Review of What We Have Learned so Far”. *Molecular Simulation* **2021**, *47*, 831–845.

(187) Dünweg, B.; Kremer, K. “Molecular Dynamics Simulation of a Polymer Chain in Solution”. *The Journal of Chemical Physics* **1993**, *99*, 6983–6997.

(188) Moultos, O. A.; Zhang, Y.; Tsimpanogiannis, I. N.; Economou, I. G.; Maginn, E. J. “System-Size Corrections for Self-Diffusion Coefficients Calculated from Molecular Dynamics Simulations: The Case of CO2, n-Alkanes, and Poly(Ethylene Glycol) Dimethyl Ethers”. *The Journal of Chemical Physics* **2016**, *145*, 074109.

(189) Jamali, S. H.; Hartkamp, R.; Bardas, C.; Söhl, J.; Vlugt, T. J. H.; Moultos, O. A. “Shear Viscosity Computed from the Finite-Size Effects of Self-Diffusivity in Equilibrium Molecular Dynamics”. *Journal of Chemical Theory and Computation* **2018**, *14*, 5959–5968.

(190) Hu, Z.; Jiang, J. “Assessment of Biomolecular Force Fields for Molecular Dynamics Simulations in a Protein Crystal”. *Journal of Computational Chemistry* **2010**, *31*, 371–380.

(191) Zhang, Y.; Maginn, E. J. “Direct Correlation between Ionic Liquid Transport Properties and Ion Pair Lifetimes: A Molecular Dynamics Study”. *The Journal of Physical Chemistry Letters* **2015**, *6*, 700–705.

(192) Humbert, M. T.; Zhang, Y.; Maginn, E. J. “PyLAT: Python LAMMPS Analysis Tools”. *Journal of Chemical Information and Modeling* **2019**, *59*, 1301–1305.

(193) Gullbrekken, Ø.; Røe, I. T.; Selbach, S. M.; Schnell, S. K. “Charge Transport in Water–NaCl Electrolytes with Molecular Dynamics Simulations”. *The Journal of Physical Chemistry B* **2023**, *127*, 2729–2738.

(194) Blazquez, S.; Abascal, J. L. F.; Lagerweij, J.; Habibi, P.; Dey, P.; Vlugt, T. J. H.; Moultos, O. A.; Vega, C. “Computation of Electrical Conductivities of Aqueous Electrolyte Solutions: Two Surfaces, One Property”. *Journal of Chemical Theory and Computation* **2023**, DOI: 10.1021/acs.jctc.3c00562.

(195) Ramasubramani, V.; Dice, B. D.; Harper, E. S.; Spellings, M. P.; Anderson, J. A.; Glotzer, S. C. “Freud: A Software Suite for High Throughput Analysis of Particle Simulation Data”. *Computer Physics Communications* **2020**, *254*, 107275.

(196) The HDF Group *Hierarchical Data Format, Version 5* https://www.hdfgroup.org/HDF5/ (accessed 12/06/2023).

(197) 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* **2007**, *19*, 335206.

(198) Benedict, W. S.; Gailar, N.; Plyler, E. K. “Rotation-Vibration Spectra of Deuterated Water Vapor”. *The Journal of Chemical Physics* **1956**, *24*, 1139–1165.

(199) Thomsen, B.; Shiga, M. “Ab Initio Study of Nuclear Quantum Effects on Sub- and Supercritical Water”. *The Journal of Chemical Physics* **2021**, *155*, 194107.

(200) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. “Benchmark Oxygen-Oxygen Pair-Distribution Function of Ambient Water from x-Ray Diffraction Measurements with a Wide Q-range”. *The Journal of Chemical Physics* **2013**, *138*, 074506.

(201) Gaiduk, A. P.; Gygi, F.; Galli, G. “Density and Compressibility of Liquid Water and Ice from First-Principles Simulations with Hybrid Functionals”. *The Journal of Physical Chemistry Letters* **2015**, *6*, 2902–2908.

