Ghent University

Model-Informed Training Data Curation for Reactive All-Atom Potentials

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Science: Physics

Author:	Loïc Dumortier
Promotors:	Prof. Dr. ir. Toon Verstraelen Dr. Jelle Vekeman Dr. Theodorus De Bruin Dr. Benoît Creton
Department Chair:	Prof. Dr. Jan Ryckebusch
Academic year:	2024-2025

Center for Molecular Modeling
Department of Physics and Astronomy
Faculty of Sciences

23 april 2025

Members of the Examination Committee

Chair

Prof. Dr. Bartel Van Waeyenberge	Ghent University

Reading Committee

Prof. Dr. Jean-Francois Paul	Université de Lille
Prof. Dr. Jonathan Leliaert	Ghent University
Prof. Dr. ir. Sven Rogge	Ghent University
Dr. Jenna Mancuso	Ghent University

Promotors

Prof. Dr. ir. Toon Verstraelen	Ghent University
Dr. Jelle Vekeman	Ghent University
Dr. Theodorus De Bruin	IFP Energies Nouvelles
Dr. Benoît Creton	IFP Energies Nouvelles

“”

–

Contents

Preface

As I begin presenting this work, the result of nearly five years of effort, I wanted to share a few reflections. My passion for science goes back to early childhood, sparked by encyclopedias and books about the stars that captured my imagination. Becoming a scientist was always the dream, though I couldn't have truly imagined finding myself where I am today. But getting here was certainly not a solo journey.

First and foremost, I want to extend my sincere gratitude to Toon for his fantastic supervision throughout this project. I deeply appreciate the time you consistently make for your students—something truly valuable and never to be taken for granted. I could always count on your theoretical insights and technical expertise to guide me through challenges and brainstorm ideas.

My thanks also go to my co-supervisors, Benoit, Theo, and Jelle. Your willingness to help whenever I needed it was invaluable.

Reflecting on my time in Rueil-Malmaison, I want to thank everyone who made my stay so enjoyable and fun. A huge thanks to the entire department of molecular modeling and thermodynamics at IFP. I consistently felt welcomed and accepted, even though connecting in person wasn't always simple during that period. Special thanks to Rosa, who was always up for a walk during a time when even that felt borderline illicit! And cheers to all the doctoral students at IFP School for the great parties at the annual doctoral days.

For my time in Ghent, my gratitude goes out to all my colleagues at CMM for creating such a wonderfully pleasant and stimulating work environment. I genuinely enjoyed coming to work and our time together at CMM, especially our after-work events. Special shout-outs to Jelle, Nikitha, and Lukas, with whom I shared the majority of my office hours. Thanks to Siebe and Ruben for the (far too frequent, but always welcome) work interruptions! And to Sven, who I had the distinct privilege of waving to multiple times a day—the wining and dining was always thoroughly enjoyed! A big thank you to Jenna for tagging along (or letting me tag along) to countless music events and festivals!

I'd also like to thank those who helped by reading drafts and offering comments: Lukas, your feedback was much appreciated.

My heartfelt thanks go to my girlfriend, Flo. You consistently offered comfort and stood firmly by my side during what was possibly (and hopefully remains!) the most stressful period of my life.

Lastly, and most significantly, I want to thank my parents. You have been my bedrock for all 28 years of my life. I honestly wouldn't have achieved the smallest fraction of my dreams without your unconditional love and unwavering support. I'll repeat the words I wrote in my Master's thesis acknowledgements, hoping you can read them this time: You are the greatest parents a son could ever wish for.

–Loïc

Summary

This PhD thesis focuses on the development of predictive models, known as reactive potentials, used extensively in Molecular Dynamics. These potentials aim to simulate chemical reactions accurately and efficiently at the atomic scale, making it possible to study complex systems and processes over length and time scales that are computationally intractable for higher-accuracy ab initio methods. This task presents challenges, particularly concerning the high cost and scarcity of the high-quality reference data needed to build these models. The work concentrates on pioneering "model-informed training data curation" strategies to enhance the development of two major types of reactive potentials: the empirical Reactive Force Fields (ReaxFF) and the data-driven Machine Learning Interatomic Potentials (MLIPs). Two primary methodological advancements within this curation framework are presented in this thesis, demonstrating techniques tailored for ReaxFF and MLIPs, respectively, but holding potential for broader application.

Firstly, it presents the "Balanced Loss" cost function, a novel approach designed to streamline the complex task of ReaxFF parameter optimization. Traditional optimization methods rely on manually weighting the diverse and often imbalanced data types found in training sets, such as energies, bond lengths, and angles. Balanced Loss reformulates this by allowing users to specify physically meaningful "tolerances," representing the desired accuracy for different data categories. Its Log-Sum-Exp formulation automatically focuses the optimization on minimizing errors in the categories deviating most from their tolerances, providing direct feedback on the model's capabilities and making the process more systematic and less reliant on user intuition. This technique was successfully applied to reparameterize ReaxFF for simulating water adsorption on alumina, yielding a force field with markedly improved accuracy.

Secondly, the thesis details an efficient active learning framework aimed at automating the creation of training datasets for dedicated MLIPs. This work focuses on building tailored, more efficient MLIPs for specific chemical systems. The proposed active learning strategy employs Kernel Density Estimation on predicted atomic "node energies" to intelligently identify the most valuable new data points required to refine the MLIP. This framework facilitated the development of a dedicated MACE potential for the complex hydrated silicate ionic liquid system. It demonstrated that such tailored, actively learned MLIPs can achieve computational performance exceeding even traditional ReaxFF for large-scale reactive Molecular Dynamics simulations.

In essence, this research provides innovative, systematic, and data-efficient methodologies applicable within and beyond the immediate context of reactive potential development. By offering the Balanced Loss function and a node energy-based active learning scheme, it equips the field with valuable tools to advance the field of reactive Molecular Dynamics simulations. The Balanced Loss function shows potential for broader multi-objective optimization problems, while the active learning approach might be adapted for other models involving atomic energy decomposition. As such, they make the development of accurate and efficient models more rational and accessible. The utility of these methods is validated through successful applications to complex, relevant chemical systems: alumina-water interactions and hydrated silicate ionic liquids.

Samenvatting

Dit proefschrift richt zich op de ontwikkeling van voorspellende modellen, bekend als reactieve potentialen, die veelvuldig worden gebruikt in moleculaire dynamica. Deze potentialen hhebben als doel het nauwkeurig en efficiënt simuleren van chemische reacties op atomair niveau, wat het mogelijk maakt om complexe systemen te bestuderen over tijd- en systeemschalen dat computationeel niet haalbaar zijn met ab initio methodes. Deze taak brengt uitdagingen met zich mee, met name wat betreft de hoge kosten en schaarste van de hoogwaardige referentiedata die nodig zijn om deze modellen te bouwen. Het werk concentreert zich op baanbrekende strategieën voor "model-geïnformeerde trainingsdata-curatie" om de ontwikkeling van twee belangrijke types reactieve potentialen te verbeteren: de empirische Reactive Force Fields (ReaxFF) en de data-gedreven Machine Learning Interatomic Potentials (MLIPs). Twee primaire methodologische vorderingen binnen dit curatie-framework worden in dit proefschrift gepresenteerd, waarbij technieken worden gedemonstreerd die respectievelijk zijn toegespitst op ReaxFF en MLIPs, maar die potentieel breder toepasbaar zijn.

Ten eerste presenteert het de "Balanced Loss" kostenfunctie, een nieuwe aanpak ontworpen om de complexe taak van ReaxFF parameteroptimalisatie te stroomlijnen. Traditionele optimalisatiemethoden vertrouwen op het handmatig wegen van de diverse en vaak ongebalanceerde datatypes in trainingssets, zoals energieën, bindingslengtes en hoeken. Balanced Loss herformuleert dit door gebruikers in staat te stellen fysisch betekenisvolle "toleranties" te specificeren, die de gewenste nauwkeurigheid voor verschillende datacategorieën vertegenwoordigen. De Log-Sum-Exp formulering focust de optimalisatie automatisch op het minimaliseren van fouten in de categorieën die het meest afwijken van hun toleranties wat directe feedback geeft over de capaciteiten van het model, het proces systematischer maakt en minder afhankelijk van de intuïtie van de gebruiker. Deze techniek werd succesvol toegepast om ReaxFF te herparametriseren voor het simuleren van wateradsorptie op aluminiumoxide, wat resulteerde in een krachtveld met aanzienlijk verbeterde nauwkeurigheid.

Ten tweede beschrijft het proefschrift een efficiënt kader voor actief leren, gericht op het automatiseren van de creatie van trainingsdatasets voor specifieke MLIPs. Dit werk richt zich op het bouwen van op maat gemaakte, effi-

ciëntere MLIPs voor specifieke chemische systemen. De voorgestelde actieve leerstrategie maakt gebruik van Kernel Density Estimation op voorspelde atomaire “node energieën” om intelligent de meest waardevolle nieuwe datapunten te identificeren die nodig zijn om de MLIP te verfijnen. Dit kader faciliteerde de ontwikkeling van een specifieke MACE potentiaal voor het complexe gehydrateerd silicaat ionische vloeistof systeem. Het toonde aan dat dergelijke op maat gemaakte, via actief leren ontwikkelde MLIPs computationele prestaties kunnen bereiken die zelfs traditionele ReaxFF potentialen overtreffen voor grootschalige reactieve moleculaire dynamica simulaties.

In essentie biedt dit onderzoek innovatieve, systematische en data-efficiënte methodologieën die toepasbaar zijn binnen en buiten de directe context van de ontwikkeling van reactieve potentialen. Door de Balanced Loss functie en een op “node energieën” gebaseerd actief leren schema aan te bieden, voorziet dit werk het vakgebied van waardevolle tools om het terrein van reactieve moleculaire dynamica simulaties vooruit te helpen. De Balanced Loss functie toont potentieel voor bredere multi-objectief optimalisatieproblemen, terwijl de actief leren-aanpak mogelijk aangepast kan worden voor andere modellen die gebruikmaken van atomaire energie-decompositie. Als zodanig maken ze de ontwikkeling van nauwkeurige en efficiënte modellen rationeler en toegankelijker. Het nut van deze methoden wordt gevalideerd door succesvolle toepassingen op complexe, relevante chemische systemen: aluminiumoxide-water interacties en gehydrateerde silicaat ionische vloeistoffen.

List of Symbols

$\boldsymbol{R}$	Configuration vector
$\boldsymbol{x}$	Vector containing degrees of freedom of system
E	Energy
$\hat{H}$	Hamilton operator
Ψ	Many-body wave function
ρ	Electron density
T	Kinetic energy
v_{ext}	External potential
v_{eff}	Correction interaction potential
v_{XC}	Exchange-correlation potential
ψ	Kohn-Sham orbital
ε_i	Kohn-Sham orbital energy
$\hbar$	Reduced Planck's constant
BO'	Uncorrected bond order
$\boldsymbol{p}$	ReaxFF parameter vector
$\mathcal{L}$	Loss function
P_a	Acceptance probability
$m^{(g)}$	Mean of generation g
$C^{(g)}$	Covariance of generation g
λ	Population size
$\mathcal{U}$	uniform distribution
$\mathcal{N}$	normal distribution
$\mathcal{O}$	Computational complexity in big O notation
$\boldsymbol{h}_i^{(t)}$	Feature vector for atom i at layer t
z_i	Atomic number of atom i
$\boldsymbol{r}_i$	Position of atom i
d_{ij}	Distance between i and j
r_{cut}	Cutoff distance
$\mathcal{K}$	Collection of atoms within cutoff distance of an atom
M_t	Learnable message function
$\boldsymbol{m}_i^{(t)}$	Message at layer t arriving at atom i
U_t	Learnable update function

R_t	Learnable readout function
$D(g)$	Matrix representation of rotation g
$Y_{\ell m}$	Spherical Harmonics
$L_{\max}$	Maximum degree order

List of Abbreviations

ACSFs	Atom-Centered Symmetry Functions
BL	Balanced Loss
BO	Bond order
CMA-ES	Covariance Matrix Adaptation Evolutionary Strategy
D3(BJ)	Becke-Jonhson Dispersion Correction
dDsC	Density-dependent Dispersion Correction
DFT	Density Functional Theory
FF	Force Field
MD	Molecular Dynamics
MCFF	Monte Carlo Force Field
GA	Genetic Algorithms
GGA	Generalized Gradient Approximation
GPU	Graphics Processing Unit
HF	Hartree-Fock
HSIL	Hydrated Silicate Ionic Liquid
ISI	Identity-Sum-Identity
KDE	Kernel Density Estimation
LDA	Local Density Approximation
LSE	Log-Sum-Exp
MLIP	Machine Learning Interatomic Potential
MLP	Machine learning Potential
NMR	Nuclear Magnetic Resonance
OOD	Out-Of-Distribution
PBE	Perdew-Berke-Ernzerhof Functional
QM	Quantum Mechanical
QM/MM	Quamtum Mechanical / Molecular Mechanics
RDF	Radial Distribution Function
RMSE	Root-Mean-Square Error
RSS	Root-Sum-Squars
SCF	Self-Consistent Field
SOAP	Smooth Overlap of Atomic Positions
SOPPE	Successive One-Parameter Parabolic Extrapolation
SSE	Sum-of-Squarres Error
SWA	Stochastic Weight Averaging
vdW	van der Waals

Part I

Model-Informed Training Data Curation for Reactive All-Atom Potentials

1

Introduction

"If you want to find the secrets of the universe,
think in terms of energy, frequency and vibration."
– Nikola Tesla

Understanding and predicting the fundamental nature of matter is a cornerstone of modern science. Human curiosity has always driven us to understand the world around us, from the vast cosmos down to the smallest constituents of matter. For centuries, this quest relied on observation, philosophical reasoning, and increasingly, careful experimentation. Landmark discoveries provided frameworks—fundamental laws—that allowed us to predict how objects move and interact, transforming our understanding from purely descriptive to quantitatively predictive. This pursuit of knowledge, blending insight with rigorous measurement, laid the foundation for modern science.

The advent of the digital computer introduced a shift in how science could be approached. Suddenly, complex equations derived from the fundamental laws of quantum mechanics, describing how atoms and electrons interact, could be solved numerically—calculations previously intractable by hand. Phenomena that were difficult or impossible to probe directly in the lab could be simulated at the atomic scale. This computational revolution opened a new window onto the atomic and molecular world. It became possible to build "virtual laboratories" to watch atoms move, interact, and rearrange, which gained the scientific community unprecedented insights into the microscopic mechanisms.

One of the most fascinating and challenging frontiers in this domain is understanding the dynamics of chemical reactions: the process by which molecules transform, bonds break, and new ones form. Capturing the intricate choreography of atoms during a chemical reaction is crucial for fields ranging from materials science and catalysis to drug discovery and environmental chemistry. Computer simulations hold great promise here. Specifically, computational models built using 'reactive potentials' or 'reactive force fields'—mathematical functions designed explicitly to describe the energy changes associated with chemical bond breaking and formation—offer the potential to unravel complex reaction networks, design new catalysts, or predict the stability of novel materials from the bottom up.

However, realizing this potential faces a fundamental tension. On the one hand, we desire simulations based on fundamental principles, such as ab initio methods derived directly from the laws of quantum mechanics without empirical input. These methods offer high accuracy by faithfully replicating the subtle quantum mechanics governing atomic interactions. On the other hand, we need simulations that are computationally feasible, capable of handling the large numbers (e.g., thousands) of atoms involved in realistic systems and spanning timescales long enough to observe meaningful events, like a slow catalytic process or material degradation. Achieving both simultaneously remains a formidable challenge. The most accurate ab initio simulation methods are often computationally prohibitive for large, complex systems. This necessitates the development of interatomic potentials, such as classical force fields (FF) or machine learning interatomic potentials (MLIPs), which approximate the interactions based on functional forms fitted to reference data. Empirical approaches offer significant computational speed-up but may sacrifice accuracy, especially when describing the complexities of chemical bond changes inherent to reactive systems.

Focusing now on these empirical interatomic potentials the challenge of balancing accuracy and efficiency becomes deeply intertwined with the data required to build these sophisticated simulation models. Unlike ab initio methods, which derive directly from first principles without needing extensive pre-existing data for their construction, these empirical approaches must be "taught" how atoms interact, especially during reactions, using reliable reference information This often takes the form of highly accurate results derived from solving the equations of quantum mechanics for specific atomic arrangements (often referred to simply as ‘quantum calculations’ or QM calculations), or alternatively, precise experimental measurements of energies, forces, and structures for relevant atomic configurations. Here, the field of reactive molecular modeling through empirical potentials diverges sharply from many other modern scientific domains awash in “big data”. Unlike astronomical surveys generating terabytes of observational data, or genomic studies benefiting from low-cost sequencing, generating high-fidelity reference data for reactive chemistry via computationally expensive QM calculations is painstaking. Each single data point represents a significant investment of resources. This data requirement contrasts sharply with ab initio methods, which, by definition, do not rely on pre-existing data for their construction.

This reality of expensive, scarce data for empirical reactive models brings us to the central questions motivating this thesis: When high-quality data is the bottleneck for building these models, how do we proceed most effectively? Which specific atomic arrangements should we invest our computational budget in calculating to gain the most useful information? How can we design strategies to extract the maximum possible knowledge from every precious data point we generate? Simply put, how can we be smarter and more efficient in gathering the reference data needed to build better empirical reactive models?

This work explores solutions to this critical problem through the development and application of Model-informed Training Data Curation methodologies. The core idea is to move beyond brute-force data generation towards a more intelligent, iterative process where insights gleaned from the simulation models themselves actively guide the selection and refinement of the training data. By developing systematic strategies to identify the most informative data points and to effectively balance the diverse types of information needed, we aim to make the entire process of building accurate and efficient reactive potentials more rational, tractable, and ultimately, more successful. This thesis presents novel techniques designed to address these data curation challenges, paving the way for more powerful simulations of complex, reactive chemical systems. While demonstrated here for reactive potentials, the underlying principles possess broader applicability. They can be applicable to a broad range of interatomic potentials where the total energy prediction can be decomposed to atomic contributions, and even to multi-variate optimization problems in general.

Goal and Outline

This thesis focuses on advancing the methodology for creating accurate and efficient reactive all-atom potentials, specifically addressing the challenges in training data curation and model development. The primary goal is to develop and demonstrate novel techniques that improve the parameterization of Reactive Force Fields and the training of MLIPs for complex chemical systems.

Following this introductory chapter, **Chapter 2: Molecular Modeling**, provides the necessary theoretical background. It introduces the landscape of computational modeling methods, discussing the trade-offs between accuracy and computational cost, and details key methods including Ab Initio approaches like Density Functional Theory, interatomic potentials like classical Force Fields, Reactive Force Fields (e.g. ReaxFF), and MLIPs.

Chapter 3: Molecular Modeling Applications, then introduces the specific chemical systems that are investigated in this thesis: Alumina (Al_2O_3), discussing its structure, surface properties, catalytic relevance, and role as a binder, and Hydrated Silicate Ionic Liquids (HSILs), exploring their role in zeolite synthesis.

The core research contributions are presented in the subsequent chapters. **Chapter 4** corresponds to the first publication titled **Paper I: Balanced Loss**, focuses on ReaxFF development. It introduces the “Balanced Loss” function, a novel method to address the challenges of weighting heterogeneous and imbalanced training data during ReaxFF parameter optimization. The effectiveness of this approach is demonstrated through a reparameterization of ReaxFF for water adsorption on alumina. **Chapter 5** corresponds to the first publication titled **: Paper II: Node Energy Density Active Learning**, details the second major contribution, concentrating on MLIP development. It introduces an efficient active learning framework using node energy density estimation to automate training set generation for dedicated MLIPs like MACE. This methodology is applied to the complex HSIL system, showing that tailored MLIPs can achieve superior computational efficiency compared to traditional ReaxFF potentials for large reactive simulations.

Finally, **Chapter 6: Conclusions & Outlook**, summarizes the key findings and discusses potential future research directions. The thesis concludes with Appendices containing supplementary materials such as (pre-)publications, conference contributions, software details, and supporting information for the papers.

2

Molecular Modeling

"...if I were again beginning my studies,
I should follow the advice of Plato and start with mathematics."
– Galileo Galilei

In the quest to understand and predict chemical and physical behavior at the atomic scale, computational chemists and physicists have developed a wide range of modeling techniques. Each method strikes a balance between accuracy, computational cost, and the spatial and temporal scales it can reach. The choice of method is therefore not arbitrary, it depends critically on the type of system under study, the properties of interest, and the level of approximation one is willing to accept.

Fundamentally, these approaches diverge based on their reliance on pre-existing data or parameters. Ab initio (Latin for 'from the beginning') methods aim to predict properties directly from the fundamental laws of quantum mechanics without relying on empirical input fitted to experimental or prior computational results. In contrast, interatomic potentials, often referred to as force field-based approaches, retain an atomistic description of matter but use mathematical functions with fitted parameters, specific to each atom type and interaction, to approximate the physical behavior. These parameters are determined by adjusting them until the model reproduces reference data, such as experimental measurements or results from high-level ab initio calculations. The effectiveness of such an empirical force field depends strongly on the quality and scope of this reference data and how well the fitted parameters capture the relevant chemistry of the system.

This distinction underlies the spectrum of available techniques for molecular modeling: from the quantum-mechanical precision of ab initio calculations, capable of describing just a few atoms with high accuracy, to the major simplifications of coarse-graining models, which represent groups of atoms as single entities.[1] In between lies a rich variety of techniques, each grounded in different assumptions and suited to different problems. The accuracy-efficiency tradeoff is illustrated in Figure 2.1.

Despite their differences, all these approaches share a common goal: to map the positions of atoms in space, denoted by the configuration vector $\boldsymbol{R}$, to a corresponding potential energy E:

$$\boldsymbol{R} = (\boldsymbol{r_1}, \boldsymbol{r_2}, ..., \boldsymbol{r_N}) \in \mathbb{R}^{3\mathrm{N}} \tag{2.1}$$

$$f : \boldsymbol{R} \to E \tag{2.2}$$

Once the energy is known, thermodynamic properties can be derived through its gradients and higher-order derivatives. [2,3]

This work centers specifically on the development and application of advanced empirical reactive potentials. In order to position the research chapters in this all-atom modeling landscape, the following sections guide the reader through this hierarchy of methods. We begin with a brief overview of ab initio approaches, with particular attention to Density Functional Theory (DFT). Next, we turn to interatomic potentials, introducing their general structure before delving deeper into Reactive Force Fields. Finally, we explore the emerging class of MLIPs, which promise to combine the accuracy of quantum methods with the speed of classical force fields.

2.1 Quantum Mechanical Electronic Structure Methods

In computational chemistry and materials science, ab initio methods refer to a class of modeling approaches that aim to predict the properties of matter directly from fundamental physical principles, without relying on empirical parameters. These methods are rooted in quantum mechanics, specifically in solving the many-body Schrödinger equations for electrons and nuclei in a given system.

The time-independent, non-relativistic Schrödinger equation for a molecular system is expressed as:

$$\hat{H}\Psi(x, R) = E\Psi(x, R) \tag{2.3}$$

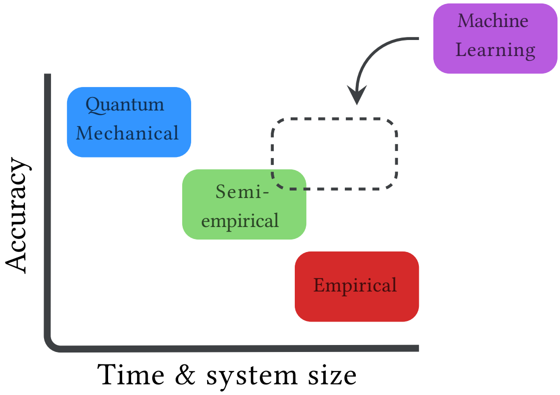

Figure 2.1: General overview of computational modeling methods. The plot illustrates the general trade-off between model accuracy and the achievable simulation time and system size, which correlates with computational cost. Recently developed machine learning potentials aim to disrupt this typical trade-off by providing, once trained, models that can approach QM accuracy while remaining computationally efficient enough for large systems and long timescales.

where $\hat{H}$ is the hamiltonian operator, $\Psi(x, R)$ is the many-body wavefunction depending on the electronic coordinates and spin x and nuclear coordinates R, and E is the total energy of the system. Solving this equation analytically for systems with more than 2 bodies is not possible, and numerically intractable beyond a dozen particles due to the exponential scaling of the wavefunction's complexity with the number of electrons. To make the approach computationally feasible, ab initio approaches use a series of approximations. The most common first step, is the Born-Oppenheimer approximation, which decouples the motion of the heavy nuclei from the electrons.

Not only the nuclei are subject to approximations; most ab initio methodologies approximate the solution of the electronic Schrödinger equation as well. One of the simplest methods that imposes the anti-symmetry requirement of the electronic wavefunction is the Hartree-Fock (HF) method, which uses a single Slater determinant to represent the many-electron wavefunction. Because the exact wavefunction is a linear combination of Slater determinants, HF searches for solutions in a restricted space and overestimates the true ground state energy.[4,5] HF is said to neglect the correlation energy, which post-HF methods, such as Configuration Interaction and Coupled Cluster theories, aim to recover.[6,7] This family of wavefunction theory methods can systematically improve the accuracy of the correlation energy. However, a practical limitation of these methods is that their cost scales steeply with increasing number of electrons or one-particle orbitals, making them impractical for large systems.

DFT offers an alternative approach to approximate the correlation energy at a lower computational cost. It reformulates the problem of solving the electronic ground state energy in terms of the electron density rather than the full many-body wavefunction. Hohenberg and Kohn demonstrated that there is a one-to-one relation between the external potential $v_{\text{ext}}(r)$ and the electron density $\rho(r)$. [8] The ground-state energy is a unique functional of the external potential and thus also of the density. Density Functional Approximations (DFAs) model $E[\rho]$ directly without constructing the full electronic wavefunction as an intermediate step. This "shortcut" lies at the heart of DFT's efficiency, but the exact and explicit functional of the energy as a function of the density, i.e., without constructing the N-electron wavefunction as an intermediate step, is unknown.

In its strict sense, DFT can be applied without any wavefunction, but such DFAs turn out to be too crude to describe chemical bonding.[9] A fundamental difficulty is that DFAs of the kinetic energy make too large systematic errors. To address this limitation, the Kohn-Sham formalism reintroduces an auxiliary wavefunction into DFT, consisting of a single Slater determinant of non-interacting fermions. [10] The kinetic energy of the interacting system is approximated by T_s, the ground state energy of the wavefunction of the non-interacting fermions, subject to the constraint that its density matches that of the interacting system. The total energy functional $E[\rho]$ is then rewritten using T_s, grouping all the remaining quantum-mechanical terms (the difference between the true kinetic energy $T[\rho]$ and $T_s[\rho]$, plus all non-classical exchange and correlation effects) into the exchange-correlation functional $E_{\text{XC}}[\rho]$:

$$E[\rho] = T_s[\rho] + \int v_{\text{ext}}(r)\rho(r)dr + E_{\text{Hartree}}[\rho] + E_{\text{XC}}[\rho] \tag{2.4}$$

where $E_{\text{Hartree}}[\rho]$ is the classical Hartree repulsion energy.

Applying the variational principle to this energy expression with respect to the auxiliary orbitals $\psi_i(r)$ leads to the effective single-particle Kohn-Sham equations:

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2+v_{\mathrm{eff}}(r)\right)\psi_i(r)=\varepsilon_i\psi_i(r) \tag{2.5}$$

where ε_i are the Kohn-Sham orbital energies, and $v_{\mathrm{eff}}(r)$ is the effective potential experienced by the non-interacting electrons:

$$\begin{aligned} v_{\mathrm{eff}}(r) &= v_{\mathrm{ext}}(r)+v_{\mathrm{H}(r)}+v_{\mathrm{XC}}(r) \\ &= v_{\mathrm{ext}}(r)+\int\frac{\rho(r)}{|r-r'|}\mathrm{dr}+v_{\mathrm{XC}}(r) \end{aligned} \tag{2.6}$$

It consists of the external potential v_{ext}, the Hartree potential v_H, and the exchange-correlation potential v_{XC}, which implicitly contains all the non-trivial many-body quantum mechanics. In terms of computational cost and implementation, the Kohn-Sham equations are similar to HF, but they have two main advantages: (i) they are still the result of an exact theory and (ii) the correlation energy can be approximated by a computationally efficient functional of the electron density.

The electron density of the system is constructed from the occupied Kohn-Sham orbitals:

$$\rho(r)=\sum_{i=1}^{N}|\psi_i(r)|^2 \tag{2.7}$$

This formulation highlights an important implication: the Kohn-Sham orbitals themselves do not necessarily have a direct physical meaning. Their primary role is to approximate the kinetic energy and reconstruct the electronic density.

Since the Hartree and exchange-correlation potentials depend on the electron density $\rho(r)$, which in turn is calculated from the orbitals ψ_i that are solutions to the Kohn-Sham equations involving v_{eff}, an iterative approach is required. This self-consistency problem necessitates the use of a self-consistent field (SCF) procedure to find the final solution. The self-consistent field procedure, as shown in Figure 2.2, is initiated by a guess of the electron density $\rho(r)$. Subsequently, the Kohn-Sham hamiltonian $\hat{H}_{\mathrm{KS}}$ is constructed using this density. Solving the Kohn-Sham equations yields a new set of single-particle orbitals ψ_i and eigenvalues ε_i. From these orbitals, an updated electron density $\rho_{\text{out}}(r) = \sum_i^N |\psi_i(r)|^2$ is computed. This output density ρ_{out} represents the electronic distribution resulting from the potential generated by the input density ρ_{in}. The output density then becomes the new input density ρ_{in} in the next iteration of the SCF cycle.

This iterative process continues until self-consistency is achieved, meaning that the input density and the output density differ negligibly between iterations. Convergence can be monitored using the change in the density or potential directly, or using measures related to the residual of the Kohn-Sham equations. Frequently, multiple criteria must be satisfied simultaneously before the calculation is considered converged, for instance, requiring both the change in total energy and the change in the sum of Kohn-Sham eigenvalues to fall below predefined tolerances. The electron density can now be used to obtain physical and electronic properties of the system.

In DFT, evaluating the exact exchange-correlation energy is inherently complex, as it must encapsulate all the non-trivial QM effects of electron exchange and correlation. While the exact exchange-correlation functional

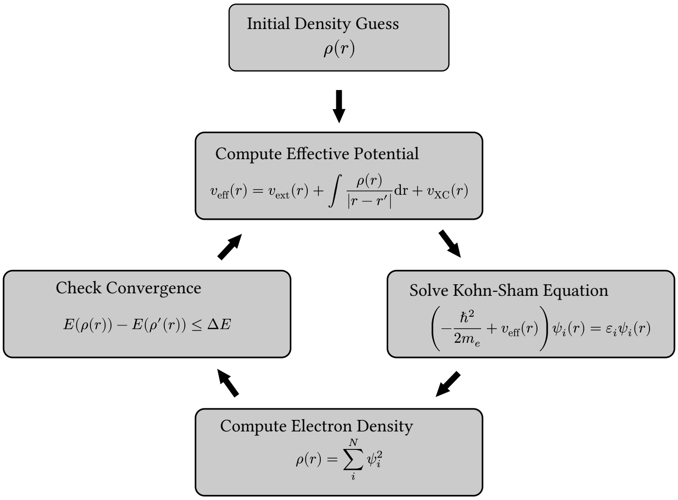

Figure 2.2: SCF loop in Kohn-Sham DFT. The process begins with an initial guess for the electron density $\rho(r)$, from which the effective potential $v_{\text{eff}}(r)$ is computed. The Kohn-Sham equations are then solved to obtain single-particle orbitals $\psi_{\text{i}}(r)$, which are used to update the electron density. The cycle repeats until convergence is achieved.

is formally defined, this formal definition does not provide a known, explicit mathematical expression solely in terms of the electron density. [11] Devising better approximations of the exact functional remains a central challenge to the field of DFT-based molecular modeling. Fortunately, even relatively simple approximations to the exchange-correlation functional have proven remarkably effective and often yield acceptable, or even highly accurate, results for a wide range of systems and properties. The first and simplest set of functionals is the Local Density Approximation (LDA). [12] LDA describes the exchange-correlation by approximating each point of the system by a homogeneous electron gas, with the same electron density at each point. [13] This approximation yields reasonable results for systems where the electron density is relatively uniform or slowly varying, such as in many simple metals. The inherent error cancellation in such systems sometimes masks the deficiencies of the local approximation.

In systems with significant density inhomogeneity, such as molecules, the LDA's assumption of local uniformity becomes problematic. The electron density changes rapidly, particularly in and around chemical bonding regions, which are critical for determining chemical properties. To address this, more sophisticated functionals have been developed. A widely used improvement is the Generalized Gradient Approximation (GGA), and one of its most popular implementations is the Perdew-Burke-Ernzerhof (PBE) functional. [14,15] GGA approaches improve upon LDA by incorporating not just the local density $\rho(r)$, but also its gradient $\nabla\rho(r)$. This makes GGAs semi-local: the inclusion of the gradient allows the functional to sense variations in density in the immediate vicinity of point r, better responding to the inhomogeneities found in molecules and materials. However, this dependence remains inherently local and is insufficient to capture long-range electron correlation effects, which manifest as dispersion forces (van der Waals interactions). It is also important to note that while foundational functionals like LDA and PBE (a popular GGA functional) are largely derived from theoretical principles (e.g., satisfying known exact constraints or based on the homogeneous electron gas model), many other widely used density functionals, particularly those favoured for high accuracy in molecular chemistry, incorporate parameters fitted to experimental or high-level computational data, making them empirical models. LDA and GGA stand out as important non-empirical rungs on the hierarchy of DFT approximations. Meta-GGAs add dependence on the kinetic energy density, hybrid functionals mix a portion of exact HF exchange with GGA or meta-GGA functionals (often significantly improving predictions for properties like reaction barriers), and double-hybrid functionals further incorporate correlation effects from wave function theory methods (like MP2). [16] While these higher-rung functionals often provide increased accuracy for specific chemical properties, they typically come with substantially higher computational cost and, in many cases, increased reliance on empirical parameterization. For generating the large datasets often required for fitting reactive potentials, the balance of reasonable accuracy and computational efficiency offered by GGAs like PBE remains highly relevant, provided their limitations are addressed.

Both LDA and GGA, including PBE, inherently lack long-range dispersion forces (van der Waals interactions), stemming from their (semi-)local nature. These interactions are essential for accurately describing weak interactions between distant fragments, such as physisorbed molecules on surfaces. DFT implementations overcome this functional limitation with additional empirical or semi-empirical dispersion correction terms that are applied post-hoc to the DFT total energy. Two commonly used approaches are:

- Grimme's D3 Correction (D3(BJ)): This method adds an atom-pairwise dispersion energy correction to the DFT total energy.[17–20] Additional damping, such as the Becke-Johnson damping scheme, ensures that these corrections do not interfere with short-range interactions captured by the base functional.
- Density-dependent dispersion correction (dDsC): This is an extension of the exchange-hole dipole model, which derives atomic C_6, C_8, and C_{10} coefficients from a Hirshfeld partitioning of the electron density.[21] dDsC modifies the partitioning scheme to dominant Hirshfeld and proposes a refined density-dependent model for short-range damping.[22] While this method has been successfully benchmarked across a wide range of systems, it also has some limitations. For example, dominant Hirshfeld (and thus also dDsC) is rather sensitive to the numerical quadrature grids in the DFT code. Moreover, the density dependence of the forces is neglected, potentially hampering geometry convergence of complex systems.

2.2 Interatomic Potentials

interatomic potentials describe molecular systems using classical physics rather than quantum mechanics, and are widely used to study large, many-particle systems efficiently. Unlike ab initio methods, interatomic potentials do not explicitly consider electrons[1] or solve the Schrödinger equation. Instead, each atom is treated as a single particle, and its key properties—mass, van der Waals radius, partial charge, and so on—are encoded directly in the model. [23] As a result, the explicit distinction between nuclei and electrons is lost, but their effects are embedded in the parameterization.

A defining feature of interatomic potentials is that they rely on *a priori* knowledge. That is, they are constructed by fitting to experimental data or results from higher-level calculations. This is in sharp contrast to ab initio methods, which attempt to predict molecular behavior from first principles, without relying on pre-existing data. Consequently, the accuracy and transferability of interatomic potentials depend heavily on the quality and scope of its training set. One fundamental consequence of this classical, parameterized approach is that chemical bonds become explicit entities in the model. This contrasts with QM descriptions like DFT, where bonds are emergent features of the electron density. In empirical potentials, bonds are modeled directly as mechanical constructs, typically as springs, whose energetic contributions depend on deviations from their equilibrium lengths or angles.

The total energy of a molecular system in Molecular Modeling is computed using a potential energy function, commonly referred to as a force field. The name "force field" comes from the fact that the negative gradient of the potential energy yields the force acting on each atom. Given the forces, the system can evolve by minimizing its energy or propagating in time through Molecular Dynamics (MD), where particle positions and velocities are updated based on Newton's laws of motion.

A typical force field includes both bonded and non-bonded terms:

$$E_{\text{total}} = \sum E_{\text{bonded}} + \sum E_{\text{non-bonded}} \tag{2.8}$$

Bonded terms capture interactions between atoms connected by chemical bonds and typically include bond stretching, angle bending and torsional

---

[1] There are exceptions, such as eReaxFF and eMLPs, which explicitly consider (some) electrons

rotations. Each of these is often represented by a harmonic potential of the form:

$$E(x) = \frac{k}{2}(x - x_{\mathrm{ref}})^2 \tag{2.9}$$

where k is a force constant, x is a geometric variable such as a bond lenght or angle, and x_{ref} is its equilibrium value.

Non-bonded terms model longer-range interactions between atoms that are not directly connected via bonds. [24] These include van der Waals forces, often modeled with a Lennard-Jones potential (shown in Figure 2.3):

$$E_{\mathrm{vdW}}(i, j) = 4\varepsilon_{\mathrm{ij}}\left(\left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{12} - \left(\frac{\sigma_{\mathrm{ij}}}{r_{\mathrm{ij}}}\right)^{6}\right) \tag{2.10}$$

where ε and σ are tunable parameters depending properties of particle i and j. Electrostatic interactions between partially-charged particles, according to Coulomb's law of elestrostatics:

$$E_{\mathrm{charge}}(i, j) = k_e \frac{q_i q_j}{r_{\mathrm{ij}}} \tag{2.11}$$

with k_e being Coulomb's constant.

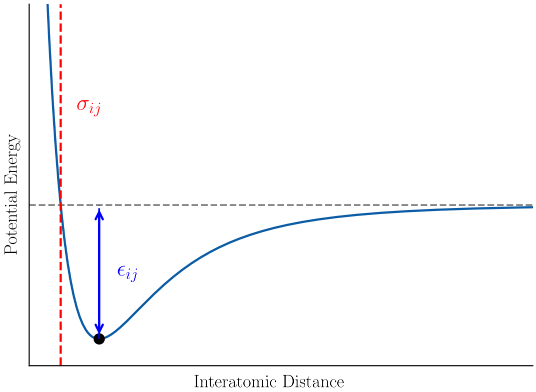

Figure 2.3: The Lennard-Jones potential, illustrating the balance between short-range repulsion ($\propto r^{-12}$) and long-range attraction ($\propto r^{-6}$). The equilibrium distance lies near the minimum of the potential well, where attractive and repulsive forces cancel. This functional form is commonly used in interatomic potentials to describe non-bonded van der Waals interactions.

Despite their simplicity, such force fields remain widely used in molecular simulations due to their efficiency. Well-established frameworks like CHARMM, AMBER, and the force fields employed in GROMACS rely on similar functional forms, but achieve accuracy through careful parameterization and extensive benchmarking. [25–28]

However, classical force fields have inherent limitations. Beyond their predefined bond topology, another significant challenge lies in accurately capturing the relative energies between different molecular conformations. This can be particularly difficult when conformations differ by changes in multiple dihedral angles simultaneously, as the standard, often simple, functional forms for dihedral terms may not fully represent the complexity of the true potential energy surface. Furthermore, classical force fields typically approximate interactions as pairwise additive, neglecting or only implicitly accounting for crucial many-body effects like electronic polarization and non-additive dispersion or exchange interactions, which can significantly influence conformational preferences. Another primary limitation of classical force field, especially relevant for this work, is their inability to describe chemical reactions. Since bonds are predefined and modeled using harmonic potentials, they cannot break or form dynamically. This makes standard interatomic potentials unsuitable for systems where reactivity is important.

2.3 Reactive Force Fields

To address the limitations of classical force fields in simulating chemical reactions, reactive force fields have been developed. These potentials employ specialized functional forms that allow chemical bonds to form and break dynamically during a simulation, based on the evolving local atomic environment. Influential early approaches, such as the Tersoff potential, utilized environment-dependent bond order (BO) functions to modulate interactions, finding success particularly in modeling covalent materials. [29–31] However, simulating complex chemistry across a much wider range of systems necessitated a more flexible and broadly transferable approach than the Tersoff potential allowed.

Addressing this need, ReaxFF, introduced by van Duin and Goddard in 2001, provided a significant breakthrough. [32–34] At the core of ReaxFF lies the concept of BO, which dynamically adjusts based on interatomic distances. This mechanism allows the energy contribution of a bond to decrease smoothly as atoms move apart and increase as they approach, enabling continuous transitions between bonded and non-bonded states. The flexible design of ReaxFF grants remarkable adaptability, leading to its widespread adoption for simulating complex reactive phenomena across diverse chemical environments and making it the central methodology explored in this thesis.

The BO between two atoms in ReaxFF is calculated using a sum of three exponential terms, each representing a different type of covalent bond (σ, π, and $\pi\pi$). The (uncorrected) BO[2] between atoms i and j is given by:

$$\begin{aligned} \mathrm{BO}'_{\mathrm{ij}} &= \mathrm{BO}'_{\mathrm{ij},\sigma} + \mathrm{BO}'_{\mathrm{ij},\pi} + \mathrm{BO}'_{\mathrm{ij},\pi\pi} \\ &= \exp\left(p_{\mathrm{bo},1} \left(\frac{r_{\mathrm{ij}}}{r_0^{\sigma}} \right)^{p_{\mathrm{bo},2}} \right) + \exp\left(p_{\mathrm{bo},3} \left(\frac{r_{\mathrm{ij}}}{r_0^{\pi}} \right)^{p_{\mathrm{bo},4}} \right) \\ &+ \exp\left(p_{\mathrm{bo},5} \left(\frac{r_{\mathrm{ij}}}{r_0^{\pi\pi}} \right)^{p_{\mathrm{bo},6}} \right) \end{aligned} \tag{2.12}$$

Each term decays exponentially with the interatomic distance r_{ij} with different equilibrium bond lengths r_0^{σ}, r_0^{π} and $r_0^{\pi\pi}$, and parameters p_{bo} that depend on the element types involved. These parameters are fitted during ReaxFF training to match quantum chemical or experimental reference data. This formulation allows ReaxFF to describe single, double, and triple bonds within the same consistent framework, and is illustrated in Figure 2.4.

Due to its flexibility, ReaxFF includes a large number of tunable parameters, far greater than the few described above. The size of a parameter set scales with the number of elements involved, as shown in Table 2.1

---

[2] ReaxFF requires additional corrections to the BO to avoid over-coordination *i.e.* assigning bonds where there are none.

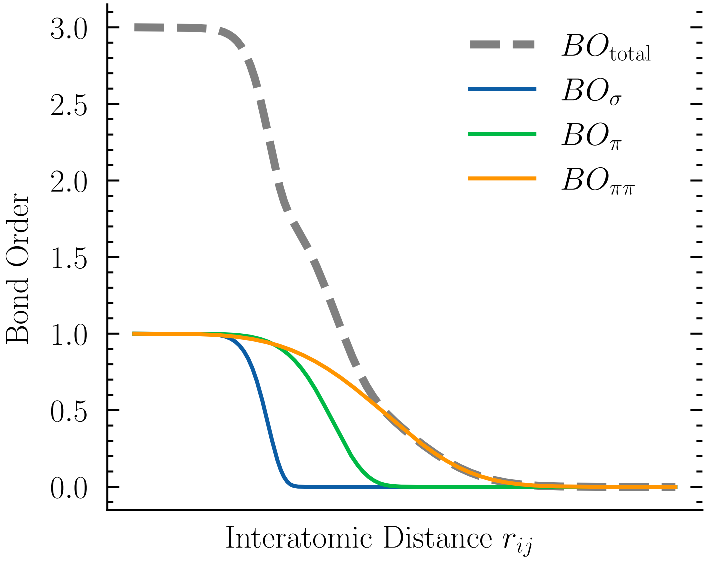

Figure 2.4: The three different ($\mathrm{BO}'_{\mathrm{ij},\sigma}$, $\mathrm{BO}'_{\mathrm{ij},\pi}$, $\mathrm{BO}'_{\mathrm{ij},\pi\pi}$) contributions towards the total BO are shown.

Parameter Category	Instance Definition	Atom Types Involved	Parameter count (per instance)
Global	System wide	0	41
Atomic	Per atom type	1	32
Bond	Per bond type	2	16
Valence angle	Per angle type	3	7
Torsion	Per torsion type	4	7
Off-diagonal	Per atom type pair	2	6
Hydrogen bond	Per H-bond type	3	4

Table 2.1: Overview and Scope of Parameter Categories in a ReaxFF Force Field. (*Parameter counts are approximate and based on typical ReaxFF implementations*)

The total system energy in ReaxFF is expressed as:

$$\begin{aligned} E_{\text{System}} = E_{\text{Bond}} + E_{\text{Over}} + E_{\text{Under}} + E_{\text{Val}} + E_{\text{Tors}} \\ + E_{\text{vdW}} + E_{\text{Charge}} + E_{\text{Specific}} \end{aligned} \tag{2.13}$$

where E_{Bond} corresponds to bond formation and breaking terms, E_{Over} and E_{Under} are correction terms for over- and under-coordination respectively. E_{Val} is the valence angle energy and E_{Tors} is the torsional angle energy between four different particles. E_{Charge} and E_{vdW} are non-covalent terms, the former being the charge interaction and the latter the van der Waals interaction. The charge interactions take into account polarization and Coulomb forces. All of the above-mentioned terms are required in each ReaxFF parametrization. The last term of the equation, E_{Specific}, is a collection term of optional additional contributions dependent on the model and system. An overview of the total system energy computation and the dependence on the BO is shown in Figure 2.5.

2.3.1 Optimization of a ReaxFF Force Field

Effectively using ReaxFF to model a system requires careful and system-specific parameter tuning. A typical ReaxFF parameter set contains around 100 parameters per element type, many of which need to be optimized, either

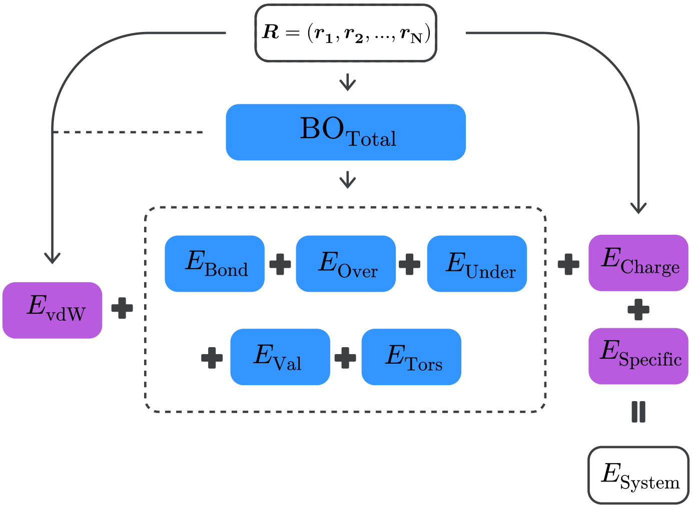

Figure 2.5: Schematic overview of energy calculation in the ReaxFF reactive force field. Atomic coordinates (R) determine bond orders (BO_{Total}), which mediate bonded energy terms (E_{Bond}, $E_{\text{Over}}/E_{\text{Under}}$, E_{Val}, E_{Tors}). These are combined with non-bonded, indirectly BO-dependent van der Waals (E_{vdW}) and electrostatic (E_{Charge}) interactions, plus any system-specific terms (E_{Specific}), to yield the total system energy (E_{System}).

manually or through an optimization algorithm. [35] Furthermore, beyond the large number of parameters, parameters governing different aspects are often heavily correlated and intricately influence each other's effects on the potential energy surface. This makes the search for a reliable and transferable parameter set a highly non-trivial task.

Given the vast parameter space, potential correlations between parameters, the lack of property representation in the training data and the high computational cost of evaluating the force field's predictions, optimizing all parameters simultaneously is often infeasible or can lead to ill-defined results. Therefore, a common practical approach is to optimize only a carefully selected subset of the parameters that are most sensitive to the properties of interest or that are deemed crucial for improving the description of specific chemical phenomena, like reaction barriers or bond dissociation. The remaining parameters might be kept fixed, often using values from previously published parameter sets or earlier stages of optimization. To guide the optimization of this selected parameter subset, a quantitative measure of agreement between the force field's predictions and high-quality reference data is needed. This is achieved by defining and minimizing a cost function or loss function $\mathcal{L}(\boldsymbol{p})$:

$$\mathcal{L}(\boldsymbol{p}) = \sum_{i=1}^{n} \left(\frac{x_{i,\text{reaxff}}(\boldsymbol{p}) - x_{i,\text{ref}}}{\sigma_i} \right)^2 \tag{2.14}$$

Here, x represents a molecular property such as the energy, bond lengths, forces, ... The term σ_i is a scale factor that accounts for the acceptable deviation for each property. p is then the suggested ReaxFF parameter set. $x_{i,\text{reaxff}}(\boldsymbol{p})$ and $x_{i,\text{ref}}$ are the value predicted by ReaxFF using parameter set $\boldsymbol{p}$ and the the reference value, respectively. The sum then runs over all entries n in the training set. The goal is then to find a subset $\boldsymbol{p}$ that minimizes $\mathcal{L}$ given a certain training set:

$$\boldsymbol{p}_{\text{opt}} = \arg\min_{\boldsymbol{p}} \mathcal{L}(\boldsymbol{p}) \tag{2.15}$$

Although significant improvements have been made to the smoothness of ReaxFF's functional form, the potential energy surface defined by its parameters remains numerically unstable and often exhibits discontinuities. [36–38] Furthermore, the loss function landscape often becomes rugged because calculating target properties frequently requires intermediate geometry optimizations; even small parameter changes can lead to different optimized minima and thus non-smooth changes in the loss. These characteristics severely limit the applicability of gradient-based optimization methods, which rely on smooth and continuous loss landscapes. As a result gradient-free approaches are strongly preferred for ReaxFF parameter fitting.

Successive One-Parameter Parabolic Extrapolation (SOPPE) is a simple and intuitive optimization strategy, originally proposed for ReaxFF parameter tuning.[33] In SOPPE, all parameters are fixed except one, which is sampled at multiple values—typically three—to fit a local quadratic model of the cost function. The method is easy to implement and interpret, making it appealing for non-expert users or when parameter interactions are assumed to be minimal. However, ReaxFF's parameter space is highly coupled and nonlinear, meaning that changes in one parameter often disturb the optimality of others. SOPPE does not account for such interactions, requiring repeated passes over all parameters, which becomes inefficient in high dimensions. Moreover, its reliance on local quadratic approximations breaks down if samples are too far apart or if the error surface is noisy. Early ReaxFF implementations relied on SOPPE with iterative cycling through parameters until convergence. While transparent and user-friendly, the method scales poorly and has largely been replaced by more robust global optimization algorithms.

Genetic algorithms (GAs) were the first ReaxFF optimization methods that moved beyond SOPPE approaches.[39,40] These algorithms apply selection, crossover, and mutation to evolve parameter sets. Their main advantage lies in their global search capability and minimal requirement for user intuition. Since their introduction, GAs and related evolutionary strategies have been applied in various ReaxFF studies, including multi-objective and parallel implementations. [41,42] Despite their versatility, results emphasize the importance of well-defined optimization criteria, parameter interdependencies, and cross-system transferability.