(202) Wang, J.; Román-Pérez, G.; Soler, J. M.; Artacho, E.; Fernández-Serra, M.-V. “Density, Structure, and Dynamics of Water: The Effect of van Der Waals Interactions”. *The Journal of Chemical Physics* **2011**, *134*, 024516.

(203) Guo, Y.-j.; Xu, H.-b.; Guo, F.; Zheng, S.-l.; Zhang, Y. “Density and Viscosity of Aqueous Solution of K2CrO4/KOH Mixed Electrolytes”. *Transactions of Nonferrous Metals Society of China* **2010**, *20*, s32–s36.

(204) Muñoz-Santiburcio, D. “Accurate Diffusion Coefficients of the Excess Proton and Hydroxide in Water via Extensive Ab Initio Simulations with Different Schemes”. *The Journal of Chemical Physics* **2022**, *157*, 024504.

(205) Kühne, T. D.; Krack, M.; Parrinello, M. “Static and Dynamical Properties of Liquid Water from First Principles by a Novel Car-Parrinello-like Approach”. *Journal of Chemical Theory and Computation* **2009**, *5*, 235–241.

(206) Willow, S. Y.; Salim, M. A.; Kim, K. S.; Hirata, S. “Ab Initio Molecular Dynamics of Liquid Water Using Embedded-Fragment Second-Order Many-Body Perturbation Theory towards Its Accurate Property Prediction”. *Scientific Reports* **2015**, *5*, 14358.

(207) Marcus, Y. “Ionic Radii in Aqueous Solutions”. *Chemical Reviews* **1988**, *88*, 1475–1498.

(208) Galamba, N. “Mapping Structural Perturbations of Water in Ionic Solutions”. *The Journal of Physical Chemistry B* **2012**, *116*, 5242–5250.

(209) Megyes, T.; Bálint, S.; Grósz, T.; Radnai, T.; Bakó, I.; Sipos, P. “The Structure of Aqueous Sodium Hydroxide Solutions: A Combined Solution x-Ray Diffraction and Simulation Study”. *The Journal of Chemical Physics* **2008**, *128*, 044501.

(210) Imberti, S.; Botti, A.; Bruni, F.; Cappa, G.; Ricci, M. A.; Soper, A. K. “Ions in Water: The Microscopic Structure of Concentrated Hydroxide Solutions”. *The Journal of Chemical Physics* **2005**, *122*, 194509.

(211) Bruni, F.; Ricci, M. A.; Soper, A. K. “Structural Characterization of NaOH Aqueous Solution in the Glass and Liquid States”. *The Journal of Chemical Physics* **2001**, *114*, 8056–8063.

(212) Botti, A.; Bruni, F.; Imberti, S.; Ricci, M. A.; Soper, A. K. “Ions in Water: The Microscopic Structure of Concentrated NaOH Solutions”. *The Journal of Chemical Physics* **2004**, *120*, 10154–10162.

(213) Gilliam, R. J.; Graydon, J. W.; Kirk, D. W.; Thorpe, S. J. “A Review of Specific Conductivities of Potassium Hydroxide Solutions for Various Concentrations and Temperatures”. *International Journal of Hydrogen Energy* **2007**, *32*, 359–364.

(214) McCall, D. W.; Douglass, D. C. “The Effect of Ions on the Self-Diffusion of Water. I. Concentration Dependence”. *The Journal of Physical Chemistry* **1965**, *69*, 2001–2011.

(215) Yuan-Hui, L.; Gregory, S. “Diffusion of Ions in Sea Water and in Deep-Sea Sediments”. *Geochimica et Cosmochimica Acta* **1974**, *38*, 703–714.

(216) Jamali, S. H.; Wolff, L.; Becker, T. M.; de Groen, M.; Ramdin, M.; Hartkamp, R.; Bardow, A.; Vlugt, T. J. H.; Moultos, O. A. “OCTP: A Tool for On-the-Fly Calculation of Transport Properties of Fluids with the Order-n Algorithm in LAMMPS”. *Journal of Chemical Information and Modeling* **2019**, *59*, 1290–1294.

A

Results at Experimental Density

The results at experimental density (MLMD 100 ps ρ_{Exp}) are compared with the results at the calculated density (MLMD 100 ps). The experimental density is 1048 kg m^{-3} and the calculated equilibrium density is 1335 kg m^{-3}. Although the structural and transport properties at experimental density perform better compared to experimental results, the average pressure during the simulations is $(-12.95 \pm 0.07) \times 10^3$ bar.

Table A.1: The produced structural results of the KOH simulations at experimental density (MLMD 100 ps ρ_{Exp}) compared to the results at the calculated equilibrium density (MLMD 100 ps). $d_{i\text{O}}$ indicates the shortest typical distance to a water molecule from the center of molecule type i. The oxygen atoms are used as the center of H_2O and OH^- molecules. The coordination numbers are calculated using equation (2.9) with an r_{shell} of 4, 3.5, and 3.3 Å for n_{OO}, $n_{\text{K}^+\text{O}}$ and $n_{\text{OH}^-\text{O}}$, respectively. Except for the density, all other relevant simulation settings are shown in table 2.1 and the simulations are performed using the RPBE-D3 DF.

	$d_{\text{OO}}/[\text{Å}]$	n_{OO}	$d_{\text{K}^+\text{O}}/[\text{Å}]$	$n_{\text{K}^+\text{O}}$	$d_{\text{OH}^-\text{O}}/[\text{Å}]$	$n_{\text{OH}^-\text{O}}$
MLMD 100 ps	2.758 ± 0.004	13.386 ± 0.007	2.818 ± 0.004	8.73 ± 0.03	2.601 ± 0.006	6.33 ± 0.06
MLMD 100 ps ρ_{Exp}	2.776 ± 0.002	10.12 ± 0.02	2.843 ± 0.003	6.80 ± 0.08	2.59 ± 0.01	4.80 ± 0.09
Experimental	2.8[208]	-	2.79 ± 0.08[207]	6-8[207]	2.67 ± 0.07[209]	3-5[209]

Table A.2: The produced transport properties of the KOH simulations at experimental density (MLMD 100 ps ρ_{Exp}) compared to the results at the calculated equilibrium density (MLMD 100 ps), showing the viscosity (ν) and self-diffusivity of H_2O, K^+, and OH^- ($D^{\infty}_{\text{H}_2\text{O}}$, $D^{\infty}_{\text{K}^+}$, and $D^{\infty}_{\text{OH}^-}$) in the thermodynamic limit. The electric conductivity σ is calculated using the Nernst-Einstein relation shown in equation (2.13). Except for the density, all other relevant simulation settings are shown in table 2.1 and the simulations are performed using the RPBE-D3 DF.

	$\nu/[\text{mPa s}]$	$D^{\text{corr}}/[\times 10^{-9}\ \text{m}^2\ \text{s}^{-1}]$	$D^{\infty}_{\text{H}_2\text{O}}/[\times 10^{-9}\ \text{m}^2\ \text{s}^{-1}]$
MLMD 100 ps	8.64 ± 0.07	0.0695 ± 0.0005	1.16 ± 0.08
MLMD 100 ps ρ_{Exp}	6.8 ± 0.2	0.081 ± 0.003	2.43 ± 0.03
Experimental	1.005[203]	-	2.45[214]

	$D^{\infty}_{\text{K}^+}/[\times 10^{-9}\ \text{m}^2\ \text{s}^{-1}]$	$D^{\infty}_{\text{OH}^-}/[\times 10^{-9}\ \text{m}^2\ \text{s}^{-1}]$	$\sigma_{\text{NE}}/[\text{S m}^{-1}]$
MLMD 100 ps	0.6 ± 0.1	6 ± 2	29 ± 7
MLMD 100 ps ρ_{Exp}	1.79 ± 0.03	6 ± 2	29 ± 6
Experimental	1.96[215]	5.27[215]	26[213]

B

MLMD 100 ps: properties over time

Some properties of the 100 ps MLMD production runs that are sampled over time are presented in this appendix. The system state (energies, temperature and pressure) are illustrated together in appendix B.1. The atom index of the oxygen atom in the OH^- molecule is shown in appendix B.2. From this, the reaction rate, as well as the two different options after reactions can be seen. Appendix B.3 visualizes the MSDs, used to determine the self-diffusion coefficients.