Another commonly used global optimization method for ReaxFF optimization is the Monte Carlo Force Field (MCFF) optimizer with simulated annealing. [41,43,44] This statistically driven optimization method is inspired by thermodynamics and uses a temperature-like parameter T to control a gradual “cooling” process. Cooling down the system (or annealing) helps the algorithm escape local minima and eventually converge toward a global optimum. [45,46] MCFF is based on the Metropolis-Hastings algorithm, where new parameter sets are proposed by taking random steps around the current values:

$$\boldsymbol{p}_{i,\text{new}} = \boldsymbol{p}_i + \delta\boldsymbol{p}_{i,\max}(2u-1); u \sim \mathcal{U}(0,1) \tag{2.16}$$

If the new parameter set yields a lower error ε, it is always accepted. Otherwise, it may still be accepted with a probability depending on the temperature parameter T:

$$P_a(\varepsilon_{\text{new}}, \varepsilon_{\text{old}}, T) := \begin{cases} 1 \text{ if } \varepsilon_{\text{new}} \leq \varepsilon_{\text{old}} \\ \exp\left(\frac{\varepsilon_{\text{old}}-\varepsilon_{\text{new}}}{T}\right) \text{ if } \varepsilon_{\text{new}} > \varepsilon_{\text{old}} \end{cases} \tag{2.17}$$

At high temperatures, even suboptimal steps are often accepted, allowing broad exploration of the parameter space. As the temperature decreases, the algorithm becomes more selective, favoring lower-error configurations. While MCFF is conceptually simple and effective for exploring rugged error surfaces, its success depends heavily on the annealing procedure. Cooling too quickly can lead to premature convergence to local minima. In theory, a reversible temperature decrease guarantees convergence to the global minimum, but practical implementations often trade this guarantee for efficiency.

These approaches—SOPPE, GA, and MCFF—represent some of the most widely used methodologies for ReaxFF parameter optimization. However, many other strategies exist. The current state-of-the-art in ReaxFF optimization is the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) algorithm, which is also used in this work and described in more detail in the following section.

2.3.2 Covariance Matrix Adaptation

CMA-ES is a powerful, derivative-free global optimization algorithm. Originally proposed by Hansen *et al.*, CMA-ES is rooted in self-adaptive evolutionary strategies.[47–51] It operates by iteratively adapting a multivariate normal distribution to model the loss function's surface spanned by the ReaxFF parameters in this case.

At each generation, CMA-ES produces a population of candidate solutions (individuals), evaluates their performance, and uses the best-performing individuals to update the mean and covariance matrix of the sampling distribution for the next generation. An illustration of this with two parameters is shown in Figure 2.6. The algorithm is initialized with an initial guess for the parameter set and an initial covariance matrix. One iteration ($g \to g+1$) of the basic CMA-ES procedure involves the following steps [41,48,49]:

1. **Sampling:** Generate λ candidate parameter vectors (individuals) from a multivariate normal distribution:

   $$\boldsymbol{p}_i^{g+1} \sim \mathcal{N}\left(m^{(g)}, C^{(g)}\right) \tag{2.18}$$

   Here, $m^{(g)}$ and $C^{((g)}$ represent the current mean and covariance matrix at generation g. The population size λ is typically chosen proportional to $\ln\left(N_{\text{par}}\right)$, where N_{par} is the number of tunable parameters.
2. **Evaluation:** Evaluate the loss function $\mathcal{L}\left(\boldsymbol{p}_i^{g+1}\right)$ for each individual $k = 1, ..., \lambda$. In ReaxFf optimization, this means running the ReaxFF engine with parameter set $\boldsymbol{p}_k^{(g+1)}$ against the training set and comparing to reference values.
3. **Selection and Weighting:** Individuals are ranked by performance (lower $\mathcal{L}$ is better), and the top μ (where $\mu < \lambda$) individuals contribute to the next generation's updates. Assign weights w_i to these individuals, typically with higher weights to better performing individuals. The weights are often normalized such that $\sum_{i=1}^{\mu} w_i = 1$.
4. **Update of Mean and Covariance Matrix:** Update the parameters of the sampling distribution for the next generation ($g+1$) based on the weighted contributions of the selected μ individuals from the current generation. The new mean $m^{(g+1)}$ is calculated as the weighted average of the selected individuals, effectively shifting the center of the search towards promising regions. The new covariance matrix $C^{(g+1)}$ update adapts the shape and orientation of the sampling distribution. This sophisticated update mechanism is crucial to CMA-ES's performance. Conceptually, it learns from:
   - Current Successes: The primary information comes from the distribution of the steps ($p_i^{(g+1)} - m^{(g)}$) taken by the successful individuals in the current generation. This is often called the rank-μ update. It adapts the covariance to reflect the observed variance of successful search directions.
   - History: To ensure robustness and avoid reacting too strongly to noise in a single generation, the update typically blends information from the previous covariance matrix C^g using a learning rate.

- Search Path (Cumulation): CMA-ES maintains "evolution paths" that accumulate the average direction of movement of the mean over several generations. This path information helps to detect and exploit consistent directions of progress, allowing the covariance matrix to elongate along successful search paths. This is related to the rank-one update and accelerates convergence, particularly on narrow valleys or ridges in the loss landscape. The final update combines these different sources of information.

5. **Convergence Check:** Evaluate termination criteria to determine convergence. A common criterion checks if the cost function for all parameters along the principal axes of the covariance matrix drops below a threshold. This is critical in ReaxFF optimization, where a single misconfigured parameter may lead to non-physical behavior. Another criterion checks whether the improvement remains below a set loss threshold for multiple step sizes.

Multiple CMA-ES variants exist, employing different strategies for updating the covariance matrix adaptation. Some use approximate covariance updates to avoid the computational cost of matrix inversion, while others introduce elitism by allowing high-performing individuals to persist across generations. These variants improve scalability and convergence speed, especially in high-dimensional parameter spaces.

CMA-ES is particularly attractive because of its robust performance on difficult optimization problems and its relative ease of use, only requiring an initial guess and covariance matrix. Furthermore, its structure lends itself well to parallelization as the evaluation of λ individuals can often be done independently. Efficient covariance approximations can reduce the computational complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(2N)$, making the algorithm highly feasible for optimization problems with many parameters.[50]

A potential drawback off CMA-ES, and to many stochastic optimizers, is sensitivity to the initial guess. Different initializations might lead the search to converge in different local minima of the loss function, which often necessitates multiple independent runs to ensure robust results.

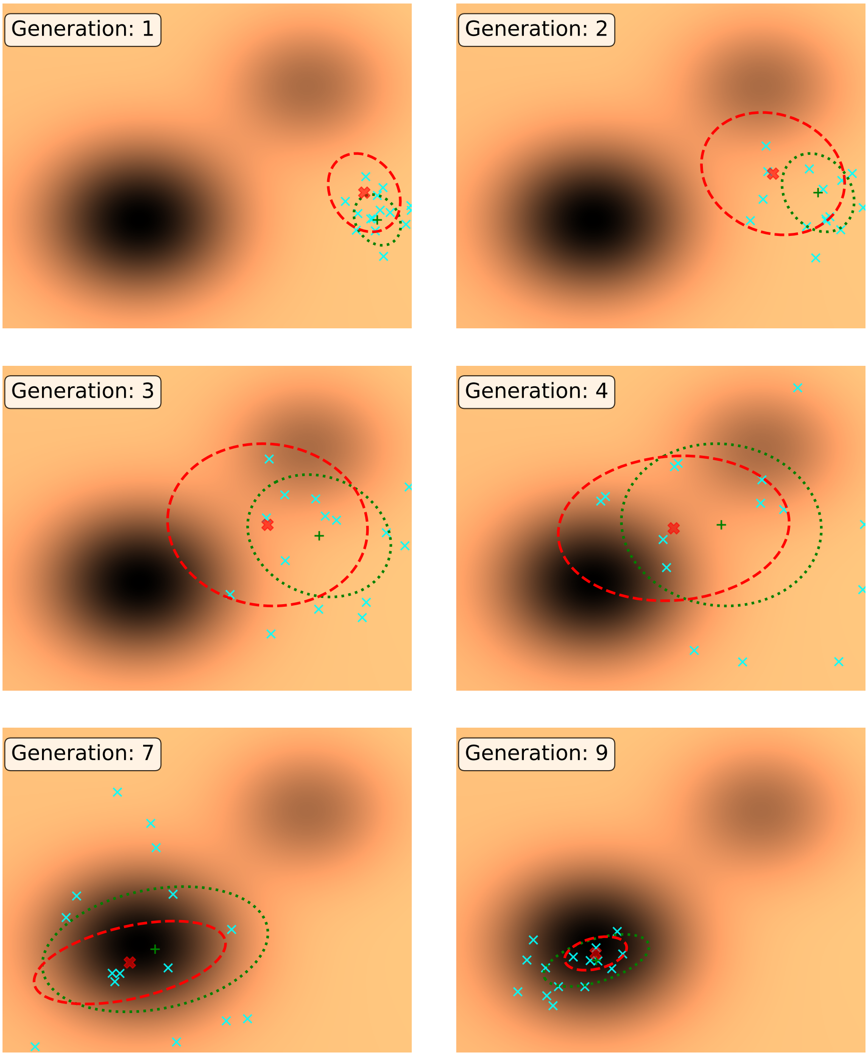

Figure 2.6: An example of the CMA-ES algorithm optimizing a two-dimensional potential energy function. Each frame displays: (1) The sampling distribution for the current generation in green dotted ellipse centered on the green '+' (2) The λ candidate solutions drawn from this distribution in cyan 'x' markers (3) The resulting updated distribution, based on weighted recombination of the evaluated samples, marked in red dashed ellipse centered on the new mean 'X'

2.4 Machine Learning Interatomic Potentials

MLIPs have emerged as a transformative approach over the past ~15 years, developed to achieve high fidelity in simulating atomic interactions without the prohibitive computational cost of traditional QM methods. The core concept of MLIPs is to employ powerful function approximation techniques from machine learning, such as Gaussian Process Regression or neural networks, to learn the complex relationship between the 3D arrangement of atoms (that is, using the atomic coordinates r_i and atomic numbers z_i) and the system's potential energy surface directly from a dataset of reference QM calculations. By training on QM energies and forces, MLIPs can capture intricate many-body interactions and subtle QM effects implicitly contained within the reference data, without being constrained by the fixed functional forms inherent to classical potentials.

A crucial and most important component in developing MLIPs is the representation of the atomic structure. Early and influential MLIPs often relied on descriptors encoding the local atomic environment around each atom. These descriptors, such as Atom-Centered Symmetry Functions (ACSFs) or the Smooth Overlap of Atomic Positions (SOAP), were carefully designed to be invariant to translation, rotation, and the permutation of identical atoms, capturing the essential geometry required by the underlying physics.[52–56]

More recently, inspired by advances in deep learning, architectures like graph neural networks have been developed that learn relevant atomic representations directly from the data through iterative refinement of atom-centered representations through several interactions layers ($t = 0, 1, ..., T$). This process is called message-passing between neighboring atoms.[57,58]

Each atom i is associated with a feature vector (or state) $\boldsymbol{h}_i^{(t)}$ at each layer t. This vector encodes learned information about the atom and its surrounding chemical environment as determined by the network up to that layer. The initialization process (i.e. layer $t = 0$) starts by constructing an initial feature vector $\boldsymbol{h}_i^{(0)}$ for each atom based solely on its identity. This is typically achieved using a learnable embedding: each distinct atomic number z present in the dataset is assigned a unique vector $\boldsymbol{a}_z$ of a predetermined dimension. The initial state for atom i is then simply set to this embedding:

$$\boldsymbol{h}_i^{(0)} = \boldsymbol{a}_{z_i} \tag{2.19}$$

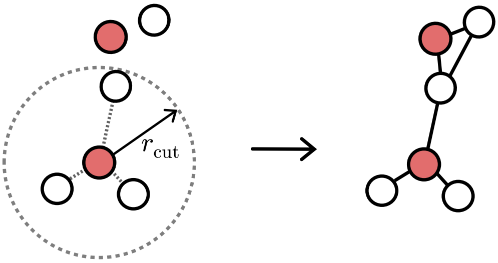

Figure 2.7: Illustration how an atomic system with 2 different atomic types are translated to a graph in message-passing MLIPs. MLIPs have no notion of chemical bonds, and will create a graph based on all atoms that fall within the r_{cut} range.

After the initialization, the feature vectors are iteratively updated in the interaction layers ($t > 0$) by exchanging information between neighboring atoms, the message-passing process. To compute the updated state $\boldsymbol{h}_i^{(t)}$, atom i aggregates information ("messages") derived from the feature vectors of its neighbors $\boldsymbol{h}_j^{(t-1)}$ within distance r_{cut} (illustrated in Figure 2.7) from the previous layer.

Let's define the collection of atoms j surrounding atom i within the cutoff distance as $\mathcal{K}(i)$. For each neighbor $j \in \mathcal{K}(i)$, a "message" $\boldsymbol{m}_{j \to i}$ is computed using a learnable message function M_t. This function usually depends on the feature vectors of both the central atom i and the neighboring atom j from the previous layer $(t-1)$, as well as their relative geometric position vector $\boldsymbol{r}_{\text{ij}} = \boldsymbol{r}_i - \boldsymbol{r}_j$ (or derived quantities like distance d_{ij} and direction $\vec{r}_{\text{ij}}$):

$$\boldsymbol{m}_{j \to i}^{(t)} = M_t\left(\boldsymbol{h}_i^{(t-1)}, \boldsymbol{h}_j^{(t-1)}, \boldsymbol{r}_{\text{ij}}\right) \tag{2.20}$$

The messages arriving at atom i from all its neighbors are then aggregated into a single vector $\boldsymbol{m}_i^{(t)}$, typically using a permutation-invariant operation like summation:

$$\boldsymbol{m}_i^{(t)} = \sum_{j \in \mathcal{K}(i)} \boldsymbol{m}_{j \to i}^{(t)} \tag{2.21}$$

The feature vector (or state) of atom i for the current layer t is then constructed using an update function U_t. This function combines the atom's state from the previous layer $\boldsymbol{h}_i^{(t-1)}$ with the aggregated message $\boldsymbol{m}_i^{(t)}$ computed in the current step. The update function U_t itself is typically a learnable neural network module

$$\boldsymbol{h}_i^{(t)} = U_t\left(\boldsymbol{h}_i^{(t-1)}, \boldsymbol{m}_i^{(t)}\right) \tag{2.22}$$

Iterating this message construction, aggregation, and state update process through T layers allows the network to build increasingly complex representations of the local atomic environment by effectively propagating information through the neighborhood graph defined by the input geometry.

Once the final feature vectors $\boldsymbol{h}_i^T$ are obtained after the last interaction layer, a final set of layers R_t map these learned features to atomic energy contributions E_i, often called "node" energies. While these individual E_i values lack a strict physical interpretation, their sum yields the total potential energy E_{total} for the atomic configuration:

$$E_{\text{total}} = \sum_i E_i = \sum_i R_t\left(\boldsymbol{h}_i^{(T)}\right) \tag{2.23}$$

An overview of such a workflow is illustrated in Figure 2.8.

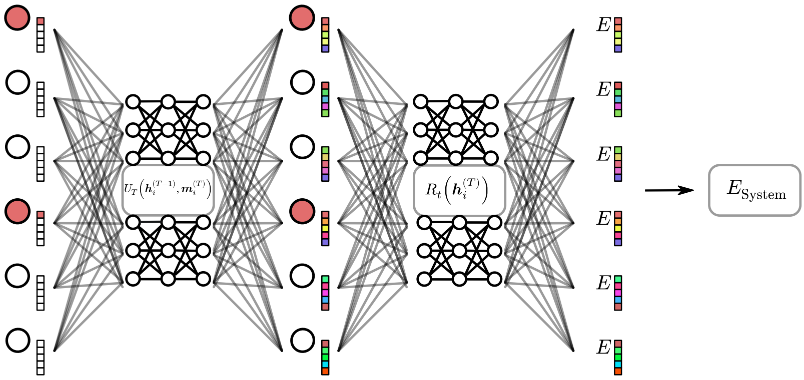

Figure 2.8: Schematic overview of the data-flow in a MLIP. On the very left of the figure, we can see the atoms and their *learnable* embedding vector, only dependent on their atomic number z_i. Through a message passing neural network, the atomic vectors or *features* are updated based on atoms in their local environment, cut off at r_{cut}. Once the atomic features are built, a final set of layers (often called the readout network) uses those atomic features $\boldsymbol{h}_i$ and maps them to individual "node" energies. Summing those "node" energies then results in the final energy E_{tot} of the system.

A crucial aspect in designing MLIPs is ensuring they respect the fundamental symmetries of physics. The potential energy E must be invariant under translations, rotations, and reflections (collectively the $E(3)$ group) of the system, as well as permutations of identical atoms. Atomic forces $\vec{F}_i = -\vec{\nabla}_{r_i} E_i$, however, must be equivariant; they should rotate and reflect along with the system.

Many early MPNNs, like SchNet, ensured rotational invariance by using only scalar features (e.g., interatomic distances) in the message and update functions.[59,60] While successful, this approach inherently limits the model's ability to represent and propagate directional information, which can be crucial for describing complex interactions. For instance, distinguishing conformations based on relative orientations can be difficult using only distances and angles, especially with small cutoffs.

More recently, $E(3)$-equivariant MPNNs, such as NequIP, PaiNN, and MACE, have been developed. [61–65] These networks utilize internal features that are geometric tensors (vectors and higher-order tensors) which transform predictably under rotations and reflections of the atomic coordinates. Formally, a feature $\boldsymbol{h}_L$ is equivariant if, considering rotations $g \in SO(3)$ with matrix representation $D(g)$, it satisfies:

$$\boldsymbol{h}_L(g \cdot x) = D_L(g)\boldsymbol{h}_L(x) \tag{2.24}$$

where L labels the irreducible representation of the rotation group $SO(3)$, indexed by a non-negative integer $L = 0, 1, 2, ...$ Within these models, features are typically constructed using spherical harmonics $Y_{\ell m}$ with a maximum degree of order $L_{\max}$. A feature component $\boldsymbol{h}_L$ (where $L < L_{\max}$) transforms under rotation in the same characteristic way as the spherical harmonics $Y_{\ell m}$ of order L: $L = 0$ features are scalers, transforming like a monopole moment; $L = 1$ features are vectors, transforming like a dipole moment; $L = 2$ are rank-2 tensors, transforming like a quadrupole moment, and so on for higher L up to the maximum $L_{\max}$ used by the model.

The final energy prediction is obtained by contracting equivariant features into an invariant scalar at the very last step in R_t. By utilizing features that transform predictably under rotation in addition to invariant scalars, these models capture a richer, more physically informative picture of the local atomic environment compared to approaches relying solely on invariant inputs like interatomic distances. Furthermore, the network operations themselves are constructed to strictly adhere to these geometric transformation rules. This built-in equivariance acts as a strong physical constraint, guiding the learning process within a relevant subspace and preventing the model from needing to learn fundamental rotational physics from scratch. Together, this enables the model to learn the complex mapping between structure and energy more effectively and generalize better from limited data. Training accurate (equivariant) MLIPs on a few hundred reference structures is not uncommon.

3

Molecular Modeling Applications

"Nothing is lost, nothing is created,
everything is transformed."
–Linus Pauling

The development and refinement of computational modeling techniques, such as those discussed in the previous chapter, ultimately find their purpose in application. To test, validate, and demonstrate the novel methodologies for training data curation and potential development central to this thesis, it is essential to apply them to challenging and relevant chemical systems. The systems chosen serve not only as demanding testbeds for assessing the accuracy, efficiency, and robustness of reactive models but are also selected for their intrinsic scientific interest and industrial importance.

This chapter introduces the two primary chemical systems investigated in this work: Alumina (Al_2O_3) and Hydrated Silicate Ionic Liquids (HSILs). Each presents unique complexities in terms of structure, reactivity, and dynamics, providing diverse and realistic environments in which to evaluate and showcase the core contributions of this dissertation to ReaxFF parameterization and training of MLIPs. The following sections provide the necessary background on the structure, properties, and significance of these materials, setting the stage for the two subsequent chapters where the methodological developments are applied directly to these systems.

3.1 Alumina: Structure, Properties and Catalytic Relevance

Aluminum oxide (alumina, Al_2O_3) is a cornerstone material in industrial catalysis, extensively used in catalytic processes, particularly in the automotive and petrochemical sectors.[66–69] Its utility stems from a combination of tunable textural properties and surface chemistry. However, alumina exhibits significant complexity due to its polymorphism and sensitivity to synthesis conditions and hydration state.

3.1.1 Alumina as a Binder

Alumina, particularly the gamma phase, is widely utilized as a binder to impart mechanical strength and shape to catalyst formulations, enabling their use in industrial reactors by improving handling and managing pressure drop. [67,70,71] This binding function significantly enhances the crush strength and attrition resistance of the catalyst particles, with the final mechanical properties depending heavily on preparation parameters like binder concentration and compaction pressure. [72] However, alumina binders are not inert; they dilute the active catalytic component and modify the composite's overall physicochemical properties, including texture, porosity, and surface acidity (primarily adding Lewis sites). These modifications, along with potential interactions that redistribute the active phase, consequently impact catalytic performance, often reducing overall activity while potentially altering reaction selectivity and deactivation pathways. Optimizing binder use therefore involves balancing the required mechanical integrity against these inevitable effects on catalytic function.

3.1.2 Structure and Polymorphism

Alumina exists in numerous crystalline forms beyond the thermodynamically stable α-Al_2O_3. [73–75] Several metastable "transition aluminas" can be formed, often depending on the precursor and processing conditions. These are generally based on either a face-centered cubic (fcc) or hexagonal close-packed (hcp) oxygen sublattice, with Al^{3+} cations occupying interstitial octahedral and tetrahedral sites to varying degrees.

Among the transition aluminas, the gamma phase (γ-Al_2O_3) is particularly important for catalytic applications. It is commonly produced by calcining aluminum hydroxides or oxyhydroxides, such as boehmite (γ-

$AlOOH$) or bayerite (α-$Al(OH)_3$), typically between 350 and 1000 °C. [76] The transformation sequence often follows boehmite $\rightarrow \gamma \rightarrow \delta \rightarrow \theta \rightarrow \alpha$-$Al_2O_3$, although the exact phases and transition temperatures depend significantly on factors like precursor crystallinity and heating conditions.

The structure of γ-Al_2O_3 is frequently described as a defect cubic spinel. [75] The oxygen anions form an approximately fcc lattice, while the Al^{3+} cations occupy only a fraction of the octahedral and tetrahedral sites to maintain stoichiometry. The resulting cation vacancies per unit cell are often considered to be disordered, primarily over the tetrahedral sites, although debate continues regarding their exact distribution and potential ordering. Some studies also suggest a tetragonal distortion (deviation from perfect cubic symmetry) for γ-Al_2O_3 derived from crystalline precursors like boehmite, potentially related to residual hydroxyls or cation ordering. Other metastable phases like δ, θ, and the more recently identified θ', θ'', and λ phases also arise from different ordering patterns of Al cations within the fcc oxygen framework. [77,78]

3.1.3 Surface Chemistry and Reactivity

The surface chemistry of alumina is dominated by the presence of hydroxyl groups (aluminols) under ambient or aqueous conditions, formed by the dissociative chemisorption of water. These surface OH groups, along with coordinatively unsaturated Al^{3+} and O^{2-} sites exposed upon dehydration, confer both Brønsted and Lewis acid-base properties to the material.[79]

- **Lewis Acidity:** Attributed to exposed Al3+ cations, primarily tetracoordinated (Al^{IV}) or pentacoordinated (Al^{V}) Al sites are considered strong Lewis acids. Lewis acidity increases significantly upon dehydration (high activation temperatures) as hydroxyl groups are removed.
- **Basicity/Brønsted Acidity:** Associated with surface oxygen anions and hydroxyl groups. The acid/base character of OH groups depends on their coordination and the coordination of the Al atoms they are bonded to. Surface basicity is reported to be maximal on fully hydroxylated surfaces (formed at lower activation temperatures, 140 °C) and decreases upon further dehydration. The point of zero charge varies widely (pH 6-9) depending on synthesis, hydration, and impurities, reflecting surface heterogeneity. [67]

Interaction with water is critical. Initial adsorption on a dehydrated surface is strongly exothermic and dissociative, forming hydroxyl groups. Subsequent water layers interact via hydrogen bonding, eventually forming liquid-like films at high relative humidity. Edge sites, containing low-coordinated Al and O atoms, exhibit distinct reactivity, including strong water adsorption and selective reaction with probes. Prolonged exposure to liquid water can lead to phase transformation, with γ-$\mathrm{Al_2O_3}$ slowly converting to more stable hydrated phases like bayerite (β–$\mathrm{Al(OH)_3}$), often via an amorphous intermediate. This transformation alters the surface area and significantly reduces surface reactivity.

3.1.4 Surface Properties and Energetics

Industrially relevant γ-$\mathrm{Al_2O_3}$ possesses desirable textural properties, including high specific surface area (often >150 $\mathrm{m^2\,g^{-1}}$, sometimes exceeding 300 $\mathrm{m^2\,g^{-1}}$) and significant pore volume, which are crucial for dispersing active catalytic species. The high surface area implies a significant contribution from surface energy to the material's overall thermodynamics. MD simulations suggested that γ-$\mathrm{Al_2O_3}$ might have lower surface energies than α-$\mathrm{Al_2O_3}$, potentially explaining its stability at high surface areas. However, calorimetric measurements indicate that while nanocrystalline aluminas are destabilized relative to bulk, the hydrated surface energies of γ-$\mathrm{Al_2O_3}$ and α-$\mathrm{Al_2O_3}$ are actually quite similar. The measured anhydrous surface energy of γ-$\mathrm{Al_2O_3}$ is relatively high, around 1.5 $\mathrm{J\,m^{-2}}$, decreasing significantly upon hydration. It is proposed that highly reactive sites on anhydrous surfaces, especially on α-$\mathrm{Al_2O_3}$, are effectively "cured" or stabilized by water adsorption, masking potential differences in anhydrous surface energies. The heat of water adsorption is strongly correlated with the anhydrous surface energy, with higher energy surfaces exhibiting more exothermic adsorption. The surface of alumina nanoparticles is heterogeneous, exposing different crystallographic facets (primarily (110), (100), and (111) for γ-$\mathrm{Al_2O_3}$) as well as edges and corner sites. These different sites exhibit distinct coordination environments and reactivity.