B.1. System State Over Time

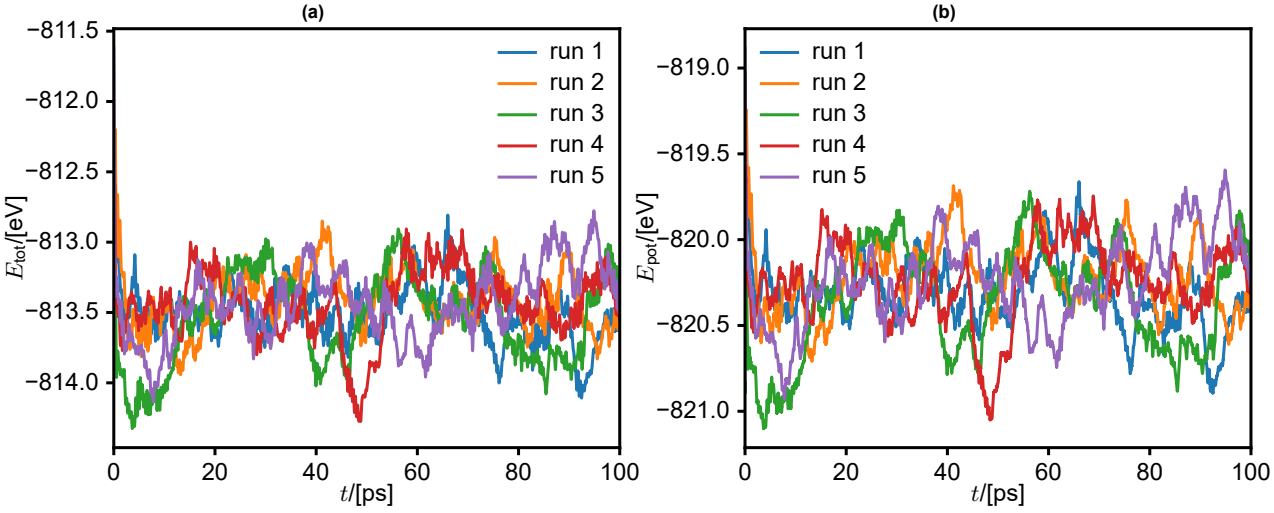

Figure B.1: The (a) total (E_{tot}) and (b) potential (E_{pot}) energies of the 100 ps MLMD production runs as a function of time (t). A Gaussian moving average filter with a width of 2.5 ps is used to smoothen the fluctuations and visualize possible energy drifts. All simulations show an energy drift in the first 5 ps, after which natural fluctuations occur.