3.1.5 Challenges in Simulating Water-Alumina Interfaces

Simulating these complex surface phenomena, particularly the dynamics of water interaction, dissociation, surface reconstruction, and potential phase transformations at the interface, poses significant challenges for computational modeling. Accurately capturing the subtle energy differences and bond breaking/formation events demands accurate reactive models, while understanding processes like surface wetting or etching requires simulations spanning considerable time and length scales. Standard classical potentials lack the necessary reactivity description, and direct ab initio MD remains computationally prohibitive for the required scales. This highlights the critical need for accurate and efficient reactive potentials. Developing robust parameterization strategies for potentials like ReaxFF, capable of handling these bond changes, is therefore essential. The 'Balanced Loss' methodology presented in Section 4.3, applied specifically to this water-alumina system, directly tackles the challenge of building such a potential by providing a more systematic and insightful way to incorporate diverse reference data, ultimately aiming to enable physically realistic MD simulations that can unravel the atomistic mechanisms governing alumina surface chemistry.

3.2 Hydrated Silicate Ionic Liquids

Ionic Liquids (ILs) are broadly defined as salts that exist in the liquid state below a certain temperature, commonly set at 100°C. [80] While high-temperature molten salts have long been used in inorganic synthesis, the modern focus is often on ILs containing at least one organic component, which typically exhibit low melting points due to the size and asymmetry of the organic ions. [81] These ILs possess unique properties such as negligible vapor pressure, high polarity, good thermal stability, and often a high degree of structural organization through extended hydrogen-bonding networks. These characteristics have made ILs attractive as 'green' alternatives to volatile organic solvents in fields like catalysis and electrochemistry, but also offer distinct advantages as media for materials synthesis, including nanoparticles and inorganic nanostructures. The term "Ionothermal Synthesis" describes the use of ILs as a solvent, directly paralleling the established hydrothermal (water solvent) and solvothermal (organic solvent) methods.

3.2.1 Ionothermal Zeolite Synthesis Using Organic Ionic Liquids

In the context of zeolite synthesis, organic ILs were initially explored for their potential to act simultaneously as both the solvent and the template, or structure-directing agent. [82,83] Many common organic IL cations are chemically similar to traditional SDAs used in hydrothermal synthesis. [84,85] This dual role eliminates competition between solvent and template molecules for interaction with the growing inorganic framework. Ionothermal syntheses using organic ILs, such as those based on imidazolium cations, have been particularly successful for preparing aluminophosphate based molecular sieves, yielding both known and novel structures. However, applying this approach to aluminosilicate or silicate zeolites proved challenging, primarily due to the low solubility of silica precursors in many common organic ILs. Successful silicate zeolite synthesis often required modifications, such as the addition of mineralizers like fluoride or hydroxide ions, or small amounts of water, to enhance solubility and promote crystallization. In such cases, the role of the organic IL often reverts to primarily being an SDA, and complications like IL decomposition under synthesis conditions can arise. The IL anion, typically chosen to be weakly coordinating, plays a significant role in determining the IL's physical properties (like hygroscopicity) and can indirectly influence the synthesis outcome ('induction effect') but generally does not act as a template itself.

3.2.2 Hydrated Silicate Ionic Liquids: A New Paradigm

A significant conceptual shift occurred with the development of Hydrated Silicate Ionic Liquids , which are purely inorganic ILs based on alkali silicates. [86,87] Discovered by Taulelle and co-workers, HSILs are defined as room-temperature, hypo-hydrated melts of alkali silicates ($MOH\text{-}SiO_2\text{-}H_2O$, where M = Na, K, Cs). The term 'hypo-hydrated' signifies a low water content where water molecules primarily reside in the coordination shells of the alkali cations rather than acting as a bulk solvent. Unlike organic ILs in ionothermal synthesis, the silicate anions within HSILs are not inert bystanders but constitute the primary source of framework-building units for zeolite formation. HSILs are characterized by strong ion-pairing interactions between the alkali cations and small, deprotonated silicate oligomers (typically monomers to tetramers). The liquid structure, speciation, and resulting properties (conductivity, viscosity) are highly sensitive to the cation type, water content, and alkalinity (Si/OH ratio). These liquids exist in a delicate balance; sufficient alkalinity is needed to deprotonate silanols and enable interaction with cations, while excessive water disrupts ion-pairing and can lead to the formation of colloidal suspensions or gels.

3.2.3 Synthesis of HSILs and HSIL-based Zeolites

HSILs are typically synthesized in the lab via the hydrolysis of a silicon alkoxide, most commonly tetraethoxysilane, in a concentrated aqueous solution of an alkali hydroxide (NaOH, KOH, or CsOH). [88] Under specific stoichiometric conditions (typically MOH/H_2O ratios leading to < 8-10 water molecules per cation after hydrolysis, and SiO_2/MOH ratios around 0.5-1.25), the hydrolysis product, ethanol, acts as a coacervating agent, inducing spontaneous liquid-liquid phase separation. [89] This yields a dense, lower phase consisting of the hypo-hydrated alkali silicate ionic liquid, essentially free of ethanol, and an upper phase containing the ethanol byproduct and excess water. The HSIL phase can be easily separated by decantation. To prepare a zeolite synthesis medium, this 'native' HSIL is then mixed with additional alkali hydroxide, water, and an aluminum source. The aluminum source can be aluminum hydroxide (like gibbsite), sodium aluminate solution, or even aluminum alkoxides (which may also induce phase separation upon hydrolysis). By carefully controlling the final molar ratios (SiO_2, Al_2O_3, MOH, H_2O), homogeneous, transparent, and often aggregate-free precursor liquids can be obtained within specific compositional windows. Hydrothermal treatment of these Al-doped HSILs, typically by heating in sealed autoclaves at moderate temperatures (e.g., 60-170°C) for periods ranging from hours to days, leads to the crystallization of zeolites. [90,91,91–93]

3.2.4 Benefits of Using HSILs in Zeolite Science

The development of HSIL-based synthesis represents a significant advance for investigating zeolite crystallization and offers several key benefits over traditional methods:

- **Homogeneity and Reduced Complexity:** HSIL precursors are typically true, homogeneous liquids, free from the gels or colloidal aggregates often present in conventional synthesis mixtures. This eliminates complications arising from phase boundaries, concentration gradients, and solid-phase aging effects, allowing for a clearer correlation between the bulk liquid composition and the crystallization outcome.
- **Enhanced Characterization:** The optical transparency, relatively low viscosity, and monophasic nature of HSILs make them exceptionally well-suited for investigation by a wide range of in situ analytical techniques. Techniques like liquid-state NMR, conductivity measurements (especially

using MEEIS for corrosive media), pH measurements, dynamic light scattering (DLS), and synchrotron X-ray scattering (SAXS/WAXS) can be applied to monitor the evolution of (alumino)silicate speciation, ion-pairing, potential aggregate formation, and crystal nucleation and growth in real-time. [94]
- **Ideal Model Systems:** The simplified and well-defined nature of HSILs provides realistic yet manageable systems for developing and validating theoretical and computational models (e.g., MD simulations, DFT calculations) of zeolite precursor liquids. This opens pathways to understanding aspects like ion association, oligomer stability, and the molecular steps involved in nucleation at an atomistic level.
- **Synthesis Control & Insights:** HSILs allow for systematic and controlled variation of key synthesis parameters (cation type, hydration level, alkalinity, Si/Al ratio). This facilitates the mapping of phase diagrams and the study of how these parameters influence zeolite phase selection, framework composition (Si/Al ratio), crystal size, morphology, and defect concentration. Through this approach, reproducible syntheses of highly crystalline, ordered materials has been achieved, providing fundamental insights into the structure-directing role of inorganic cations and the thermodynamic factors governing phase selection in inorganic media.

3.2.5 Challenges in Simulating HSILs

From a simulation perspective, the rich chemical complexity of HSILs—involving strong ionic interactions, dynamic silicate speciation, explicit water coordination, and potential aggregation leading to nucleation—presents not only an experimental and computational challenge but also an ideal test case for reactive models. Accurately capturing these interacting phenomena with high fidelity over large system sizes and long timescales is crucial for understanding their role in processes like zeolite synthesis, yet it pushes the limits of current computational methods. Simulating such complex, reactive inorganic systems has previously only been doable with models like ReaxFF, which, as explored in Section 4.3, require intense, system-specific (re)parameterization efforts. The demands of the HSIL system therefore make it an excellent benchmark to evaluate whether modern MLIPs can provide a viable pathway to simulating such complex reactivity accurately and efficiently. Addressing this, the active learning methodology detailed in Section 4.2.2

proposes a route to rapidly and efficiently develop dedicated MLIPs tailored for these challenging conditions.

4

Paper I: Balanced Loss

"So, the thing is this:
it's just like riding a bicycle. It's a balance trick."
– Alan Watts

Background and Motivation

As established in Section 2.3, developing a reliable ReaxFF potential is a complex undertaking, primarily due to the large number of interdependent parameters that must be optimized to reproduce reference data accurately. While sophisticated optimization algorithms like CMA-ES have been employed to navigate the challenging, often noisy potential energy landscape defined by these parameters, a significant and often less systematically addressed challenge lies in the construction and utilization of the training data itself within the optimization's loss function.

Conventionally, ReaxFF optimization minimizes a loss function based on the weighted sum of squared errors between ReaxFF predictions and reference QM data. However, ReaxFF training sets are inherently heterogeneous, containing a diverse mix of data types such as bond lengths, angles, energy differences, and potentially atomic forces, each with distinct physical units and magnitudes. Furthermore, these datasets are often highly imbalanced, with geometrical data points potentially outnumbering energy-related data points by orders of magnitude. Properly weighting these disparate data entries in the loss function to achieve a good compromise—ensuring that critical energy barriers, for example, are not neglected due to the sheer number of bond length constraints—is crucial but notoriously difficult. Historically, this weighting process has often relied on empirical adjustments and expert intuition, a tedious procedure that lacks transparency and offers little guidance when the model fails to meet expectations for certain properties. Is poor performance due to incorrect weighting, or is it an intrinsic limitation of the ReaxFF model for that specific property?

The work presented in Paper I directly confronts this critical bottleneck in ReaxFF development by introducing the "Balanced Loss" function and an accompanying optimization workflow. This methodology fundamentally reformulates the problem by shifting the focus from assigning opaque weights to defining physically meaningful "tolerances" for different categories of training data. These tolerances represent the user's expectation for the acceptable root-mean-square error (RMSE) for each different data category (e.g., Al-O bonds, O-H bonds, specific surface adsorption energies). Setting these tolerances is typically less tedious than assigning traditional weights, as tolerances directly relate to physically meaningful accuracy goals for specific data categories and can often be set based on insight into individual properties without extensive cross-category consideration; the algorithm automatically manages the relative focus during optimization. The Log-Sum-Exp (LSE) form of the Balanced Loss function naturally prioritizes minimizing the error for the category that currently exhibits the largest deviation relative to its defined tolerance. This automated balancing mechanism not only simplifies the parametrization workflow but crucially turns the optimization into a process of validating the initial expectations (tolerances) against the inherent capabilities of the ReaxFF model. This provides clear feedback: if errors in a category remain high even when the LSE Balanced Loss inherently prioritizes it, it indicates the initially set tolerances are likely unattainable with the current ReaxFF model capabilities and the flexibility provided by the parameters set for optimization. The optimization is already focusing maximum effort on reducing the error for that category relative to its tolerance. Within this workflow, the appropriate response is to relax the tolerance, acknowledging the model's limitations and adjusting expectations for that specific category's achievable accuracy. This paper demonstrates the effectiveness of this approach through a nontrivial reparametrization of ReaxFF for water adsorption on alumina.

Contributions of Authors

L.D.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization.

T.V.: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Data Curation, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition, Corresponding Author.

C.C.: Conceptualization, Writing - Review & Editing,

B.C.: Conceptualization, validation, Writing - Review & Editing, Project administration, Supervision, Funding acquisition.

T.B.: Conceptualization, validation, Writing - Review & Editing, Project administration, Supervision, Funding acquisition.

Paper I

Managing Expectations and Imbalanced Training Data in Reactive Force Field Development: An Application to Water Adsorption on Alumina

Loïc Dumortier, Céline Chizallet, Benoit Creton, Theodorus de Bruin and Toon Verstraelen

Journal of Chemical Theory and Computation, **2024**, *20*, 3779-3797

Abstract

ReaxFF is a computationally efficient model for reactive Molecular Dynamics simulations, which has been applied to a wide variety of chemical systems. When ReaxFF parameters are not yet available for a chemistry of interest, they must be (re)optimized, for which one defines a set of training data that the new ReaxFF parameters should reproduce. ReaxFF training sets typically contain diverse properties with different units, some of which are more abundant (by orders of magnitude) than others. To find the best parameters, one conventionally minimizes a weighted sum of squared errors over all data in the training set. One of the challenges in such numerical optimizations is to assign weights so that the optimized parameters represent a good compromise between all the requirements defined in the training set. This work introduces a new loss function, called Balanced Loss, and a workflow that replaces weight assignment with a more manageable procedure. The training data is divided into categories with corresponding "tolerances", i.e. acceptable root-mean-square errors for the categories, which define the expectations for the optimized ReaxFF parameters. Through the Log-Sum-Exp form of Balanced Loss, the parameter optimization is also a validation of one's expectations, providing meaningful feedback that can be used to reconfigure the tolerances if needed. The new methodology is demonstrated with a non-trivial parameterization of ReaxFF for water adsorption on alumina. This results in a new force field that reproduces both rare and frequent properties of a validation set not used for training. We also demonstrate the robustness of the new force field with a Molecular Dynamics simulation of water desorption from a γ-Al_2O_3 slab model.

4.1 Introduction

Reactive force fields are widely used in Molecular Dynamics (MD) simulations because they combine low computational cost, close to that of Molecular Mechanics (MM) models, with the ability to describe chemical events, similar to more expensive Quantum Mechanics (QM) methods, such as Density Functional Theory (DFT). [95] Unlike hybrid QM/MM schemes reactive force fields handle many simultaneous chemical reactions throughout the simulation cell, not just in one predefined active site. This is advantageous for the direct simulation of reaction networks[96] of complex chemical processes such as combustion, [97] pyrolysis, [98,99] chemisorption, [100,101] catalysis, [101] mechanochemistry, [102] crack propagation, [103] nucleation, [104,105] and so on. ReaxFF is one of the most established reactive force fields and is efficient enough to perform multi-nanosecond MD of systems with thousands of atoms using only a single high-performance compute node. [33,106,107] Compared to other popular reactive force fields such as Tersoff, [108] AIREBO-M, [109] or COMB3, [110] ReaxFF has been parameterized for more diverse chemical spaces. [107] More recently, machine learning potentials have also been proposed for reactive MD simulations.[111–113] All of these models share the ambition to simulate complex chemical systems at a computational cost that scales like Molecular Mechanics models.

The computational efficiency of reactive force fields comes at a price. They are generally interatomic potentials, sometimes inspired by physical principles, whose parameters must be fitted to reproduce a chemistry of interest. Such parameterization is fraught with challenges: the collection of reference data sets for training and validation, the choice of numerical optimization algorithm, the selection of parameters to optimize, the computational burden of the parameter optimization, and so on. Specifically for ReaxFF, many optimization algorithms have been proposed and tested, [41,114–121] while the design of reference data sets has received much less attention. For example, ReaxFF parameters are rarely published with their training sets in a reusable form, save for a few exceptions. [41,102,118,120,122,123] However, these data are vital, as they specify the requirements for the optimized parameters and no models will ever outperform the data it was trained on.

A conventional ReaxFF training set consists of various target properties of relevant molecular or periodic structures, including internal coordinates, energy differences and atomic forces. Reaction energies and barriers are obviously important for reactive force fields, but for a training set of N systems, one has at most $N-1$ independent energy differences and many more internal coordinates and atomic forces. In the context of machine learning potentials, this imbalance is addressed by weighting data categories (typically energies and forces) inversely proportional to their prevalence, [124–126] but this practice is less established in the context of ReaxFF. Moreover, in ReaxFF, such weights are often adjusted empirically. For example, one gives more weight to an important subset of the training data in order to prioritize the performance of the trained model for that subset. [119,127,128] Conversely, the model of interest may also be inherently limited for some subsets of reference data, making it pointless to give high weight to such subsets. These subjective motivations mean that training set design requires expert judgment. To make this task more accessible to a broader audience, this paper introduces a new loss function and an intuitive workflow for reweighting training data, called Balanced Loss. It naturally takes into account data imbalance and inherent strengths and weaknesses of the model being trained. A ReaxFF parameterization is used as a case study in this paper, because we believe ReaxFF can greatly benefit from Balanced Loss, but the methodology is general enough to be applied to other (even non-chemical) parameterizations with similar challenges.[121,129–134] As software tools and algorithms for (re)parameterizating (reactive) force fields improve, [41,117,119–121,135,136] we expect more practitioners to face the challenge of data imbalance, also for machine learning potentials that are trained on increasingly large and diverse data sets. [137–139]

Alumina provides a great test case for ReaxFF parameterization because it is a versatile and widely used material in the chemical industry with a complex chemistry, [68,140–143] also at scales out of reach for DFT. Alumina has many known polymorphs, including γ-Al_2O_3, α-Al_2O_3, δ-Al_2O_3 and θ-Al_2O_3, [66] of which γ-Al_2O_3 is the most relevant for catalytic applications. [67,144] For example, alumina selectively adsorbs unwanted elements such as sulfur and can be used as a catalyst for the dehydration of alcohols to ethers and olefins. [69,145–149] However, the main application of γ-Al_2O_3 is in the automotive and petrochemical industries, where it serves as a support for other heterogeneous catalysts such as metals, metal sulfides or metal oxides. [67,150–152] Despite their massive use in industry, the design of alumina-supported catalysts is an empirical process, partly due to the limited fundamental understanding of the materials involved. For example, the exact structure of γ-Al_2O_3 is still under discussion due to its poor crystallinity. [67,153] Also, the microscopic mechanisms at the water-alumina interface during support preparation, metal phase impregnation, shaping and use as a catalyst remain unclear. [154]

The formation, stability and structure of γ-Al_2O_3 are controlled by hydration and dehydration processes. [79,142,143,155–157] The γ phase is formed upon dehydration of boehmite at temperatures between 700 K and

800 K. Once formed, the γ polymporph remains stable up to 1100 K under dry conditions. [67,76,144] γ-Al_2O_3 transitions to other polymorphs upon further increase in temperature and/or water partial pressure. Having both Lewis acid and basic sites on the surface, alumina can react with water in several ways depending on temperature, water partial pressure (for gas/solid interfaces) and pH (for liquid/solid interfaces). [140–143,156,158] Water can adsorb without dissociation by forming an Al-O bond. One of the O-H bonds of water may then dissociate and react with a surface Al-O pair, resulting in two hydroxyl groups, called aluminols. Dissociative adsorption is reported to be more prevalent at crystal surface defects, leading to an "etching"-like degradation at these positions. [159,160] It is clear that the chemistry at the H_2O/Al_2O_3 interface is highly complex and depends on an interplay of multiple microscopic mechanisms and external conditions.

Molecular simulation of the H_2O/Al_2O_3 interface is a promising but ambitious method to improve our understanding of widely used supported catalysts and to pave the way towards their rational design. DFT has often been used to model the H_2O/Al_2O_3 interface. [79,141,143,155,157,161,162] Ideally, sufficiently large atomistic models are considered to avoid artificial spatial correlations, to introduce defects at low concentrations, to include both support and catalyst, and to mimic realistic water concentrations. [163–165] Because larger models also have a larger configurational space, with many local minima on the potential energy surface, their properties can no longer be simulated with static calculations and one should resort to MD to sample all relevant configurations. [166] Linear-scaling DFT implementations[167,168] have enabled ab initio MD simulations of the H_2O/Al_2O_3 interface,[169–171] but they are still computationally demanding compared to reactive force fields. Compared to linear-scaling DFT, ReaxFF has a much lower computational cost, allowing for large-scale MD simulations of alumina.[172–180] The first alumina and water ReaxFF parameters were proposed by Zhang *et al.*, [172] and these were later refined and extended by Joshi *et al.* for aluminosilicates and water, [174,176] which is particularly relevant for simulations of alumina-supported catalysts. However, as shown in the results, the state-of-the-art ReaxFF parameters by Joshi *et al.* poorly reproduce DFT reference data for water adsorption on alumina. This motivated us to demonstrate the relevance of Balanced Loss with a reparameterization of ReaxFF for H_2O/Al_2O_3 interactions, using DFT data from the literature. [79,162,169,181]

The rest of the paper is structured as follows. Section 4.2 contains the methodological details of the study: a brief overview of ReaxFF, the generation of the training and validation data sets, the parameter selection and the optimization algorithm. The Balanced Loss function and workflow are described and motivated in detail in Section 4.3. Section 4.4 presents the results of the ReaxFF training and validation, and it demonstrates the suitability of the resulting force field for MD simulations. The last section formulates the main conclusions and gives an outlook on future work.

4.2 Methodology

4.2.1 ReaxFF Reactive Force Fields

ReaxFF was developed and introduced in 2001 by van Duin *et al.*, for reactive MD simulations, initially of hydrocarbons[33], and has since been regularly extended to other chemistries. [107] Like all force fields, it is a mathematical model of the interactions between atoms in a molecule or a condensed phase, as a function of the Cartesian coordinates of the atomic nuclei. Unlike most classical force fields, it can describe bond breaking and formation.

The ReaxFF potential energy of an atomistic model is defined as:

$$\begin{aligned} E_{\text{System}} = E_{\text{Bond}} + E_{\text{Over}} + E_{\text{Under}} + E_{\text{Val}} + E_{\text{Tors}} \\ + E_{\text{vdW}} + E_{\text{Charge}} + E_{\text{Specific}} \end{aligned} \tag{4.1}$$

<{eq:full_energy_reaxff>

where E_{bond} describes the energy of an atom pair in all relevant regimes: bonded, in transition states and dissociated. E_{over} and E_{under} are correction terms for over- and under-coordination, respectively. E_{val} is the valence angle energy and E_{tors} is the torsional angle energy between four different particles. Non-covalent interactions are modeled with E_{charge} and E_{vdW}, the charge and the van der Waals interactions, respectively. The atomic charges are variable and account for polarization and Coulomb forces. [182,183] In addition to these commonly used energy terms, ReaxFF contains additional contributions for specific use cases, grouped into E_{specific}, which are not used in this work.

The covalent terms depend on bond orders (BO), which are defined for each pair of atoms and allow ReaxFF to describe bond breaking and formation processes in chemical reactions. The uncorrected bond order of a pair of atoms consists of three terms, each corresponding to one type of covalent bond, σ, π and $\pi\pi$:

$$\begin{aligned} \mathrm{BO}'_{\mathrm{ij}} &= \mathrm{BO}'_{\mathrm{ij},\sigma} + \mathrm{BO}'_{\mathrm{ij},\pi} + \mathrm{BO}'_{\mathrm{ij},\pi\pi} \\ &= \exp\left(p_{\mathrm{bo},1}\left(\frac{r_{\mathrm{ij}}}{r_0^{\sigma}}\right)^{p_{\mathrm{bo},2}}\right) + \exp\left(p_{\mathrm{bo},3}\left(\frac{r_{\mathrm{ij}}}{r_0^{\pi}}\right)^{p_{\mathrm{bo},4}}\right) \\ &+ \exp\left(p_{\mathrm{bo},5}\left(\frac{r_{\mathrm{ij}}}{r_0^{\pi\pi}}\right)^{p_{\mathrm{bo},6}}\right) \end{aligned} \tag{4.2}$$

where $p_{\mathrm{bo},\{\mathrm{x}\}}$ represent tunable parameters that can be different for each pair of chemical elements. r_{ij} is the interatomic distance and r_0^{σ}, r_0^{π} and $r_0^{\pi\pi}$ are the element-specific σ, π and $\pi\pi$ equilibrium bond lengths, respectively. The expression for the uncorrected bond orders in Equation (4.2) features only a small subset of all the adjustable ReaxFF parameters. ReaxFF has additional equations (with more parameters) to convert uncorrected to corrected bond orders, which are then used in expressions for the covalent energy terms. A full description can be found in the AMS documentation. [184]

ReaxFF has been implemented in several software packages. The most established ones are the original "Standalone ReaxFF" distributed by van Duin, the commercial implementation in the Amsterdam Modeling Suite (AMS) [185] and the open-source version in the LAMMPS package. [186,187] In this paper, the ReaxFF implementation from AMS (release 2023.101) is used. The parameter optimization, discussed below, is implemented with ParAMS, [135,188] which is a recently developed tool in AMS for the parameterization of approximate potential energy surfaces, such as ReaxFF or Density Functional Tight-Binding (DFTB) models. [131,189] In addition, the Atomistic Simulation Environment [190] was used for processing DFT calculations in the training and validation sets. [190] Visual Molecular Dynamics (VMD) is used for the 3D visualizations in this work. [191]

4.2.2 Training Set Development

The development of the training set goes through the following steps: (i) the selection of bulk, (hydrated) surface and (hydrated) edge structures, (ii)

Group	Structure	Nr.~of structures with additional water
Bulk Structures	Bulk α-Al_2O_3	–
	Bulk boehmite	–
	Bulk γ-Al_2O_3	–
Surfaces	γ-Al_2O_3 (100)	4
	γ-Al_2O_3 (110)	6
	γ-Al_2O_3 (111)	3
Boehmite Surfaces	Boehmite (101)	1
	Boehmite (010)	0
	Boehmite (100)	2
	Boehmite (001)	1
Edges	γ-Al_2O_3 (100-110)	6
Small molecules	$[Al(OH)_4]$-H +-monomer	–
	Water	–

Table 4.1: Overview of the structures in the training set. For surface and edge structures, the number of structures with additional water adsorbed on the surface is mentioned in the last column.

periodic DFT reference calculations on these structures and (iii) the selection of properties from these calculations as training targets.

i **Structures** A training set for optimizing ReaxFF parameters requires reference structures and associated training targets, such as internal coordinates or energies, that ReaxFF should reproduce. The reference structures were taken from previous publications[79,162,169,181] and can be divided into five groups, summarized in Table 4.1 and described in more detail below. For clarity, the relevant crystal surfaces are shown in Figure 4.1. Note that γ-Al_2O_3 is industrially the most relevant material, yet other forms of (hydrated) aluminum oxide were included, most notably boehmite, to increase the diversity of the training data. A complete list of structures is provided in Table S1 of the Supporting Information.

- The group of **Bulk Structures** contains 3D-periodic models of boehmite, γ-Al_2O_3 and α-Al_2O_3 . [155]

- The group of **γ-Al_2O_3 Surfaces** is based on three different slab models, cut along the (100), (110) or (111) crystal planes. In addition to the bare surfaces, structures are included with an increasing number of water molecules adsorbed on the surface. [79] All the structures were published before a distinction was made between the lateral $(110)_\ell$ and basal $(110)_b$ surfaces of γ-Al_2O_3, as shown in Figure 4.1. [143,162] The (110)γ-Al_2O_3 surfaces in the training set are in fact all lateral $(110)_\ell$ surfaces.
- The group of **Boehmite Surfaces** contains slabs with four surface orientations: (101), (010), (100) or (001). Boehmite already contains water in its bulk structure, which is preserved upon cleaving the slabs. In addition to the bare slabs, some have additional water molecules adsorbed. [169,181]
- The group of **γ-Al_2O_3 Edges** comprises structures that represent the edge between surface orientations (100) and (110). In addition to the bare edge structure, six structures with an increasing number of adsorbed water molecules are included. [161]
- The group of **Small Molecules** contains two structures: a γ-alumina monomer, $[Al(OH)_4]^-H^+$, and water. [192]

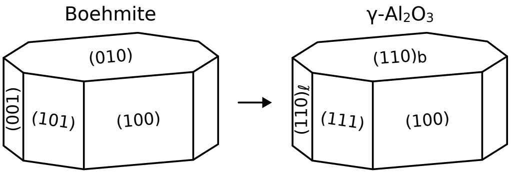

Figure 4.1: Overview of the crystal surfaces of boehmite and , with the relation between the two as proposed in Ref. [162].

ii **Periodic DFT calculations** The geometries of all structures in Table 4.1 are optimized with DFT using periodic boundary conditions and the Perdew, Burke and Ernzerhof (PBE) exchange-correlation functional as implemented in the Vienna Ab initio Simulation Package (VASP) version 5.4. [193,194] The valence interactions are described with the Projected Augmented Wave (PAW) method. [195] An additional *a posteriori* density-dependent dispersion correction dDsC is applied. [196] The plane wave basis set cutoff is 600 eV, which is increased compared to original works from which the structures were taken, [79,162,169,181] to improve the precision of the forces and to facilitate the geometry optimizations. The k-point spacing is set to 0.5 $Å^{-1}$ and a Gaussian smearing is used with a width of 0.05 eV. The convergence criterion for the self-consistent field calculation is set to 1 x 10^{-5} eV. The geometry optimization is performed with the conjugate gradient algorithm and uses a convergence criterion of 0.02 eV/Å on the forces. The cell parameters were not allowed to change and were taken from previous works. [79,162,169,181]

iii **Property Extraction** The selection of internal coordinates, used as training targets, consists of two phases: an analysis of interatomic distances to determine appropriate cutoffs for relevant atom pairs, followed by a classification and enumeration of all relevant distances and angles.

In phase 1, histograms of all interatomic distances up to 5.0 Å were constructed per pair of chemical elements, as shown in Figure S1 in the supporting information. From these histograms, cutoff distances were derived to classify and enumerate all relevant atom pairs. All pairs with a distance below 1.2 Å are classified as covalent O-H bonds. Remaining distances below 2.1 Å are identified as 0···H hydrogen bonds. The AlO distances below 2.8 Å are treated as bonds. AlAl distances below 4.0 Å are not directly bonded, but are included because they are relevant for the local structure of alumina. No other distances were included in the training set.

In phase 2, the distances defined in phase 1 are used to construct the final set of internal coordinates. In addition to distances, valence angles are constructed by combining all pairs of bonded atom pairs sharing one central atom. Dihedral angles are not included, because most O-Al-O-X quartets, where X can be H or Al, contain nearly co-linear bonds, making the dihedral angle ill-defined. Furthermore, H-O-H angles are not included because, for reasons of backward compatibility explained in the following section, the corresponding valence angle term parameters are kept fixed at the values of Joshi *et al.*[174,176] To avoid obvious redundancies in the training set, internal coordinates involving no hydrogen atoms were discarded for structures with additional water molecules, as these internal coordinates already appear in the bare surface and edge structures. The final set comprises four categories of distances (Al-O, Al-Al, O-H and O···H) and six categories of valence angles (O-Al-O, Al-O-Al,Al-O-H, Al-O···H, H-O···H, and H···O ···H).

Note that the last two categories are still relevant to include, unlike angles, because hydrogen bonding angles are sensitive to the structure of the boehmite and alumina surfaces. Histograms of the final selection of internal coordinates can be found in Figures S2 (combined) and S3 (per material) of the Supporting Information. The internal coordinates of boehmite are somewhat similar to those of γ-$\mathrm{Al_2O_3}$, with larger deviations related to hydrogen bonds. Hence, the inclusion of boehmite in the training set increases the diversity of hydrogen bonding information. The prevalence of each class of internal coordinates per structure is given in Table S1.

Energies in a ReaxFF training set are conventionally formulated as internal energy differences between reactant and product states. A full list of reaction energies in the training set is provided in Table S2 of the Supporting Information. These energies are grouped into five classes, denoted with three-letter codes: **BSH**, **GEH**, **GSH**, **SUR** and **FOR**. A short summary of the included reactions and their classification is given below.

For a given alumina surface or edge structure X, all unique pair of adsorption states are used to construct adsorption energy training data, to avoid bias towards a particular reference state. Let m_i and $m_j > m_i$ be the number of water molecules adsorbed in states X_i and X_j, respectively, then the corresponding adsorption energy in the training set is defined as:

$$\Delta_r E_{\mathrm{ads},\mathrm{X}_j,\mathrm{X}_i} = \frac{E_{\mathrm{X}_j} - E_{\mathrm{X}_j}}{m_j - m_i} - E_{\mathrm{H_2O}} \tag{4.3}$$

Thus, if there are N states with water adsorbed for a given structure, there are $(N-1)N/2$ corresponding adsorption energies in the training set. By normalizing the adsorption energies on the number of molecules added, they all have approximately the same order of magnitude. The adsorption energies were grouped into three classes: adsorption on boehmite surfaces (**BSH**), on γ-$\mathrm{Al_2O_3}$, edges (**GEH**) and on γ-$\mathrm{Al_2O_3}$, surfaces (**GSH**).

While the water adsorption energies are of primary interest, also other energies were included to diversify the training set: the transformation from bulk to a slab models (class **SUR**) and the formation of bulk and slab models from the alumina monomer (class **FOR**) are included. Such "reactions" do not correspond to specific reactive events, but they do provide useful information for the covalent ReaxFF parameters.

The resulting training set contains 12931 distances, 13409 angles, 63 water adsorption energies and 19 other energies, for a total of 26422 training targets. Notably, the geometrical features, such as angles and distances, far outnumber the energy entries in the training set. The categories of internal coordinates and energies will be used in the remainder of the paper for a detailed statistical analysis and in the construction of the Balanced Loss function.

It is worth noting that our training set is considerably larger than those of previous ReaxFF parameterizations, which typically contain hundred to a several thousand training targets. [102,116,120,122,123,197–199] To the best of our knowledge, the only exception is the ReaxFF model by Trnka *et al.* for enzyme chemistry, which was trained with 385826 entries. [118] Note that their training set consists of single-point energies and forces, making it computationally less expensive, whereas ReaxFF is normally trained by optimizing geometries at each iteration of the parameter optimization. [121]

4.2.3 Parameter Selection

When calibrating ReaxFF parameters, a careful selection of adjustable parameters must be made. Some parameters can be taken from previous works without refinement, others are not intended to be adjusted, such as the atomic mass, and some parameters are not meaningful for the application of interest. In addition, simply optimizing all parameters would result in an intractable high-dimensional optimization problem. The total number of ReaxFF parameters depends on the number of chemical elements in the system. The parameters are typically grouped into blocks, most of which can be repeated several times for different combinations of chemical elements. Parameter blocks can be independent of chemical elements (41 general parameters), defined per chemical element (32 atomic parameters), per pair (16 bond and 6 off-diagonal parameters), per triplet (7 angle and 4 hydrogen-bond parameters), or per quartet of elements (7 torsion parameters).

In this work, the selection of parameters follows a top-down approach. Initially, all parameters from a literature force field are considered,[174,176] after which several selection criteria are introduced to fix parameters to their literature values, leaving only the remainder to be refined. The selection process aims for a trade-off between an acceptable dimensionality of the optimization problem and a sufficient model flexibility to obtain a good fit.

In the following paragraphs, we motivate our selection criteria, which may be helpful for future ReaxFF calibrations.

Our starting point is the aluminosilicate force field by Joshi *et al.*[174,176] These ReaxFF parameters were calibrated to improve the description of water adsorption at acid sites, Si-O(H)-Al, in the H-ZSM-5 zeolite.[200] The literature `ffield` file contains parameters for 13 chemical elements and one dummy element, resulting in a set of 2961 parameters, many of which are irrelevant to aluminosilicates. For this work, we only consider parameter blocks that contain at least one element, and otherwise only allow O or H. All other parameters are kept fixed, including those related to Si or those describing water.

Ideally, our reparameterization would maintain backward compatibility with the Joshi *et al.* model, changing only parameters specific to alumina and irrelevant to acid sites in zeolites. However, this severely restricts the adjustable parameters to those of atom pairs and valence angles involving at least two Al atoms, namely the Al-Al pairs, and the Al-X-Al and Al-Al-X angles, where X can be Al, O or H. Of this selection, only Al-O-Al angles and Al-Al pairs (≥ 2.5 Å) appear in the training set. As a result of this mismatch, no satisfactory reproduction of our training data was possible when imposing backward compatibility. To avoid this mismatch, we made a pragmatic selection that includes some parameters related to acid sites in zeolites, but also excludes parameters that are only remotely related to our training data. This selection includes Al atom parameters, Al-X bond or pair parameters and Al-H-O, Al-O-H, Al-O-Al, H-Al-O and O-Al-O valence angle parameters.

We further narrow down the parameter selection using the recommendations from the ParAMS documentation.[135,188] In the ParAMS, each parameter is classified with one of the following three labels: “Standard”, “Expert” or “DoNotOptimize”. The first label indicates that the parameter is generally safe to optimize. The second label is used for parameters that should not be changed without a strong motivation. Parameters with the “DoNotOptimize’ label should never be touched, *e.g.*, because they contain boolean values or atomic data.[188] In this paper, only bond and angle parameters with the “Standard” label are considered for optimization.

Finally, some parameters are (de)activated for very specific reasons:

- The atomic parameter `r_0^sigma`, which is r_0^σ in Equation (4.2) for Al is deactivated because it can be overruled by corresponding pair parameters for and bonds.
- Because no bonds are present, of the parameters, only `D_e^sigma` is optimized. This introduces some freedom to tune the weak bonding interactions between pairs of atoms that are not directly bonded.
- π and $\pi\pi$ bond parameters are deactivated because no such bonds are present in our training data. This includes parameters with labels containing any of the following strings: `pi`, `p_bo3`, `p_bo4`, `p_bo5` or `p_bo6`.
- The following "Expert" parameters were activated to improve the angular energy terms: `p_val3`, `p_val4` and `p_val5`.

These selection criteria result in a subset of 36 activate parameters. For each parameter, lower and upper bounds of suitable values are determined and used to restrict the search space during the parameter optimization. For each parameter, the bounds are set equal to the corresponding range of historical values in the ReaxFF parameter database curated by Software for Chemistry & Materials B.V. (SCM). [185] Subsequently the bounds are extended to also include a window of ± 20 % around the values from the Joshi *et al.* force field. Note that such choices are subjective for lack of a better alternative: there are no established defaults for the parameter bounds. The list of active parameters and their bounds can be found in Table S3 of the Supporting Information.

4.2.4 Optimization Settings

ReaxFF parameters are typically calibrated by minimizing a loss function with a numerical optimizer. We have developed a novel loss function for this work, which will be discussed in Section 4.3. Here we focus on the details of the numerical optimization algorithm.

Several optimization algorithms have been proposed to refine ReaxFF parameters. The original method proposed by van Duin was a deterministic algorithm that optimized one parameter at a time with a parabolic extrapolation.[201] More recent algorithms are stochastic, which makes them more robust to the non-trivial structure of a standard ReaxFF loss function, such as many local minima and discontinuities.[41,114–118] These difficulties arise from the small discontinuities in the ReaxFF energy itself, and the noisy sensitivity of geometry optimizations (while training) to the ReaxFF parameters. [41,136] These difficulties are still present in this work, and therefore we use a stochastic derivative-free optimizer that has proven its effectiveness, i.e. the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES). [41,48,202,203]

The CMA-ES settings in this work follow the best practices from the literature. [48,202,203] The algorithm is repeated 40 times, starting from Joshi's parameters, to reduce the risk of getting stuck in an unfavorable local minimum. These repetitions are also used to test the robustness of the new loss function proposed in Section 4.3. The CMA population size is set to the value recommended by Hansen, $\lfloor 4 + 3 \ln N_{\text{par}} \rfloor = 14$, where $N_{\text{par}} = 36$ is the number of activated force field parameters. [48,202,203] ParAMS communicates dimensionless parameters to CMA-ES, by linearly transforming the original parameters so that their bounds all become $[0, 1]$. The initial CMA-ES step size, in these dimensionless parameters, is set to 0.2. This is sufficient to let the algorithm randomize the parameters in the first few CMA iterations, after which it starts to converge, thereby guaranteeing an initial exploration of the parameter space. Each CMA run is terminated after 1000 iterations and the parameters with the lowest loss value are selected for further analysis.

Before evaluating the loss function in each CMA iteration, all structures in the training set are optimized with the parameters generated by CMA. The maximum number of geometry iterations is set to 500, which is much higher than the default value of 30 in ParAMS. For the training set in this work, a lower setting, such as 50, 100 or 200, produces force fields that are overfitted to this lower number of geometry steps. Each CMA run is performed on 18 cores (Intel Xeon Gold 6140), with `ParallelLevels ParameterVectors=14` and `Jobs=2`. This results in a slight over-commitment of the cores, which is normally not recommended, but it improves the overall efficiency in this case, which can be understood as follows. The CMA-ES algorithm synchronizes after each iteration, resulting in idle time when the members of the population require different CPU times. This is generally the case for ReaxFF, since the number of required geometry steps depends strongly on the parameters. By over-committing the cores, the idle time is reduced, resulting in a more efficient use of resources.

4.2.5 Validation Set

i **Structures** The validation set is taken from a dataset by Raybaud *et al.*, available on NOMAD, containing alumina structures optimized with VASP, using the same level of theory as the training data, [143,204] except that a plane-wave cutoff of 400 eV was used. This set contains 53 new γ-Al_2O_3 surface structures not used for parameter optimization, with different numbers of adsorbed water molecules. The surface orientations comprise (001), (111), $(110)_\ell$ and $(110)_b$ as shown in Figure 4.1. The subscripts ℓ and b are used to distinguish between lateral and basal surfaces, respectively, which feature different Brønsted and Lewis acid sites. [143,162] The set also includes a bulk γ-Al_2O_3 model and an isolated water molecule was added in this work, using consistent VASP settings. A complete list of structures is provided in Table S4 of the Supporting Information.

Recent developments in alumina characterization have revealed an ambiguity in the terminology used in older works. In particular, earlier spinel models considered the (100), (010) and (001) surfaces to be equivalent, but it has recently been shown from non-spinel models that this is not the case. [143] To remain consistent with published datasets and with the optimizations performed in this work, the notation remains (001) for surfaces with this orientation in the validation set and (100) for surfaces with this orientation in the training set. However, they are structurally equivalent.

ii **Property Extraction** Properties are extracted using the same methodology and classification as described in Section 4.2.2. The resulting validation set contains 16745 distances, 9513 angles, 101 adsorption energies, and 6 other energies, for a total of 26365 validation targets. Figures S3, S5, S6 and S7 in the Supporting Information show the histograms of these data, whereas Table S5 lists the individual reaction energies. The final force field in this work and the original one by Joshi *et al.*[174,176] are validated by comparing these geometrical properties and energies to the VASP reference data.

4.3 Balanced Loss

ReaxFF parameters are conventionally calibrated by minimizing a loss function L, which is often a weighted Sum-of-Squares Error (SSE) or Root-Mean-Squared Error (RMSE):

$$\mathcal{L}_{\mathrm{SSE}}(\mathbf{x}) = \sum_{i=1}^{N} w_i s_i(\mathbf{x}) \tag{4.4}$$

$$\mathcal{L}_{\mathrm{RMSE}}(\mathbf{x}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} w_i s_i(\mathbf{x})} \tag{4.5}$$

where $\{s_i\}_{i=1}^{N}$ are the squared residuals:

$$s_i(\mathbf{x}) = r_i^2(\mathbf{x}) = \left(\frac{y_i - \hat{y}_i(\mathbf{x})}{\sigma_i}\right)^2 \tag{4.6}$$

and where the sum over i runs over all items in the training set. In every term, the property value i is calculated with a reference method (y_i) and ReaxFF ($\hat{y}_i$). Through the ReaxFF property values, the loss function depends on a vector of adjustable ReaxFF parameter vector $\mathbf{x}$. Note that CMA-ES is insensitive to the application of any monotonically increasing transformation of the loss function, so from CMA's perspective, $\mathcal{L}_{\mathrm{RMSE}}(\mathbf{x})$ and $\mathcal{L}_{\mathrm{SSE}}(\mathbf{x})$ are equivalent.

The constant σ_i is a configurable scaling factor with the same unit as the property of y_i to make the residual r_i dimensionless. In ParAMS, σ_i is only used as a reasonable order of magnitude for the corresponding y_i. [188] The weight w_i controls the importance of each training set entry in the total loss function. In principle, one can absorb w_i into σ_i or vice versa. The main motivation for supporting both factors in ParAMS is to cater to different user groups, some of which may prefer one over the other. This may seem surprising, since textbook treatments of the least-squares method do not mention the weights w_i and only introduce σ_i as a measurement error. However, a basic assumption of the standard least-squares method does not hold here: Our data have no measurement errors. Any discrepancies between the training data and ReaxFF are due to systematic errors, mainly in ReaxFF and, in principle, also in the model used to compute the training data. At first glance, setting the weights seems straightforward: the more important an item in the training set, the higher its weight should be. However, there are different (possibly competing) motivations for adjusting the weights.

A first purpose of the weights is to compensate for an imbalance in the training set. For example, our training set contains many more distances and angles than energies, while the energies are also important. This imbalance can be addressed by classifying the data into categories and setting the weight to the inverse of the number of elements in each category. This strategy is common in the context of machine learning potentials, e.g. when atomic forces are much more abundant than molecular energies. [124–126] A second purpose of the weights is to emphasize the importance of some residuals. For example, when an initial optimization leads to parameters for which some residuals r_i are perceived to be too large, one may increase the corresponding weights w_i and re-optimize. Assigning different weights to subsets of data is also known in the field of multi-objective optimization as the "scalarization" of multiple objectives into a single loss function. [205] Unlike scalarization methods, multi-objective evolutionary algorithms do not assume any tradeoffs between categories *a priori* and instead find many Pareto-optimal solutions. [206]

While seemingly intuitive, manual weight adjustment becomes intractable when many weights need to be adjusted differently. Due to the non-linear response of the residuals to the weights,[129] multiple combinations of weights must be tried before one reaches the residuals of interest. If the model has insufficient functional flexibility, it may even be impossible to reach the desired residuals. In addition, when some residuals of interest decrease, others will inevitably increase. It is difficult to predict which residuals will increase and by how much, and this can force the operator to keep adjusting the weights. In practice, this resembles a cat-and-mouse game between weights and residuals. Manually adjusting the weights also provides little insight into the optimization problem: If some residuals are large, there is no straightforward way to understand whether their weights should be increased further, or whether the model is simply unable to reproduce the training data.

It is clear that the development of a training set alone is rarely sufficient to find the optimal parameters. Only if the training data are completely homogeneous, one can simply set all weights equal. This is typically not the case for ReaxFF training set, which contain different types of data, such as distances, angles and energies in this work. Assigning weights to the data is therefore an unavoidable and potentially tedious task before and during a ReaxFF parameter optimization. [119,127,128]

To simplify the tedious adjustment of weights, we introduce a new method, hereafter called “Balanced Loss”. As will be demonstrated in the results, this method allows for a swift balancing of the training set, and we believe that this methodology will be equally beneficial for other optimization problems facing similar challenges. The Balanced Loss method introduces a new loss function and an intuitive workflow to balance the data and to gain more insight into how well the model can reproduce subsets of the training data.

Balanced Loss requires a classification of the training data into categories. Technically, the categories are C mutually exclusive and exhaustive subsets: $S_c \, \forall c \in \{1...C\}$. They should be defined so that residuals within a category respond in roughly the same way to a change in the model parameters. For example, one might expect that all bond lengths in a training set, while not exactly the same, do respond similarly to changes in the ReaxFF parameters. With this partitioning, the Balanced Loss function is defined as:

$$\mathcal{L}_{\mathrm{BL}} = \tau f^{-1}\left(\sum_{c=1}^{C} f\left(\frac{R_c}{\tau}\right)\right) \tag{4.7}$$

where R_c is the RMSE on the entries in category c:

$$R_c = \sqrt{\frac{1}{|S_c|}\sum_{i\in S_c} s_i} \tag{4.8}$$

$\mathcal{L}_{\mathrm{BL}}$ is dimensionless by construction. The function f and its inverse must be monotonically increasing functions, and by default $f(x) = \exp(x)$ is used, which will be denoted as the Log-Sum-Exp (LSE) form, referring to the mathematical operations in Equation (4.7). To illustrate the benefits of the Log-Sum-Exp form, all parameterizations will be repeated with two other forms of f: $f(x) = x^2$, denoted as Root-Sum-Square (RSS) and $f(x) = x$, denoted as Identity-Sum-Identity (ISI). Note that the RSS form makes Balanced Loss formally equivalent with a standard loss function in Equation (4.5), with $w_i = \frac{N}{S_c}$, where c is the category to which training set item i belongs.

By default, $\mathcal{L}_{\mathrm{BL}}$ is thus a Log-Sum-Exp function, a well-established smooth approximation of the maximum over multiple inputs. It is popular in the machine learning context [207] and it has been used for scalarization of multi-objective problems. [208] Here, the inputs to Log-Sum-Exp are all the R_c values. The hyperparameter τ, sometimes called the effective temperature, controls the smoothness of the approximation of the maximum. In the "cold" limit $\tau \to 0$, Log-Sum-Exp loses its smoothness and reduces to the maximum over all R_c. The parameter τ appears in two places, such that $\mathcal{L}_{\text{BL}} = \mathcal{L}_{\text{RMSE}}$ in the trivial case of one category and $w_i = 1 \; \forall i$. Throughout this paper we have used $\tau = 1$.

Unlike $\mathcal{L}_{\text{SSE}}$ and $\mathcal{L}_{\text{RMSE}}$, user-defined weights w_i are missing from $\mathcal{L}_{\text{BL}}$, which implies that σ_i must play a slightly different role. We propose to set each σ_i to the desired accuracy of the corresponding entry y_i in the training set, which resembles its meaning in conventional least-squares methods. To make the distinction with σ_i in other contexts, we call them "tolerances" in the context of Balanced Loss, because they represent the magnitudes of residuals one is willing to tolerate. At this stage, simply defining tolerances may seem like wishful thinking, but it will become clear later that Balanced Loss helps finding a consistent set of ReaxFF parameters and tolerances. Our definition of σ_i (as the desired accuracy) also facilitates the interpretation of R_c: It expresses, in the RMS sense, the average ratio between the actual and the desired accuracy. In the ideal case, after optimizing the parameters, one obtains $R_c = 1 \; \forall c$.

Given the interpretation of R_c, the Log-Sum-Exp form of Balanced Loss is easily motivated. If one of the R_c values is much higher than all others, one finds $\mathcal{L}_{\text{BL}} \approx R_c$, *i.e.*, category c dominates the loss function. If the optimization algorithm explores a region of the parameter space where category c dominates, it will focus only on reducing R_c, with other categories acting at best as a form of regularization. This is a desirable feature, since category c is then the worst reproduced subset of the training data, and therefore deserves the optimizer's full attention.