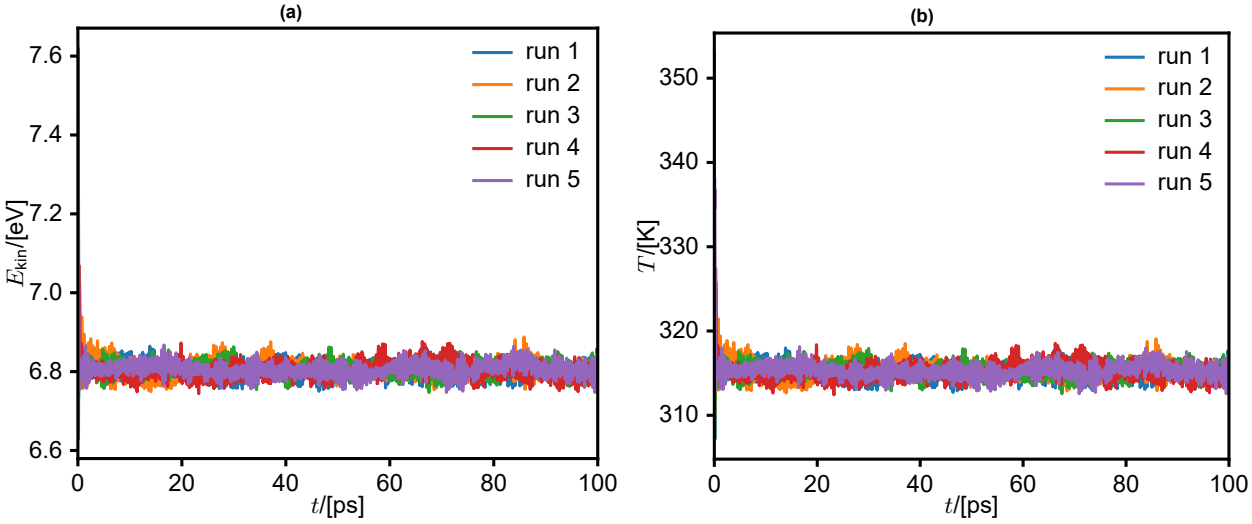

Figure B.2: The (a) kinetic energy (E_{kin}) and (b) temperature (T) of the 100 ps MLMD production runs as a function of time (t). A Gaussian moving average filter with a width of 2.5 ps is used to smoothen the fluctuations and visualize possible drifts. The kinetic energy shows no drift and is equivalent to the temperature.

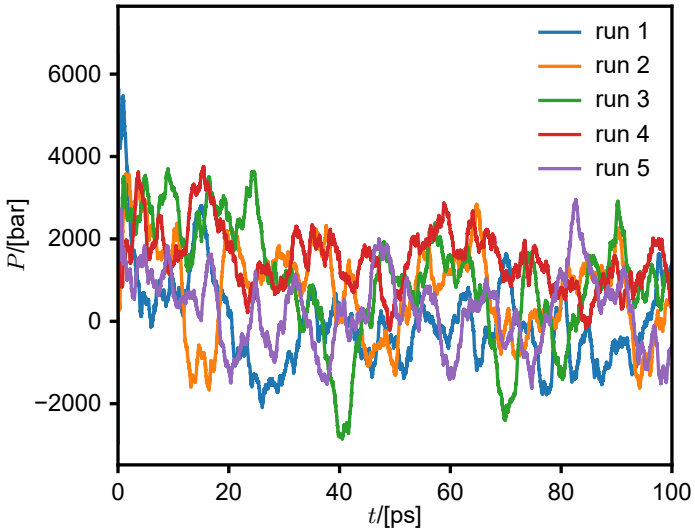

Figure B.3: The pressure (P) of the 100 ps MLMD production runs as a function of time (t). A Gaussian moving average filter with a width of 2.5 ps is used to smoothen the fluctuations and visualize possible drifts.