One may also understand the effect of Log-Sum-Exp by comparing the gradients of $\mathcal{L}_{\text{BL}}$ and $\mathcal{L}_{\text{RMSE}}$ with respect to the ReaxFF parameters:

$$\begin{aligned} \frac{\partial \mathcal{L}_{\text{BL}}}{\partial x_k} &= \sum_{i=1}^{N} \frac{\partial \mathcal{L}_{\text{BL}}}{\partial s_i} \frac{\partial s_i}{\partial x_k} \\ \frac{\partial \mathcal{L}_{\text{RMSE}}}{\partial x_k} &= \sum_{i=1}^{N} \frac{\partial \mathcal{L}_{\text{RMSE}}}{\partial s_i} \frac{\partial s_i}{\partial x_k} \end{aligned} \tag{4.9}$$

Both loss gradients are linear combinations of the gradients of squared residuals, $\frac{\partial s_i}{\partial x_k}$, but they combine them with different "weights":

$$\begin{aligned}\frac{\partial \mathcal{L}_{\text{BL}}}{\partial s_i} &= \frac{\exp\left(\frac{R_d}{\tau}\right)}{\sum_{c=1}^{C}\exp\left(\frac{R_c}{\tau}\right)}\frac{\tau}{2\mid S_d\mid R_d} \text{ with } i\in S_d \\ \frac{\partial \mathcal{L}_{\text{RMSE}}}{\partial s_i} &= \frac{w_i}{2N\mathcal{L}_{\text{RMSE}}}\end{aligned} \tag{4.10}$$

In the case of $\mathcal{L}_{\text{RMSE}}$, the weight $\frac{\partial \mathcal{L}_{\text{RMSE}}}{\partial s_i}$ is simply proportional to the user-defined weight w_i. For Balanced Loss, however, the weight $\frac{\partial \mathcal{L}_{\text{BL}}}{\partial s_i}$ contains a new and crucial factor:

$$P_d = \frac{\exp\frac{R_d}{\tau}}{\sum_{c=1}^{C}\exp\frac{R_c}{\tau}} \tag{4.11}$$

with

$$\sum_{d=1}^{C} P_d = 1. \tag{4.12}$$

This factor is known as SoftMax, a continuous generalization of the ArgMax function, used to identify the position of a maximum in an ordered list. [209] This shows how Balanced Loss borrows a strategy from reinforcement learning, known as the Gradient Bandit Algorithm: At each iteration in the optimization, the most violated subset of the training data determines the action, [210] in this context action being the direction in which the parameters must evolve. The analogy between P_d (in $\mathcal{L}_{\text{BL}}$) and the user-defined weights w_i (in $\mathcal{L}_{\text{RMSE}}$) also suggests another interpretation. Instead of a human operator tuning the weights w_i, as in the cat-and-mouse metaphor introduced above, Balanced Loss adjusts the weights algorithmically within a single optimization run.

So far, we have assumed that one simply sets the tolerances σ_i to the desired accuracy of the corresponding y_i. However, such a choice may be subjective and incompatible with the capabilities of the ReaxFF model to be trained. In practice, we recommend such "naive" tolerances σ_i as a first guess. Parameter optimization can then be used to test these expectations. To do so, we recommend the following workflow:

W.1 First gather all the elements of a conventional parameter optimization: (i) the model, (ii) the training data, (iii) a selection of parameters to optimize and their bounds, (iv) an initial guess of the parameters, and

(v) an optimization algorithm. In this paper, all these elements are described in Section 4.2.

W.2 Then define the additional elements needed for Balanced Loss: the categories of training data and an initial configuration of the tolerances σ_i. Such choices are domain-specific, but a few general recommendations can be given, in addition to the ones discussed above. It is convenient to have data with consistent units within one category c and to assign the same tolerance to all its members, for which the symbol σ_c will be used below. Furthermore, it is useful to introduce categories for data that deserve special attention, e.g. with key properties for the intended application of the force field, or with properties that are harder to reproduce than others. By placing these data in separate categories, their RMSEs are easily monitored and large errors within these categories will be prioritized during the optimization.

For ReaxFF, one can introduce different categories for distances, angles and energies. In this paper, the categories are more fine-grained: All internal coordinates are classified by the chemical elements and the bond types involved. In fact, we categorize all training data as they were introduced in Section 4.2.2: the 4 bond categories are Al-O, Al-Al, O-H and O···H, the 6 angle categories are O-Al-O, Al-O-Al, Al-O-H, Al-O···H, H-O···H and H···O···H, and the energy categories are **BSH**, **GEH**, **GSH**, **SUR** and **FOR**. Note that a regular covalent bond is denoted by minus sign (—) and a hydrogen bond by three dots (···). The tolerances σ_c will be described in Section 4.4.

W.3 Finally, minimize the Balanced Loss function and analyze the R_c values of the optimal parameters. When one category keeps dominating the loss function throughout the optimization, the only possible explanation is that the corresponding tolerances σ_c were set too small. There is no way to lower R_c because $\mathcal{L}_{\mathrm{BL}}$ already ignores all other categories. The only option left is to accept that the model cannot reproduce items in category c with the desired accuracy, and to adjust one's expectations by increasing the corresponding tolerances σ_c. One can now repeat the parameter optimization and re-evaluate the result, possibly repeating the exercise a few times. Unlike tuning the weights in a conventional loss function, these repeated optimizations provide insight: They inform the human operator about the capabilities of the model and help manage expectations. It may also happen that some R_c end up well below 1, in which case we do not recommend decreasing the corresponding σ_c. Such a fortuitous outcome should not affect the desired accuracy.

Note that steps W2 and W3 in the above workflow involve (possibly subjective) human decisions, and therefore cannot be replaced by an autonomous algorithm. This is an unavoidable aspect of multi-objective problems: One has to decide on a compromise between different categories. The overall goal of Balanced Loss is to facilitate finding suitable compromises.

4.4 Results & Discussion

4.4.1 Balanced Loss Optimization Procedure

This section illustrates how the optimization workflow of Balanced Loss leads to a competitive ReaxFF parameterization, using the alumina training set as a realistic example. In addition to the final force field, the intermediate steps provide insight into the capabilities of ReaxFF.

The data in the training set, described in Section 4.2.2, have been grouped into categories as described in Section 4.3 and as summarized in the leftmost column of Table 4.2. The optimizations are carried out in two stages, initial and final, which differ only in the tolerances σ_c.

Each entry is given an initial chemically relevant tolerance, σ_c, equal to the default sigma value from ParAMS, as shown in Table 4.2. With these tolerances, the ReaxFF parameters were optimized 40 times using different random seeds to produce independent solutions.

Figure 4.2(a) shows the evolution of the Balanced Loss during the 40 CMA runs. Figure 4.2(c) presents the $R_c\sigma_c$ values (category RMSEs with units) for the 40 optimized parameter vectors. The curves and data points are colored according to the loss value of the best parameter vector of each run. Of the 40 CMA runs in the initial stage, two are clearly worse than all others, presumably converging to unfavorable local minima. All 38 remaining runs produce comparable RMSE values, but are not identical, which is the expected behavior. ReaxFF loss functions are known to exhibit many local minima and apparent noise due to high sensitivity of the geometry optimizations in the training set to the ReaxFF parameters. [41] Adsorption energies in the categories **BSH** and **GSH** are the highest relative to their tolerance,

Category	Unit	Initial σ_c	Final σ_c
Al-O	Å	0.05	0.07
Al-Al	Å	0.05	0.10
O-H	Å	0.05	0.05
O ··· H	Å	0.05	0.12
Al-O-Al	deg	2.0	5.0
Al-O-H	deg	2.0	5.0
O-Al-O	deg	2.0	5.0
Al-O···H	deg	2.0	7.0
H-O···H	deg	2.0	7.0
H···O···H	deg	2.0	7.0
BSH	mol^{-1}	1.25	4.0
GEH	mol^{-1}	1.25	4.0
GSH	mol^{-1}	1.25	4.0
SUR	mol^{-1}	1.25	3.0
FOR	mol^{-1}	1.25	3.0

Table 4.2: Tolerances used in the Balanced Loss optimization

σ_c. The performances in all other categories have less effect on the optimized parameters, simply because these errors are closer to their tolerance. This means that, in this initial stage, the CMA runs train almost exclusively on the adsorption energies. It is therefore highly unlikely to find ReaxFF parameters that can further lower RMSE on the adsorption energies, let alone reach the tolerance of 1.25 $\mathrm{kcal\,mol}^{-1}$. Also, for all other categories, the initial tolerances seem too optimistic, which will be addressed in the next stage.

For the second (and final) stage, the tolerances are revised, as shown in the last column of Table 4.2, to be more consistent with what ReaxFF could achieve in the initial stage. Without Balanced Loss, these tolerances can only be set by expert judgment, which is greatly facilitated here by the feedback from the initial stage. Figure 4.2(b) shows that the revised tolerances result in more consistent loss values across all 40 parameterizations. This is also reflected in the RMSEs in Figure 4.2(d), most of which exhibit less scatter. In other words, for most categories, the 40 parameterizations in the final stage are comparable, with the category **GEH** being the most notable exception.

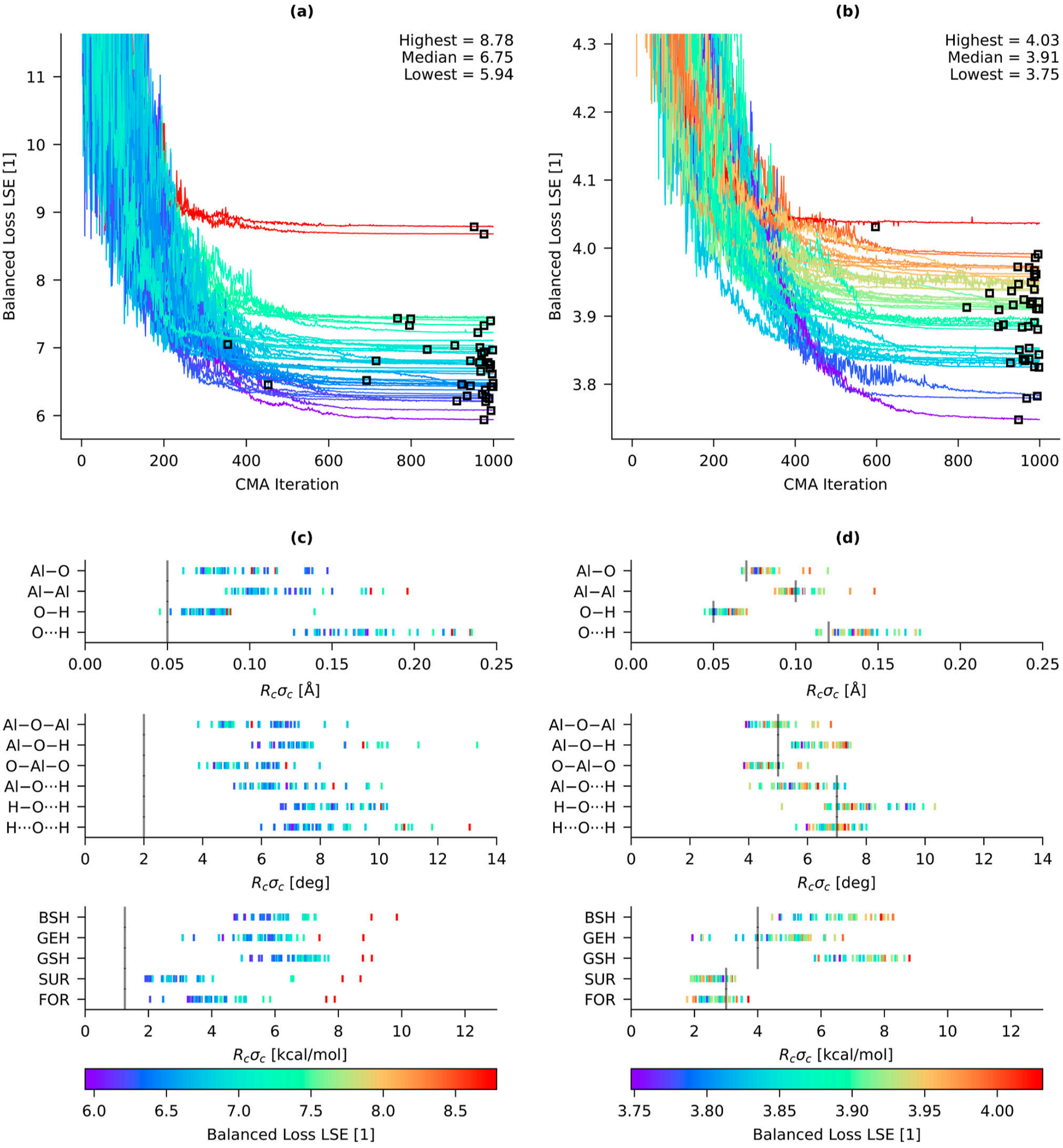

Figure 4.2: Balanced Loss as a function of CMA iteration (lowest value within the population) of the 40 parameterizations in the initial (a) and final (b) stage. The lowest value along each trajectory is indicated by a black square. The dimensioned category RMSEs, $R_c\sigma_c$, of the best parameter vector of each of the 40 parameterizations in the initial (c) and final (d) stage. All data are color coded by the loss value of the best point along the trajectory. The color bar of panel (c)/(d) is also applicable to panel (a)/(b). Grey vertical lines in panel (c)/(d) denote the tolerances for the corresponding category.

Note that the absolute value of the Balanced Loss values cannot be compared between Figure 4.2(a) and Figure 4.2(b), because the two optimization stages use different tolerances. One could slightly tweak the tolerances further to bring them closer to the errors on the training set, but this would amount to relatively small adjustments that we do not expect to lead to significant improvements.

To illustrate the importance of the Log-Sum-Exp form of Balanced Loss, we have performed the 40 CMA optimizations in six different ways: with initial and final tolerances and using different functions $f : f(x) = \exp(x)$ (Log-Sum-Exp or LSE, the default, same results as above), $f(x) = x^2$ (Root-Sum-Square or RSS) and $f(x) = x$ (Identity-Sum-Identity or ISI). For comparison, the LSE form of Balanced Loss is computed for all the 240 optimized parameter vectors, and their distribution is shown in Figure 4.3(a). In the initial stage, the function f has a significant influence. The choice of the function f determines the compromise between the RMSEs of the individual categories: In the case of RSS and ISI, the parameterization no longer exclusively prioritizes the adsorption energies in the categories **BSH** and **GSH**, resulting in higher RMSEs for these categories, as illustrated in Figure 4.3(b). The results in this figure do not reveal whether the poor performance of ReaxFF for the categories **BSH** and **GSH** can be remedied by giving these categories a higher weight in the loss function, or whether they are high due to intrinsic limitations of the model. The Log-Sum-Exp form automatically resolves this ambiguity. Since this form approximates the maximum over all R_c, this loss function is dominated by the categories **BSH** and **GSH**, resulting in the lowest possible RMSE for these categories. It is simply impossible to give these categories a higher weight, leaving no other option than to assign more humble tolerances.

Figure 4.3(a) reveals two additional insights. First, the results become less sensitive to the choice of the function f in the final stage. The RMSEs for all categories are close to the final tolerances, meaning that the argument of the function f (for the optimized parameters) is close to one, reducing the importance of the non-linearity of f. Second, with $f(x) = x^2$ (RSS), the loss function is mathematically equivalent to Equation (4.5), a standard loss functions used for ReaxFF. This implies that the optimized parameters in this work can also be found with a more conventional loss function, when the weights and sigmas are set consistently with the tolerances in Balanced Loss.

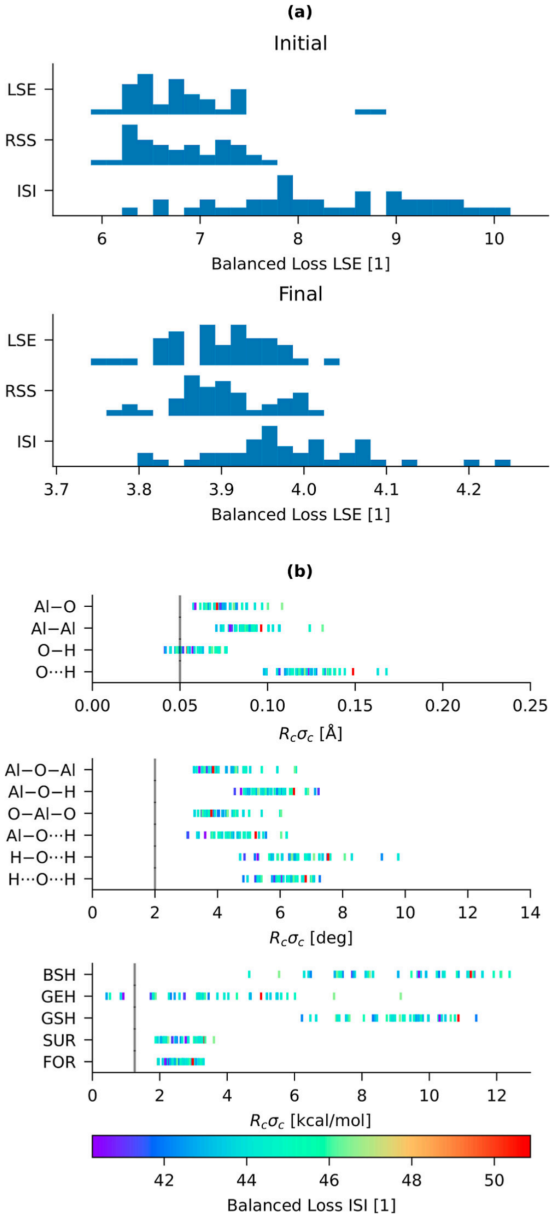

Figure 4.3: (a) Distribution of Balanced Loss values, computed with $f(x) = \exp(x)$, for parameter vectors optimized with different functions f in Equation (4.7): $f(x) = \exp(x)$ (LSE), $f(x) = x^2$ (RSS) and $f(x) = x$ (ISI). The histograms are computed for both the initial and final tolerances defined in Table 4.2. (b) The dimensioned category RMSEs, $R_c\sigma_c$, of the best parameter vector of each of the 40 parameterizations in the initial stage using $f(x) = x$ (ISI).

Hence, the added value of Balanced Loss is essentially the feedback provided from the initial stage, which facilitates the configuration of the tolerances.

Figure 4.4 shows the distribution of the 40 optimized parameter vectors, after transforming them to their dimensionless form. Although the 40 CMA runs converge to approximately the same Balanced Loss value, the corresponding parameters are not necessarily similar. Some parameters, such as `Al.O:D_e^sigma` or `Al.O:p_ovun1`, have a delineated range of optimal values. However, most parameters can be found across the entire interval of allowed values. This does not mean that all these parameters are completely random: They could be correlated, which is not apparent in the individual histograms. In any case, the optimal parameters are degenerate to some degree, which has also been observed in previous ReaxFF parameterizations on other chemical systems. [41,114,123,136]

Since ReaxFF is at least partially inspired by physical principles, one might hope that all parameters always converge to the same values. However, Sethna *et al.* have extensively shown in their work on "sloppy models" that broad parameter distributions are virtually always found for models with more than a few parameters, across different scientific disciplines. [211–214] This is a universal pattern, regardless of the degree of physical interpretation that the model parameters may have. It is observed that some degrees of freedom in the parameter space of complex models are systematically ill-defined, not due to a lack of data, but because nearly the same model predictions (for all possible inputs) are found for different parameter vectors. As a consequence, predictions on unseen data are robust, despite uncertainties in the parameters. For the Electronegativity Equalization Method (EEM), which is included in ReaxFF, this parameter degeneracy has been investigated in more detail. [215] ReaxFF also has the characteristics of "a sloppy model", as illustrated by Figure S8 in the Supporting Information: All the optimized parameter vectors of the 40 CMA runs (LSE, final stage) perform reasonably well for the validation set, despite the fact that they represent different local minima in the parameter space. This in itself is not a limitation when the model is used for simulations, but it obviously makes any direct interpretation of the parameters impossible.

It is also noticeable that several parameters have a high probability of converging close to the bounds. One might deduce that the parameter bounds are too narrow and the optimizer is trying to move the parameters to an optimum beyond the bounds, but this is not the only possible cause. Note that the components of the best parameter vector over all 40 runs, the circles in Figure 4.4, are not necessarily close to the bounds, even if the remaining near-optimal values cluster near the edges. Examples of this pattern are “Al.O:p_be1” and “Al.O.H:Theta_0,0”. A deeper investigation, beyond the scope of the current work, is needed to understand why a disproportionate number of near-optimal solutions is found near the bounds. For example, this could also be related to an optimizer inefficiency when the parameters approach their bounds, and addressing this problem may make the optimization more efficient. It should also be noted that some initial parameter values start of close to the interval bounds regardless of the boundary extension as explained in Section 4.2.3. Since some of the initial parameters have values close to zero, the effect of the boundary extension is negligible. The most prominent examples are “Al.O:p_be2”, “Al.H.O:p_val1”, “Al.H.O:p_val2” and “Al.O.H:p_val2”.

For the remainder of this work, the best parameter vector from the final stage is used for all calculations, i.e. corresponding to the lowest square in Figure 4.2(b) and the circles in Figure 4.4. The selected parameters are given in the last column of Table S3 in the Supporting Information.

4.4.2 Force field validation

Figure 4.5 offers a first visual impression of how the new force field improves the prediction of dissociative water adsorption on alumina. It shows a high-hydration structure of the γ-Al_2O_3 (001) slab model, labeled `gamma_surf-001_03w` in the validation set, geometry-optimized with the reference method (DFT calculation with VASP), the initial force field by Joshi *et al.*[176] and the new force field in this work. The Joshi *et al.* force field predicts a severe deformation of the alumina structure, and already desorbs water molecules in this static calculation while all water should be adsorbed according to reference calculation. In contrast, the new force field in this work predicts a geometry that is visually deviating only slightly from the DFT reference, and the water molecules remain adsorbed on the surface.

The performance of the new force field can be evaluated more in detail by analyzing the distributions of categorized data in the training and validation sets, and their deviations from the DFT reference. Table 4.3 shows the RMSEs between force field and reference data, for each category, for the training and

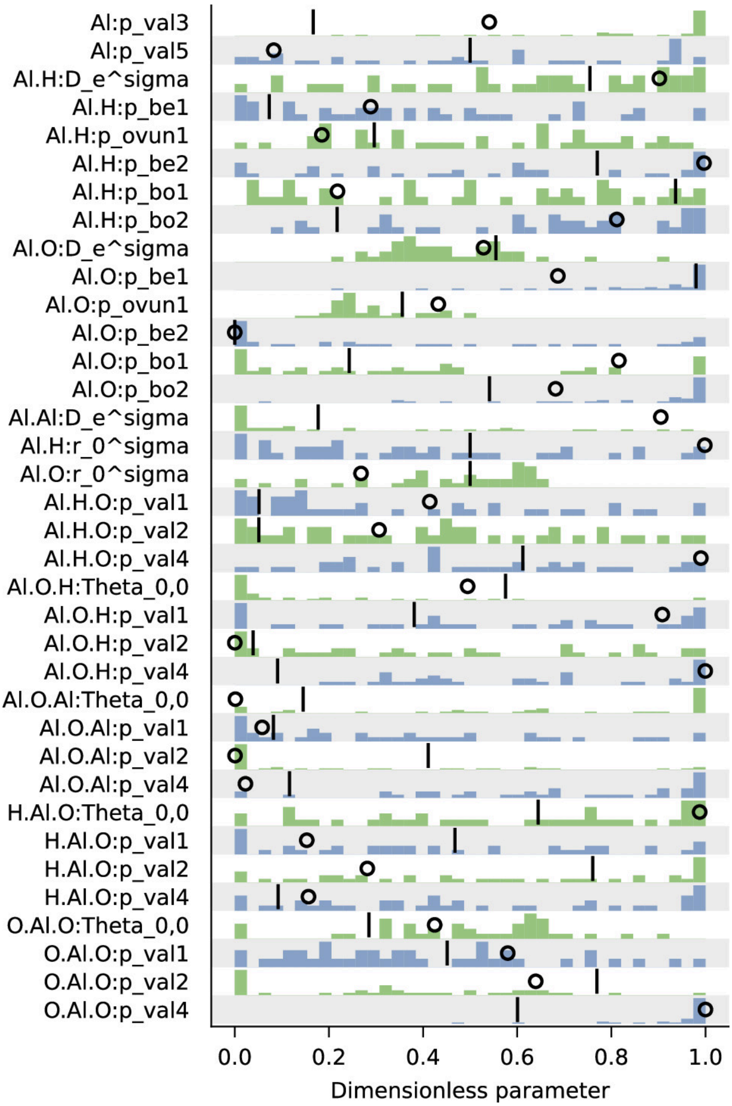

Figure 4.4: Histograms of the 36 components of the 40 optimized dimensionless parameter vectors in the final stage. The parameter components are made dimensionless by a linear transformation, such that zero corresponds to the lower bound and one corresponds to the upper bound. The bounds are listed in Table S3 in the Supporting Information. The initial values are marked with vertical black bars. The values corresponding to the lowest loss (over all 40 runs) are marked with circles.

validation set, and for the initial force field by Joshi *et al.*[174,176] and the

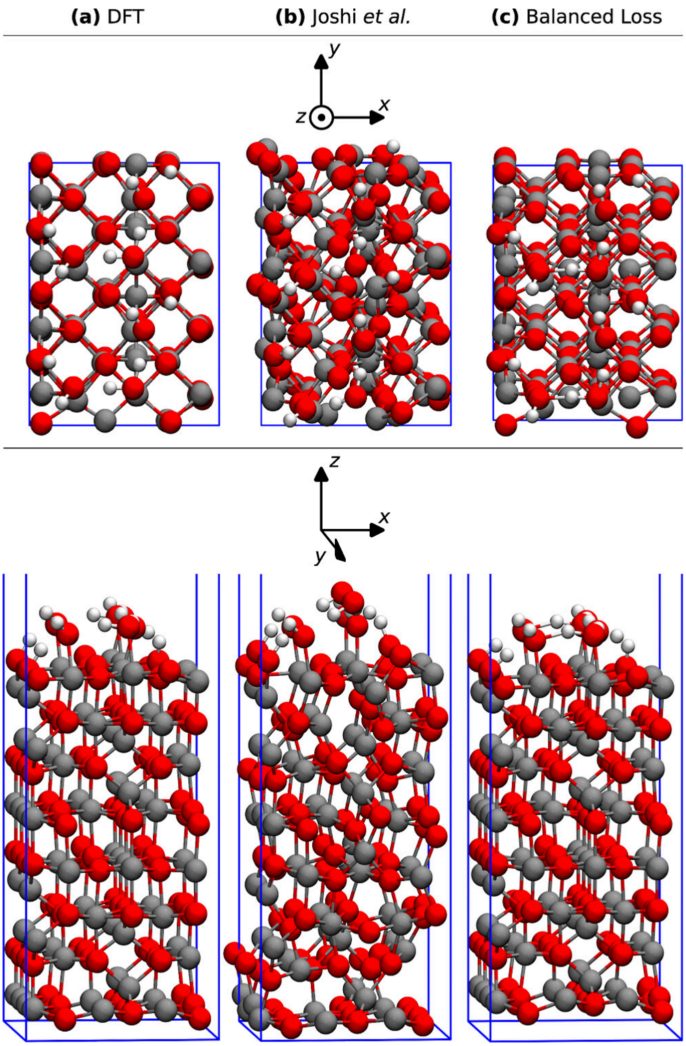

Figure 4.5: Optimized structures of the γ-Al_2O_3 (001) surface from the validation set at an coverage of 13.0 nm^{-2}, computed with different models: (a) DFT reference, (b) ReaxFF parameters by Joshi *et al.* and (c) ReaxFF parameters obtained with Balanced Loss. Al=gray, O=red, H=white.

one optimized in this work. To make the RMSEs directly comparable, only data related to γ-Al_2O_3 were taken from the training set since the validation set also contains only γ-Al_2O_3. In addition, only data categories (rows) are included that exist in both data sets. For reference, also the standard deviations on the reference data per category are reported. Figure 4.6 shows parity plots for all categories of internal coordinates in the validation set. In addition, water adsorption energies on γ-Al_2O_3 surfaces in the training and validation sets are shown in Figure 4.7. Not all data categories from the training set exist for the validation set, because the validation set is focused on adsorption on γ-Al_2O_3 only. This is reflected in Table 4.3, Figure 4.6 and Figure 4.7 by only considering γ-Al_2O_3 structures. Table S6 and Figure S3 in the Supporting Information contain the results omitted here, i.e. not involving γ-Al_2O_3 surfaces, for which a direct comparison to the validation set is not possible.

The RMSEs in Table 4.3 show that the new force field significantly reduces the errors on the bond lengths compared to the Joshi *et al.* force field. [174,176] A subset of the bonds is broken after geometry optimization with the force fields, which is not fully visible in Figure 4.6(a) and Figure 4.6(c),

category	unit	training				validation			
		ref SD	Joshi RMSE	BL RMSE	#	ref SD	Joshi RMSE	BL RMSE	#
Al-O	Å	0.16	0.45	0.09	1405	0.10	0.40	0.06	2102
Al-Al	Å	0.30	0.40	0.11	6759	0.31	0.40	0.11	13588
O-H	Å	0.03	0.31	0.07	352	0.02	0.18	0.06	704
O ⋯ H	Å	0.18	0.85	0.17	203	0.18	0.90	0.25	351
Al-O-Al	deg	21.2	17.0	4.4	1756	20.0	16.6	3.8	2744
Al-O-H	deg	12.2	20.4	6.9	598	9.3	18.7	8.3	1217
O-Al-O	deg	29.4	16.7	4.5	2943	30.9	17.4	3.9	4656
Al-O ⋯H	deg	18.4	16.9	6.5	301	17.6	17.6	6.6	554
H-O⋯H	deg	17.5	27.4	12.7	142	17.3	29.4	14.7	254
H⋯O⋯H	deg	31.1	28.9	4.4	65	38.0	33.7	7.7	88
GSH	$\frac{kcal}{mol}$	15.0	55.7	6.6	37	16.4	54.4	8.3	101
SUR	$\frac{kcal}{mol}$	6.1	11.3	2.9	8	1.4	4.7	1.4	6

Table 4.3: Comparison of Root-Mean-Square-Errors (RMSEs) of the Initial Force Field by Joshi et al. and the Force Field Optimized in This Work (BL). The RMSEs are computed for training and validation sets and are grouped per data category. For reference, the standard deviation (SD) on the reference data is included.

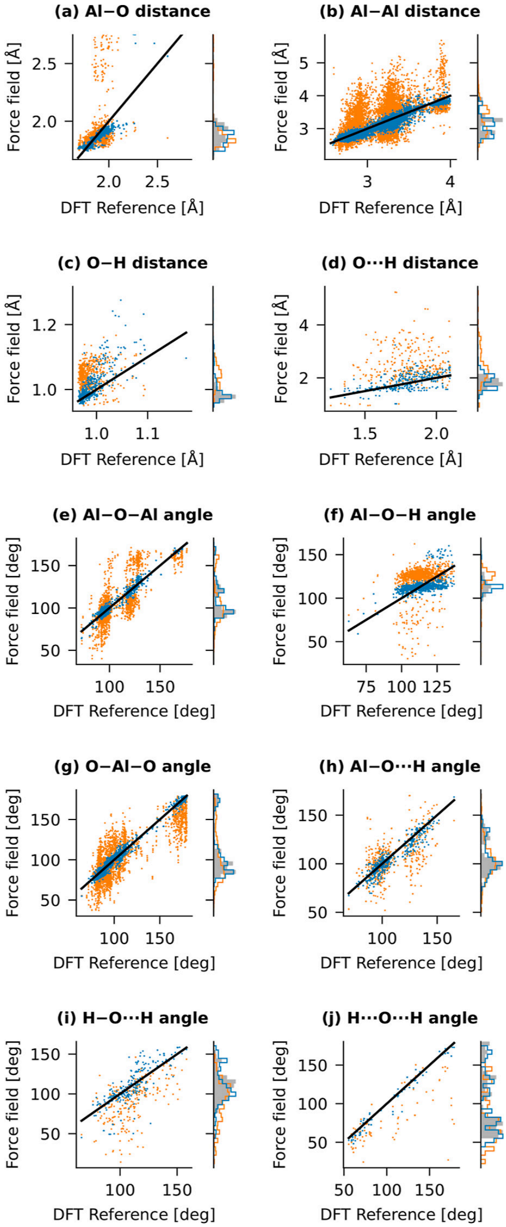

Figure 4.6: Parity plots of the reference internal coordinates from the validation set versus the force field predictions: initial parameters by Joshi *et al.*[174,176] (orange) and optimal parameters obtained with Balanced loss (blue). The parity line is plotted as a black solid line. In panels (a) and (c) the vertical axis is manually limited to only show bonded distances. The initial parameters from Joshi result in many broken bonds, which are omitted for the sake of clarity.

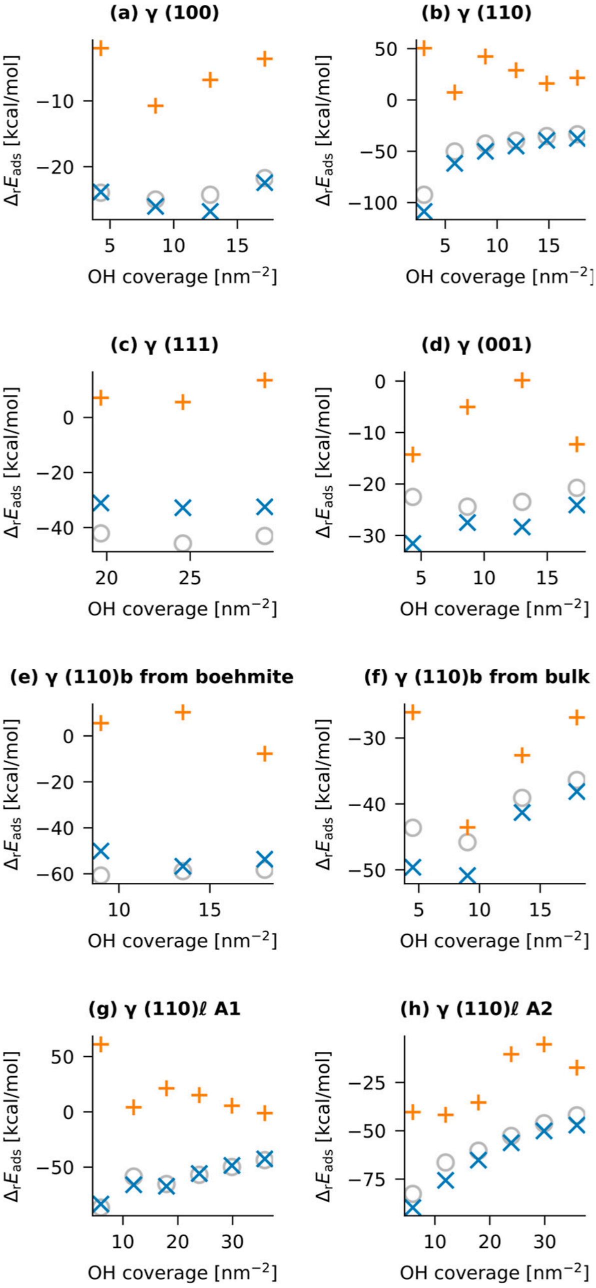

Figure 4.7: Water adsorption energy per water molecule on γ-Al_2O_3 surfaces as a function of the OH coverage, using the bare surface as the reference, i.e. using $m_i = 0$ in Equation (4.3);. Results are computed with DFT (gray circle), Joshi FF (orange plus), FF from this work (blue cross). Panels (a) to (c) are adsorption energies from the training set, whereas (d) to (h) are derived from the validation set.

because this would require an impractical scale for the vertical axes. In the validation set, 14.3 % of the Al-O bonds and 1.1 % of the O-H bonds are broken with the Joshi *et al.* force field. With the new force field proposed here, these percentages reduce to 0.0 % and 0.3 %, respectively. These percentages are consistent with the visualization in Figure 4.5(b) and confirm that the Joshi γ-Al_2O_3 surfaces force field cannot preserve the structural integrity of the (hydrated) γ-Al_2O_3 slabs. By consequence, this force field also performs poorly for other categories (angles and energies), for which a correct bonding topology is required.

The parity plots of the distances in Figure 4.6(a), Figure 4.6(c) and Figure 4.6(d) also reveal that even the new force field captures the variations in bond lengths only approximately. This is also confirmed by the fact that the RMSEs of the distances in Table 4.3 are of the same order as the standard deviation on the distances in the reference data. For the O-H and O···H distances, this was to be expected, because the corresponding bond parameters in ReaxFF were not re-optimized for the sake of compatibility with the silicate parameters in the Joshi *et al.* force field. For the O distances, the performance is slightly better, which is consistent with the fact that several Al-O parameters were re-optimized.

The new force field also improves upon the Joshi *et al.* force field in terms of valence angles, again with a somewhat better performance when no hydrogen atoms are involved. It is remarkable that the multimodal distributions of the Al-O-Al, O-Al-O and Al-O···H angles are reproduced well by the new force field, despite only having a single energy term for these angles in the ReaxFF force field. Also the improvements of the H-O···H and H···O···H angles, compared to the initial force field, are remarkable, because no corresponding valence angle of hydrogen bonding terms were reparameterized.

In line with the previous categories, also the RMSEs on the energies are significantly smaller with the new force field. Before discussing the adsorption energies, it should be noted that the category **SUR** was mainly introduced to improve the diversity of the training set. (It comprises reaction energies between bulk and slab models, normalized on the number of atoms.) A few data points in the same category can be derived from the validation structures and are included here for the sake of completeness. The error on these data points has also decreased compared to the initial force field, confirming the ability of Balanced Loss to account for underrepresented categories in the training set.

The new force field reduces the error on the adsorption energies (category **GSH**) by more than a factor 6 in the validation set. The improvements are also immediately clear in Figure 4.7, which shows the adsorption energy on γ-Al_2O_3 surfaces as a function of coverage. In most cases, the new force field predicts the correct trends in the adsorption energy, with some exceptions at low coverage in Figure 4.7(d) and Figure 4.7(f). For the (001) and the (110) $_b$ slabs in the validation set, it is unclear why these surfaces exhibit larger errors in adsorption energy at low coverage. The Joshi *et al.* force field incorrectly predicts water desorption for 12 out of 53 surface structures in the validation set, which hampers reliable energy predictions. With, the new force field, this problem is far less prevalent: Only one water molecule (out of six) from only one surface structure spontaneously desorbs. The new force field can also predict the magnitude of the adsorption energy, with an RMSE of 8.3 mol^{-1}, compared to a standard deviation of 16.4 mol^{-1} of the adsorption energies in the validation set.

For all categories of data discussed above, the errors on static calculations are very similar for the training and the validation sets, indicating that the significant improvements of the new ReaxFF force field generalize to structures not used for training. The errors on distances, valence angles and energy differences are also comparable to those reported for previous ReaxFF models. [102,118,216–219]

Because the training set only includes equilibrium geometries, it should be tested to what extent the new parameters can also reproduce non-equilibrium energies. To this end, a constant-temperature DFT Molecular Dynamics run was performed on structure `gamma_surf-110l_Al_06w` from the validation set, using the same level of theory as the training data. An elevated temperature of 1000 K stimulates the desorption of water, which is observed during the first 200 fs. Section S2 of the Supporting Information presents a detailed comparison of the DFT and ReaxFF energies computed for snapshots from this trajectory. In summary, the instantaneous DFT adsorption energy computed with Equation (4.3) as a function of time is reproduced qualitatively by the ReaxFF parameters obtained with Balanced Loss: The relative error of about 25 % over the first 200 fs is comparable to the RMSE on the training set for GSH category. Our new parameters also show a clear improvement compared to the ReaxFF energies obtained with the parameters of Joshi *et al.* The thermal energy fluctuations due to vibrations within the alumina slab are not well reproduced, which is expected, since no corresponding data was used for training.

Despite our improvements, it remains interesting to explore further refinements, e.g. to further reduce errors in adsorption energies or to improve the vibrational states of alumina. One avenue is to activate more parameters during the training that are now fixed for the sake of backward compatibility. Giving up backward compatibility would only be useful when extending the chemical space of the training set to aluminosilicates and water, such that all parameters of the Joshi *et al.* force field can be re-optimized. However, this would be a daunting enterprise, because the current training set size is already computationally demanding: A single CMA run in this work already took more than 24 hours. In addition to the increased cost of the training set, more active parameters also imply more local minima and a slower convergence of CMA-ES, further exacerbating the computational cost. This avenue is therefore only feasible when one can drastically speed up the training of ReaxFF parameters. It is encouraging that efficiency gains were reported in recent publications, e.g. by parallel optimization management, [136] or by machine learning surrogates of the loss function.[120,220,221] One may also reduce the dimensionality of the parameter space through sensitivity analysis to speed up the CMA runs.[123]

As a final check of the new force field, a Molecular Dynamics (MD) simulations is performed on $(110)_\ell$-γ-Al_2O_3 slab with a cross section of 6.0 nm^2, a thickness of 1.8 nm and surrounded by a vacuum layer of 5.7 nm wide. In the initial structure, the maximal number of water molecules is dissociatively adsorbed on both sides of the slab, such that the chemical formula is $Al_{504}O_{864}H_{216}$. The MD simulation follows a *NVT*-ensemble and uses a Nosé-Hoover thermostat [222] with a time step of 0.2 fs, a temperature of 500 K, a time-constant of 500 ps and a fixed periodic box size. Figure 4.8 shows the initial and final states of the MD trajectory, as well as the evolution of the kinetic energy, total energy (kinetic + potential) and the conserved quantity. The conserved quantity (green) exhibits a slow linear increase, which is acceptable for long ReaxFF MD simulations at a constant temperature. ReaxFF forces are imperfect due to numerical convergence of the variable charges and small discontinuities in the ReaxFF energy surface.

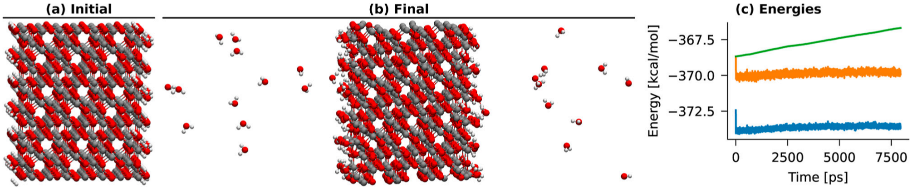

Figure 4.8: The initial (a) and final (b) states of a constant-temperature MD trajectory of a $(110)_\ell$-γ-Al_2O_3 with water initially adsorbed on the surface. Al=gray, O=red, H=white. See text for details. (c) The kinetic energy (blue), total energy (kinetic + potential, orange) and the conserved quantity (green).

Such small force errors are practically tiny random kicks on the nuclei, which slowly pump energy into the system, but this is easily compensated for by the thermostat and results in a slowly increasing conserved quantity. The 3D visualizations show that some of the water molecules desorb, as expected at a temperature of 500 K. This test shows that the new force field can also be used for MD simulations, even though it is only trained on optimized geometries. A complete study of water adsorption, with larger slabs, different alumina surfaces and temperatures, goes beyond the scope of this work. We expect that the new force field will make such simulations possible, at time and length scales that are infeasible for DFT methods.

4.5 Conclusion and Outlook

This work addresses the difficulty of assigning fitting weights (or their inverses, often called sigmas) in a conventional ReaxFF loss function. Balanced Loss is proposed as a new cost function as well as a workflow to reformulate the weight assignment in terms of more manageable concepts. One starts by classifying the training data into meaningful categories with a corresponding tolerance, which is the root-mean-square error (between ReaxFF predictions and reference data in that category) that one is willing to tolerate. When the error on one category exceeds the corresponding tolerance more than other categories, the Log-Sum-Exp form of Balanced Loss guarantees that this error will completely dominate the loss function, effectively forcing the optimizer to reduce this error first. If one or more of such dominating categories remain after the parameters converge, it is guaranteed that these errors cannot be reduced further at the expense of making larger errors in other categories. As a result, the parameter optimization also assesses whether ReaxFF can meet the expectations defined by the tolerances. If necessary, the expectations can be adjusted, followed by a new parameter optimization. The methodology is applied to a realistic and challenging reparameterization of ReaxFF for water adsorption on alumina surfaces. This not only results in a competitive force field, but also provides insight into the performance that can be expected from ReaxFF for each category of training data.

The new force field derived in this work is a refinement of the alumina parameters in the aluminosilicate force field of Joshi *et al.*[174,176] The training and validation data consisted of geometry and energy data from previous Density Functional Theory (DFT) studies of water adsorption on boehmite and γ-Al_2O_3 surfaces. While γ-Al_2O_3 is industrially the most relevant, boehmite structures were included in the training set to improve the data diversity. Parameter selection focused on maintaining backward compatibility with the ReaxFF model of Joshi *et al.* as much as possible, while still activating sufficient parameters to reproduce the training data. Covariance Matrix Adaptation (CMA) is used to minimize Balanced Loss as a function of the selected parameters. 40 independent CMA runs were performed to test the robustness of the optimized parameters. Of all these runs, the result with the lowest error on the training set is used for validation. Static calculations confirm that the optimized force field produces very similar errors on the γ-Al_2O_3 properties present in the training and validation sets. The force field can be used for MD simulations, and we expect it to be applicable to extensive simulations of water adsorption on alumina surfaces, at time and length scales inaccessible to DFT methods.

This study also revealed several challenges and new avenues for future method development. Obviously, it should be validated whether the Balanced Loss workflow is equally helpful for other ReaxFF parameterizations. Even beyond the scope of ReaxFF, Balanced Loss may facilitate optimization problems involving multiple (possibly competing) categories of training data. In addition, this study confirmed known pitfalls of ReaxFF parameterization and suggested new ones. Despite our careful and relatively small selection of parameters, the minimum of the loss function is still degenerate, suggesting that the number of active parameters can be reduced further with a sensitivity analysis. Finally, it was observed that parameters often converge to near-optimal values close to the bounds. A better understanding of this phenomenon may help to speed up the convergence to better parameters.

Data and Code availability A data set as made available at https://doi.org/10.5281/zenodo.10491516 comprising: the training and validation data, the results of the 240 CMA optimizations, and scripts used to select parameters, convert datasets and perform the training and validation.

5

Paper II: Node Energy Density Active Learning

> *"All models are wrong,*
> *but some are useful."*
> –George E. P. Box

Background and Motivation

While MLIPs have demonstrated remarkable success in providing quantum-mechanical accuracy at reduced computational cost compared to ab initio methods, their application to large-scale, long-timescale simulations, particularly those involving chemical reactions, often remains challenging. Compared to traditional reactive empirical force fields like ReaxFF, many MLIPs, especially large "universal" or "foundation" models, still present a significant computational barrier in terms of simulation speed and memory requirements. This limitation motivates the development of "dedicated" MLIPs, tailored to specific chemical systems, which can be designed with fewer parameters for greater computational efficiency. However, creating these dedicated models typically requires significant effort in generating bespoke training datasets.

The work presented in Paper II addresses this critical intersection of efficiency and accuracy in MLIP development. It introduces and validates a novel, computationally inexpensive active learning framework designed to automate the generation of diverse and efficient training sets for dedicated MLIPs with minimal human intervention. Instead of relying on computationally demanding ensemble methods or complex feature-space analyses, this approach utilizes Kernel Density Estimation (KDE) applied to predicted atomic node energies–a low-dimensional proxy for the local chemical environment–to effectively identify out-of-distribution configurations that are most valuable for improving the model.

The primary contribution of this work to the MLIP and broader scientific community is twofold. First, it provides a practical and efficient active learning strategy that significantly lowers the barrier for developing tailored, high-performance MLIPs for specific chemical problems. Second, by applying this strategy to the complex, reactive hydrated silicate ionic liquid system, it demonstrates that deliberately constructed, computationally efficient dedicated MLIPs (using the MACE architecture in this case) can achieve simulation speeds and memory usage significantly better than large universal MLIP models and can even surpass traditional reactive force fields like ReaxFF in terms of computational throughput. This challenges the common assumption that MLIPs are inherently slower than empirical potentials for large reactive simulations and positions efficient, actively learned dedicated MLIPs as a viable and potentially superior alternative for tackling complex chemical dynamics at scale.

Contributions of Authors

L.D.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Visualization.

T.V.: Conceptualization, Methodology, Software, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition, Corresponding Author.

J.V.: Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision, Funding acquisition.

B.C.: Conceptualization, Writing - Review & Editing, Supervision.

T.B.: Conceptualization, Writing - Review & Editing, Supervision.

Paper II

Scalable Reactive Molecular Dynamics with Machine Learning Potentials: Active Learning with Node Energy Density Estimation

Loïc Dumortier, Jelle Vekeman, Benoit Creton, Theodorus de Bruin and Toon Verstraelen

To Be Submitted in Journal of Chemical Theory and Computation, **2025**

Abstract

Machine learning potentials (MLPs) offer quantum-chemical accuracy with orders-of-magnitude greater efficiency than first-principle methods. However, their computational cost remains a barrier for computational chemists accustomed to empirical force fields, significantly limiting accessible time and length scales. Empirical force fields on the other hand provide great computational efficiency, while sacrificing accuracy. This work investigates how computational efficiency of MLPs can be maximized while retaining a physically meaningful and stable model. This enables studies at time and length scales on par with reactive empirical force fields like ReaxFF at an accuracy that approaches first-priniple methods. Since the efficient network architecture inherently means reduced flexibility in the model, a well-designed training set is essential. We therefore propose an active learning framework that efficiently generates a minimalistic, yet suitable and diverse training set with limited human intervention. By applying kernel density estimation to node energies, out-of-distribution structures are efficiently targeted, ensuring a well-balanced training set without the need for computationally expensive ensemble methods. The resulting model is benchmarked against ReaxFF and the universal MACE-MP-0 model, evaluating accuracy, efficiency, and reactivity. This approach is applied to a hydrated silicate ionic liquid (HSIL), a chemically complex liquid system with importance to the field of zeolite synthesis. The dedicated model achieves stable integration at a 1.2 fs timestep, making it up to 10x faster than MACE-MP-0 and 5x faster than ReaxFF for the same system, while allowing to load 10x larger system sizes in GPU memory compared to MACE-MP-0. While ReaxFF and MACE-MP-0 underestimate HSIL densities, the dedicated model accurately captures densities at diluted compositions, albeit slightly overestimating density at low water content. Large production runs show that the dedicated MLP is able to describe chemical reactivity in HSILs with low water content. This study demonstrates that dedicated MLPs can surpass empirical force fields in efficiency, making them a scalable and cost-effective alternative for reactive molecular dynamics simulation.

5.1 Introduction

Computational chemistry, more specifically molecular modeling, relies on a range of techniques, from first-principles methods such as Density Functional Theory (DFT) to classical force fields for Molecular Dynamics (MD). While DFT provides high accuracy, it is computationally expensive, limiting its applicability to small systems and short timescales. Empirical force fields, on the other hand, enable large-scale simulations, but lack the flexibility to (accurately) model bond breaking and formation. Machine Learning Potentials (MLPs) offer an alternative that bridge this gap, combining quantum-chemical accuracy with efficiencies orders of magnitude higher than quantum-chemical methods like DFT.[223–225]

However, the widespread adoption of MLPs remains limited by their computational cost in comparison to classical force fields.[226–228] Many MLPs, particularly universal models that are trained on a broad set of chemical environments, also referred as foundational models, are too expensive to run at the system sizes and timescales typically accessible with empirical force field MD simulations. While most MLPs are currently trained to reproduce DFT results to high accuracy, an important question is: how much accuracy can be sacrificed for computational efficiency while retaining physically meaningful and stable models? [229–231]

In contrast to universal MLPs, dedicated models focus on a specific subset of chemical space, require fewer parameters, and therefore lower computational resources during training and inference. Besides considerable improved system sizes and simulation speeds, in an era of rapidly growing energy demands from machine learning, this lower power consumption of dedicated MLPs cannot be overlooked. They will offer a cheaper and more sustainable alternative for targeted applications.[232]

This work departs from traditional MLP studies, which primarily focus on benchmarking MLPs against DFT for speed and accuracy. Instead, we compare MLPs with reactive force fields, exploring how far we can push simulation speed and system sizes while maintaining superior accuracy compared to those methods. As smaller models (i.e. models with fewer trainable parameters) are inherently faster than universal models, which contain orders of magnitude more parameters, we attempt to efficiently train such models. These smaller models lack the flexibility to accommodate a broad range of chemical environments like universal models and must therefore be trained for each specific system. As a result, their development time—both in terms of computational cost and manual effort—is a significant drawback, particularly when compared to readily available universal models.