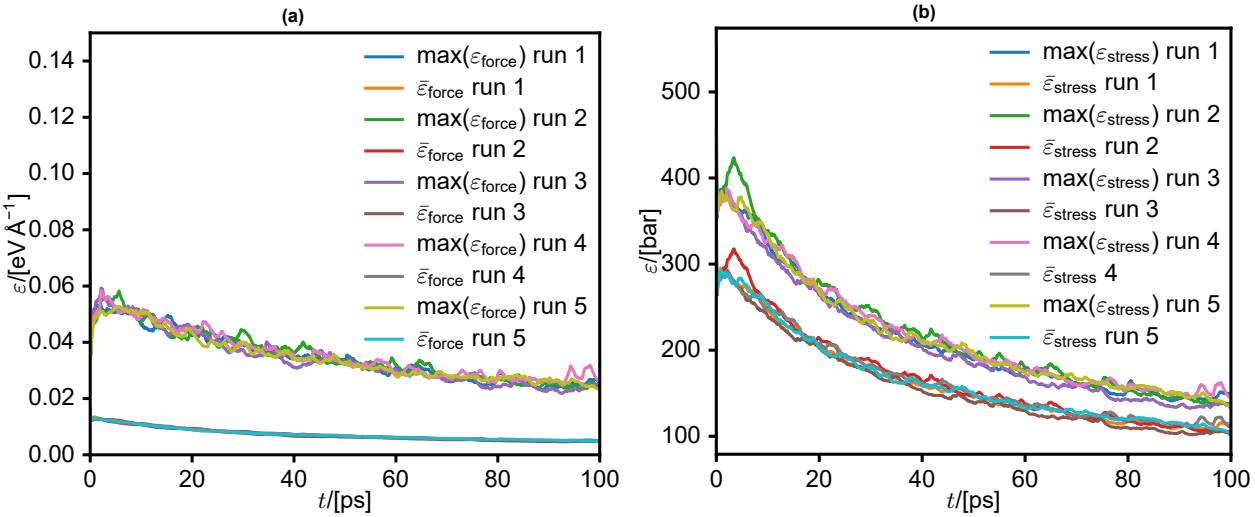

Figure B.4: The Bayesian error estimate of (a) the atomic forces (ε_{force}) and (b) the stress tensor components (ε_{stress}) as a function of time (t). Both figures show the maximum and average value.

B.2. Tracking the Atomic Index of the Hydroxide

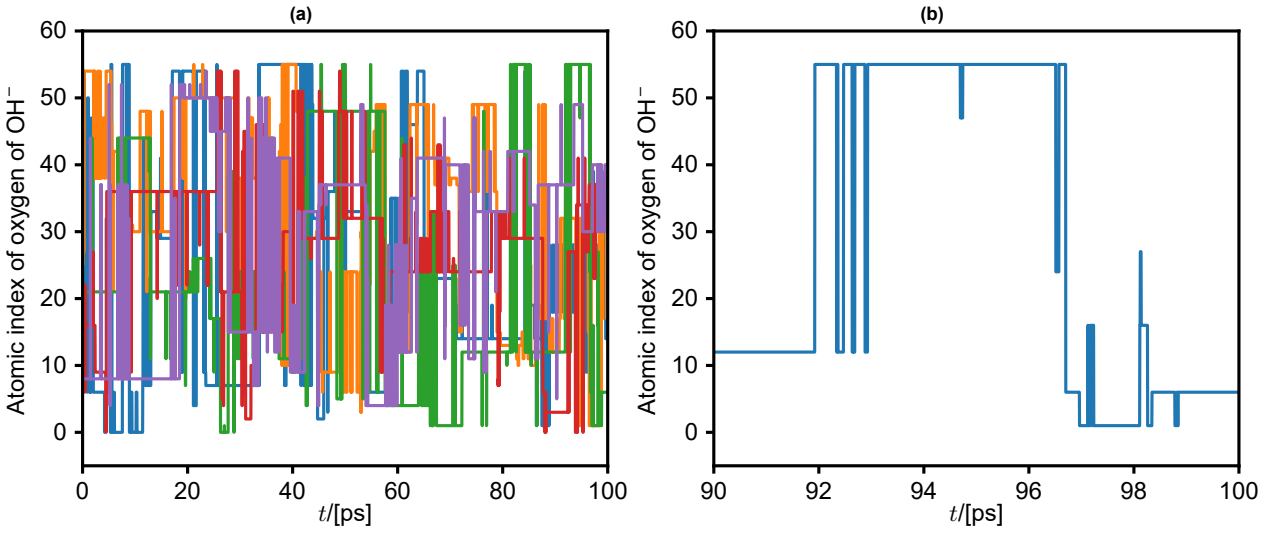

Figure B.5: The atomic index of the oxygen atom in the OH^- molecule as a function of time (t). (a) visualizes the extracted atom index of all runs and (b) shows the last 10 ps of the third run. Note that (b) shows long periods of inactivity followed by a burst of reactions. These two distinctive phases are described by Tuckerman et al. [104], and also found in this work.

B.3. MSDs of Potassium, Hydroxide and Water

In this section, the MSDs of the tracked particles are visualized. The self-diffusion coefficients are derived using these MSDs. To ensure that the results are calculated in the diffusion regime (and not the ballistic regime), lines with a slope of 1 in log-log space are fitted through this data, with the same approach as is used with the OCTP[216] post-processing code, which is developed for Blazquez et al. [194]. Note that the MSD of OH^- does not show the characteristic slope of the ballistic regime. This is caused by the proton transfer events, which occur within a few time steps. The simulation box size (L) is 10.9202 Å. Therefore, the ideal lower limit of the MSDs which the self-diffusion is fitted to is 120 $Å^2$ (L^2). Note that the MSDs of OH^- barely reaches this limit and the MSDs of K^+ clearly stays below this. This might introduce errors in the calculated self-diffusion coefficients. The MSDs of H_2O exceed this limit.

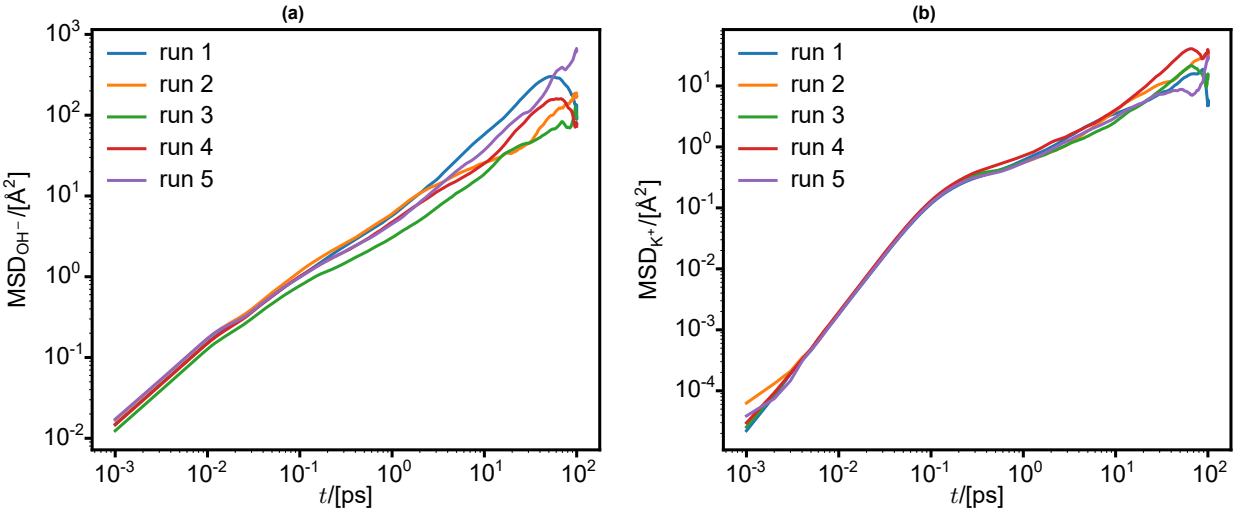

Figure B.6: The MSDs of (a) OH^- and (b) K^+. A windowed algorithm using the Freud[195] python package is used to retrieve these from the atomic positions.

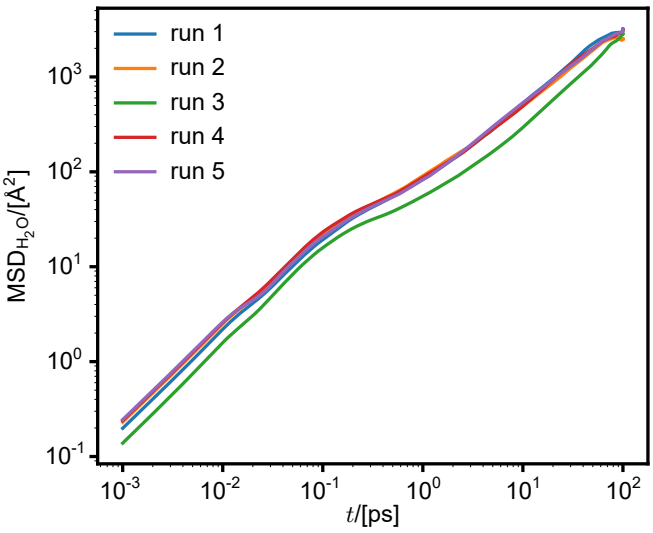

Figure B.7: The MSDs of H_2O. A windowed algorithm using the Freud[195] python package is used to retrieve these from the atomic positions.