Active learning is a well-established strategy for generating diverse training sets to develop dedicated MLPs with minimal human intervention. One of the most common approaches is ensemble-based uncertainty quantification, with query-by-committee being a widely used method.[233,234] During a query-by-committee training, ensembles of MLPs are trained on either the same training set or a subset at each iteration. The ensemble evaluates candidate configurations, and if the variance in predictions among models exceeds a threshold, the configuration is submitted for a reference calculation and added to the training set. While this method has proven to be effective in producing high-quality training sets, it introduces a significant computational overhead, as training an ensemble requires maintaining multiple MLPs —typically 5 to 10—at each iteration. [234]

An alternative approach avoids ensembles by leveraging feature-space sampling.[235–238] Here, out-of-distribution (OOD) metrics are applied to atomic descriptors or feature vectors. For example, one maximizes the high-dimensional volume of the training set in this feature space, or one adds new training data only if they contain feature vectors sufficiently far from those in the existing training set. As a result, new configurations are selected that maximize the diversity of the training set. While this method eliminates the overhead of training a committee of models, the design of a suitable OOD metric involves several ad hoc choices. For example, using a Euclidean distance in feature space assumes that all directions in this space are equally important. Since this assumption is often invalid, alternative distance metrics such as the Mahalanobis distance, which uses the data's covariance matrix to account for varying scales and correlations between features, can be employed.[239] While this provides a more statistically robust distance measure, reliable estimation of the required covariance matrix requires a significant number of feature vectors, especially in the high-dimensional spaces common in MLPs. Therefore, converging the covariance matrix itself requires large training sets, posing a challenge for this approach, especially during the early stages of active learning.

A simpler yet less targeted active learning strategy is uniform sampling from MD trajectories.[240] In this approach, one iterates through training an MLP, exploring the phase space by MD simulations, and selecting the last configuration from each trajectory for reference calculations. Additional filters or thresholds may be applied to reduce redundant configurations, but rare or short-lived events are unlikely to be captured. While this method is computationally inexpensive and minimizes training overhead, it lacks the adaptive refinement of more sophisticated selection strategies.

We propose an application of an innovative active learning procedure that iteratively refines dedicated MLPs with minimal computational overhead, automating both data selection and model improvement. Our approach applies Kernel Density Estimation (KDE) to detect outliers in the per-particle energy contributions predicted by the MLP before summation.[241] We refer to these contributions as "node energies"; they are summed within the model to obtain the total potential energy of a configuration. This terminology is intentionally used instead of "atomic energy" to emphasize that these values are intermediate outputs of the MLP architecture and do not necessarily represent physically interpretable, independent atomic properties. By analyzing the distribution of these node energies, we systematically identify OOD configurations to ensure a well-balanced training set without relying on computationally expensive ensemble-based uncertainty quantification.

To evaluate our proposed methodology, we apply it to hydrated silicate ionic liquids (HSILs)—a zeolite synthesis liquid. Composed of $\mathrm{SiO_4H_4}$, alkali hydroxides, and water, HSILs provide a homogeneous and tunable growth liquid for zeolites, contrasting with sol-gel systems where phase separation and gelation obscure mechanistic studies.[242,243] Our previous study, and to date the only computational study performed on these systems, has demonstrated the role of alkali cations in structuring HSILs, suggesting a cation-mediated assembly of silicate monomers.[94] However, these simulations relied on non-reactive classical force fields, preventing the study of condensation reactions and the formation of silicate oligomers. Traditionally, reactive force fields such as ReaxFF have been used to address the non-reactive limitations of classical force fields.[244] Unfortunately, the practical application of ReaxFF force fields is often limited by the need for extensive and labor-intensive reparameterization for each specific system. This process can be challenging, and achieving consistently high accuracy may be difficult even with the aid of advanced, dedicated optimization tools.[245–249] While developing dedicated MLPs also demands significant system-specific effort (primarily for data generation and training), the key distinction often lies in the methodology: MLP development leverages a well-established machine learning framework for systematic training, validation, and refinement, potentially offering clearer pathways to reliable, high-fidelity models compared to the intricacies associated with ReaxFF parameterization.

This study systematically assesses the accuracy, computational efficiency, and abillity to simulate reactive events of dedicated MLPs trained via active learning. These models are benchmarked against both universal MLPs and a ReaxFF force field, exploring trade-offs between accuracy, simulation speed, and system size scalability.

The remainder of this article is structured as follows: Section 5.2 introduces the active learning workflow, detailing the iterative model training and OOD sampling procedure. Section 5.3 outlines the computational methodologies, including DFT, ReaxFF, and MD simulations. Section 5.5 presents the application of our proposed OOD sampling criterion for active learning, benchmarking results, comparing model stability, timestep limitations, chemical composition evolution and computational speed. The final section discusses the broader implications of our findings and potential future directions for MLP development in reactive MD.

5.2 Node Energy Density Estimation for Out-Of-Distribution Sampling

5.2.1 Active Learning Algorithm

The development of an accurate and transferable MLP requires a diverse chemically "complete" training set that efficiently captures all relevant chemical environments, including those that may not be expected *a priori*. To achieve this, we propose an active learning algorithm that systematically refines the training set autonomously by identifying OOD configurations, performing DFT reference calculations, and retraining the potential on the augmented training set. The active learning process is divided into two main stages, as illustrated in Figure 5.1: (i) Initial training set generation (iteration

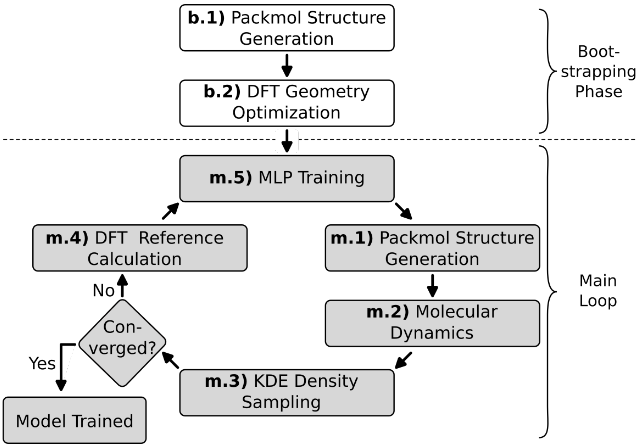

Figure 5.1: Overview of Active Learning Procedure, as explained in Section 5.2.1. The system is initialized by a configuration generation followed by a short DFT energy minimization. These cell relaxation data are used to train a coarse MLP that can handle short dynamics, that is then used to iteratively create OOD samples and retrain.

0), and (ii) Active learning main loop (iteration 1, 2, 3, ...), as detailed in the following sections.

5.2.1.1 Bootstrapping phase (iteration 0)

To construct a representative initial training set for the MLP, we first generate a diverse set of 20 random molecular HSIL configurations, each containing approximately 300 atoms, with different concentrations of SiO_4H_4, alkali-hydroxides and water. The investigated ranges of these three components of HSILs are explained in greater detail in Section 5.3.4. Each initial configuration is generated as follows: **b.1)** The molecules are placed at random positions in a periodic simulation cell using Packmol, [250] and the initial volume of the cell was estimated from the van der Waals volumes of the inidiviual molecules. **b.2)** The obtained configuration undergoes a short energy minimization using DFT, explained in Section 5.3.2 to refine the atomic positions and ensure that the training set captures a diverse range of bond angles and interatomic distances.

Since the structural snapshots from the energy minimizations are highly correlated, and the purpose of this initialization is solely training a preliminary MLP that can handle short MD runs, we limit the optimization to five ionic steps using the FIRE algorithm. [251] The energy minimizations are performed under periodic boundary conditions with isotropic cell relaxation, allowing simultaneous optimization of both atomic positions and unit cell vectors. The energies and forces of both the initial configurations and slightly relaxed configurations are included in the training set. Then, using this initial training set containing cell relaxation data, a coarse MLP is trained after which the main loop of the active learning algorithm can start.

5.2.1.2 Active learning main loop

The active learning process refines the MLP in an iterative manner, improving its accuracy and robustness after each cycle. One iteration in the active learning main loop consists of five steps:

m.1 **Configuration Generation for MD.** At the start of each iteration, new system configurations are generated using Packmol. As will be explained in the next step, the number of generated configurations decreases in later iterations to focus computational resources on MD simulations with longer timescales and higher temperatures as the model improves.

m.2 **MD Simulations with the MLP.** MD simulations are performed using the configurations from **m.1** and the MLP trained in the previous iteration. The number, duration, and temperature of MD runs are adjusted dynamically: In early iterations, many short MD's are run at a low temperature. Later, the simulation time and temperature are extended exponentially and linearly, respectively, in each iteration. This enables a broad exploration of uncorrelated, local atomic configurations, some of which will be at the edge of the MLP's capabilities, including structural conformations that were inaccessible in previous iterations. Prior to the actual selection of configuration for training data, 25 equally spaced snapshots from each MD run are stored on disk.

m.3 **Selection of OOD Configurations.** To identify which configurations should be included in the training set without direct human intervention in the process, we use a KDE selection algorithm based on node energies predicted by the MLP. Configurations that contain node energies that are not well represented in the training set are then selected for DFT evaluation. This selection criterion is explained in detail in Section 5.2.2. To avoid sampling too many configurations during the initial iterations, where many independent, short MD runs are started, a cap of $\max(30, n)$ allowed configurations for DFT evaluation is set, with n being the number of MD runs started at that active learning iteration.

m.4 **Single-Point QM Calculations and Dataset Expansion.** The selected configurations undergo single-point energy and force calculations using DFT. These DFT-calculated reference values are then added to the training dataset.

m.5 **Training of MLP.** A new MLP is trained starting from randomized weights, using the expanded training dataset, which is now enriched with newly selected DFT-evaluated configurations that were previously underrepresented.

To make sure the entire chemical space is covered, the thermodynamic conditions of the MD simulations (temperature, pressure, ...) in the active learning loop are set higher than those expected for production runs with the final MLP. Once the selection criterion fails to identify any new configurations, even under these elevated conditions, the algorithm is considered converged, and the active learning loop is terminated. At this point, the dedicated MLP is fully parameterized and ready for inference in production-scale MD simulations.

5.2.2 Selection of OOD Configurations

To identify atomic configurations that fall outside the distribution of the training dataset, we employ a selection criterion that is not based on an ensemble or on feature vectors. Instead, we select OOD configurations by sampling node energies with low probabilites compared to the training set, as computed with the current MLP. Node energies can be understood as the $1 \times N$ vector, where N is the number of atoms, from the final layers of the neural network. These node energies are summed to obtain the total energy of a configuration, meaning that this vector represents the contribution of every atom to the total energy. In essence, this vector is a low-dimensional transformed representation of the feature vectors of all atoms and closely relates to the predicted energy and forces.

Although node energies extracted from the model are not chemically interpretable, their relative distribution is informative. An atom with an energy far from the training set distribution suggests that its local atomic environment is underrepresented or missing and makes it a prime candidate for inclusion in the training set. Since MLPs, like all neural networks, struggle with extrapolation, an unseen atomic environment will likely result in node energy predictions that deviate significantly from the training set distribution.[252] In rare cases, an extrapolated node energy may fall within a high-density part of the distribution, making it “invisible” to density-based selection algorithm yet still underrepresented. However, this does not pose a problem, as it implies the MLP is not truly in an extrapolative regime, even if the local environment is not being represented.

In principle, any algorithm capable of fitting a density function to the node energies in the training set could be used. However, due to the “black box” nature of MLPs and the complexity of the chemical space they must span, the node energy probability density is a priori unknown. Therefore, a non-parametric method is required to estimate this distribution without assuming a specific functional form. In this work, we employ KDE to approximate the distribution of the node energies.[241,253]

Using the non-parametric KDE algorithm, a kernel is applied to every node energy in the training set to approximate the density function f. We use the Gaussian kernel $K(x)$:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{5.1}$$

The bandwidth h is determined using Scott’s rule [254]:

$$h = 1.06\sigma n^{-\frac{1}{5}} \tag{5.2}$$

where σ is the standard deviation of the node energies and n is the number of node energies or atoms.

The KDE value for a given node energy E_a is then computed as:

$$f(E_a) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{E_a - E_{a,i}}{h}\right) \tag{5.3}$$

To identify configurations that are OOD relative to the training dataset, we rank potential configurations based on their KDE values. Specifically, for each system configuration, we determine the minimum KDE density among all atoms:

$$f_{\min} = \min_{i=1\ldots N} \{f(E_{a,i})\} \tag{5.4}$$

where N is the number of atoms in the system configuration. If $f_{\min}$ is then lower than the k_{sel} lowest values in the current training set (with k_{sel} being a manually set threshold value), the configuration is selected for reference calculation and subsequently added to the training set. Lowering k_{sel} makes the criteria less strict and will therefore mark fewer configurations as OOD. Increasing k_{sel} then makes the algorithm mark more configurations as OOD. An additional filter is applied to discard configurations with maximum atomic forces greater than 10 $\text{eV}\text{Å}^{-1}$ when computed with DFT, thereby preventing the inclusion of potentially unphysical configurations often encountered during early MD equilibration.

We apply the proposed OOD detection procedure separately for each atomic species in the training set to estimate element-specific probability densities using a threshold value $k_{\text{sel}} = 5$. In practice, this means that if any atom in a trial configuration has a KDE value that is lower than the 5-th lowest KDE value within its element type in the current training set, the entire configuration is selected for reference calculation and incorporated into the training set of the next iteration. Once the next-generation MLP is trained on the expanded dataset, previously OOD atoms will be assigned new energies that are often no longer OOD.

A key advantage of KDE-based sampling is that it effectively explores the potential energy surface using a single model, eliminating the need for ensemble-based uncertainty quantification. Since the probability density of the training set is computed only once per iteration, the additional computational cost is negligible compared to the cost of training an MLP or performing reference calculations.

5.3 Technical Details

5.3.1 Machine Learning Potentials

All the MLPs in this work are constructed with the Multi Atomic Cluster Expansion (MACE), using mace-torch version 0.3.6.[64,223] MACE is a highly configurable framework that encompasses a range of architectures found in other MLPs.[223] It was shown to achieve high accuracy while maintaining computational efficiency.[64,225] Unlike ReaxFF and other traditional empirical force fields, MACE has no physically motivated energy terms. Instead, it learns interatomic interactions by training weights in a neural network. Because the number of weights is orders of magnitudes larger than the number of parameters in an empirical force field, MACE can typically reproduce DFT reference data with a much higher accuracy.

MACE employs a message passing architecture, also found in other MLPs.[224,255,256] Each atom in MACE is treated as a node in a graph whose state is described by a tuple:

$$\sigma_i^{(t)} = \left(\boldsymbol{r}_i, z_i, \boldsymbol{h}_i^{(t)}\right) \tag{5.5}$$

where $\boldsymbol{r}_i$ is the position of atom i, z_i its atomic number, and $\boldsymbol{h}_i^{(t)}$ its feature vector. The superscript t refers to the iteration in the message-passing scheme. The initial feature vector $\boldsymbol{h}_i^{(0)}$ consists of a set of element-specific learnable weights. Each iteration of the message-passing scheme updates the feature vector $\boldsymbol{h}_i^{(t)} \to \boldsymbol{h}_i^{(t+1)}$ by aggregating information from a set of nearby atoms within a cutoff radius r_{max}) through a so-called message function M_t:

$$\boldsymbol{m}_i^{(t)} = M_t\left(\sigma_i^{(t)}, \left\{\sigma_j^{(t)} \middle| \parallel \vec{r}_i - \vec{r}_j \parallel < r_{\text{max}}\right\}\right) \tag{5.6}$$

The message function must be permutationally invariant to the order with which neighboring atoms j are included in the calculation. The message $\boldsymbol{m}_i^{(t)}$ is then used to update the feature vector $\boldsymbol{h}_i^{(t)}$ of atom i:

$$\boldsymbol{h}_i^{(t+1)} = U_t\left(\sigma_i^{(t)}, \boldsymbol{m}_i^{(t)}\right) \tag{5.7}$$

where U_t is an update function that combines the current state of atom i with the message received from its neighbors. After T iterations through the message-passing scheme, the energy of the node is computed with a readout function R_t:

$$E_i = \sum_{t=1}^{T} R_t\left(\boldsymbol{h}_i^{(t)}\right) \tag{5.8}$$

The mathematical forms of the functions M_t, U_t, and R_t are specific to the architecture of the MLP: They can be simple pooling functions, linear transformations, tensor products, or dense neural networks, some of which contain learnable weights. A full description of these functions in MACE goes beyond the scope of this work, and we refer to the original paper for the details.[223] Here, we will provide a brief overview of the design principles behind them:

- The messages are constructed with a hierarchical body-order expansion of the atomic environment:

$$\boldsymbol{m}_i^{(t)} = \sum_j u_1\left(\sigma_i^{(t)}; \sigma_j^{(t)}\right) + \sum_{j_1,j_2} u_2\left(\sigma_i^{(t)}; \sigma_{j_1}^{(t)}, \sigma_{j_2}^{(t)}\right) \\ + \sum_{j_1,j_2,j_3} u_3\left(\sigma_i^{(t)}; \sigma_{j_1}^{(t)}, \sigma_{j_2}^{(t)}, \sigma_{j_3}^{(t)}\right) + \ldots \tag{5.9}$$

The functions u_n, for $n > 1$, are invariant to the order of neighboring atoms to ensure permutational invariance. The multi-body expansion is truncated after ν terms, with ν the correlation order, a hyperparameter of MACE.

- The functions u_n use inputs of the following form:

$$f_k(r_{ij}) Y_{\ell m}(\hat{\boldsymbol{r}}_{ij}) \tag{5.10}$$

where $f_k(r_{ij})$ are radial Bessel functions that depend on the interatomic distance and $Y_{\ell m}(\hat{\boldsymbol{r}}_{ij})$ are spherical harmonics that depend only on the unit vector $\hat{\boldsymbol{r}}_{ij} = \frac{\boldsymbol{r}_{ij}}{r_{ij}}$. This choice ensures that only relative positions are used as input, making the network translationally invariant. The maximum degree of the spherical harmonics, L, is also an important hyperparameter of MACE.

- The functions M_t and U_t are constructed to make the feature vectors $\boldsymbol{h}_i^{(t)}$ equivariant (or invariant if $L = 0$) to rotations of the atomic positions $\boldsymbol{r}_i$. This implies that a global rotation Q of the atomic positions $\boldsymbol{r}_i$ results in a well-defined unitary transformation $D(Q)$ of the feature vectors:

$$\boldsymbol{h}_i^{(t)}(Q \cdot (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)) = D(Q) \cdot \boldsymbol{h}_i^{(t)}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) \tag{5.11}$$

Earlier generations of message-passing networks were invariant, because they only used interatomic distances as inputs, making them completely insensitive to global rotations.[255] A disadvantage of an invariant network is that it discards relevant 3D information (e.g., valence angles) from the beginning and is then forced to infer such patterns from interatomic distances. Equivariant networks overcome this limitation in a more efficient way than also including all possible valence angles, dihedral angles, etc., as inputs. MACE is invariant for $L = 0$ because Y_{00} is constant and only the radial dependence of the Bessel function remains in the message function. It becomes equivariant for $L > 0$.

- The size of the feature vectors $\boldsymbol{h}_i^{(t)}$ is proportional to the number of channels, N_{channels}). Components of the feature vector are grouped per spherical harmonic degree ℓ, representing geometric information as scalars

for $\ell = 0$, vectors for $\ell = 1$, and rank-2 symmetric traceless tensors for $\ell = 2$, etc. The $N_{\text{channels}})$ hyperparameter defines the multiplicity: how many independent features are learned for each combination of indexes (ℓ, m) of the spherical harmonics. For example, $N_{\text{channels}}) = 8$ and $L = 2$ means a feature vector has $8 \times (1 + 3 + 5) = 72$ components. The initial feature vector, $\boldsymbol{h}_i^{(0)}$, is different because it consists of geometry-independent learnable weights. It therefore only contains N_{channels} components for $\ell = 0$.
- The readout function R_t only employs invariant information from the feature vector $\boldsymbol{h}_i^{(t)}$, because the energy of the system must be invariant to global rotations.

5.3.1.1 Dedicated Model

We first discuss our dedicated parametrization of MACE with settings that emphasize speed over accuracy. This choice is motivated by our interest in large-scale MD simulations, ideally at the same speed as a ReaxFF force field. A benchmark that motivates the settings is presented in Section 5.5.2. The MACE MLP was trained using two interaction layers and eight channels per interaction block, with a maximum angular momentum order $L = 2$. The total number of trainable parameters in the model was 32280. The model incorporated two-body and higher-order correlations, with a cutoff radius $r_{\text{max}} = 4$ Å to define atomic neighborhoods. For the computational efficiency benchmark, an additional model with a cutoff radius of 5 Å is also trained.

Training of the dedicated models was conducted for 200 epochs, with a batch size of 4 for both training and validation datasets. For each training, a random 10% of the training set was used as validation set. Stochastic Weight Averaging (SWA), which changes the force-energy weight ratio from 100 to 0.1, was initiated after 140 epochs to improve model generalization.[257] Model performance was evaluated every 20 epochs to monitor convergence and accuracy. The model parameters were optimized using the Adam optimizer with a learning rate of 0.01 initially and then 0.001 once SWA is enabled. Training reproducibility was ensured by fixing the seed.

5.3.1.2 Universal Model

We have used the the smallest universal model of MACE-MP-0, proposed by Batatia *et al.*[258] The model is trained for 89 elements on approximately 1.6 million bulk crystals. The model uses 3 interaction layers and has 128 channels per interaction block. The radial cutoff is set at 6 Å. The model uses

Parameter	**Dedicated**	**Universal**
Cutoff radius ($r_{\max}$) $[\text{Å}]$	4.0	6.0
Number of channels (N_{channel})	8	128
Interaction layers (T)	2	3
Maximum spherical harmonic order (L)	2	0
Correlation order (ν)	2	3
Trainable parameters	32,280	3,847,696

Table 5.1: Comparison of model parameters for the dedicated and universal MLP.

a maximum angular momentum order $L = 0$ and a correlation order of 3. A summary of the settings used in the dedicated model and the universal model can be found in Table 5.1

5.3.1.3 MD using MACE and OpenMM

MD simulations were performed using the OpenMM framework in combination with the MACE MLP. [64,223] The system topology was constructed from the provided atomic coordinates in the XYZ format, and periodic boundary conditions were applied. Atomic interactions were modeled using the MACE potential, with energy and distance units converted to kJ mol^{-}1 and nm, respectively.

Density simulations and time step stability tests were performed in the Isobaric-Isothermal Ensemble (NPT), while reaction detection simulations were performed in the canonical ensemble (NVT). Production runs employed a Langevin integrator with a time step of 1.2 fs for the dedicated model and 0.6 fs for universal MACE, as motivated in Section 5.5.3. The friction coefficient is set to 1 ps.[259,260] Because HSILs are relatively stable, the temperature was set at 425 K for both NVT and NPT simulations to promote reactive events.[80] The pressure in the NPT simulations is maintained at 5 bar using a Monte Carlo barostat to avoid phase transition of samples with high water content. These settings were chosen to represent experimental conditions using heated autoclaves. NPT simulations were conducted at 1 bar and 295 K for density comparison with experimental results. Energy minimization was performed before the production run to relax the initial configuration.

5.3.2 Periodic DFT

Reference energies and forces were computed with periodic DFT calculations using the Vienna Ab initio Simulation Package (VASP) version 6.4. [261,262] Valence interactions were described using the Projector Augmented Wave (PAW) method.[263,264] The Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional was used.[14] To account for van der Waals (vdW) dispersion interactions, the DFT-D3 correction was applied (IVDW = 12). [265] The plane wave basis set employed a cutoff energy of 500 eV, with a k-point sampling limited to the Gamma point due to the system size. Gaussian smearing with a width of 0.2 eV was used. All calculations were conducted with a convergence criterion of 1×10^{-3} eV for the self-consistent field (SCF) cycle. The maximum number of SCF iterations was restricted to 50, while a minimum of 5 iterations was enforced to ensure stability. The parallelization scheme utilized 16 cores per calculation.

5.3.3 ReaxFF Force Field

ReaxFF was introduced in 2000 by Van duin *et al.* [266–268] Unlike the majority of other empirical force fields, it effectively describes bond breaking and formation. The ReaxFF potential energy E_{System} is described as a sum:

$$\begin{aligned} E_{\text{System}} = E_{\text{Bond}} + E_{\text{Over}} + E_{\text{Under}} + E_{\text{Val}} + E_{\text{Tors}} \\ + E_{\text{vdW}} + E_{\text{Charge}} + E_{\text{Specific}} \end{aligned} \tag{5.12}$$

where E_{Bond} is the energy of an atom pair, E_{Over} and E_{Under} are correction terms for over- and under-coordination, respectively. E_{Val} is the valence angle energy and E_{Tors} is the torsional angle energy. Non-covalent interactions are modeled with E_{Charge} and E_{vdW}, the charge and the vdW interactions, respectively. The atomic charges are variable and account for polarization and Coulomb forces.[182,183]

The ReaxFF calculations in this work were performed using the open-source implementation in the LAMMPS MD package. As retraining and validating a new ReaxFF force field is outside the scope of this work, the simulations employed the ReaxFF force field parameterized by Hahn *et al.* for systems containing hydrogen, sodium, oxygen, and silicon.[269,270] This force field was parameterized for $\text{NaSiO}_x/\text{H}_2\text{O}$ systems and should be applicable to HSILs.

The system was equilibrated and maintained at 300 K using an NPT ensemble using the Nose-Hoover thermostat.[271] Pressure is controlled isotropically at 1 bar according to the equations of motions proposed by Martyna *et al.*[272,273] The relaxation times for temperature and pressure were set to 99.0 fs and 1000.0 fs, respectively.

To evaluate the scalability of the simulations, benchmarks were conducted using 16, 32, and 64 processor cores. These benchmarks were performed on systems with five different concentrations of HSILs to calculate density and assess the performance of ReaxFF under varying system sizes and computational loads.

5.3.4 HSIL Ternary Phase Diagram and Sampled Compositions

We systematically sampled compositions across the relevant range in the ternary phase diagram of SiO_4H_4, NaOH, and H_2O, as shown in Figure 5.2.

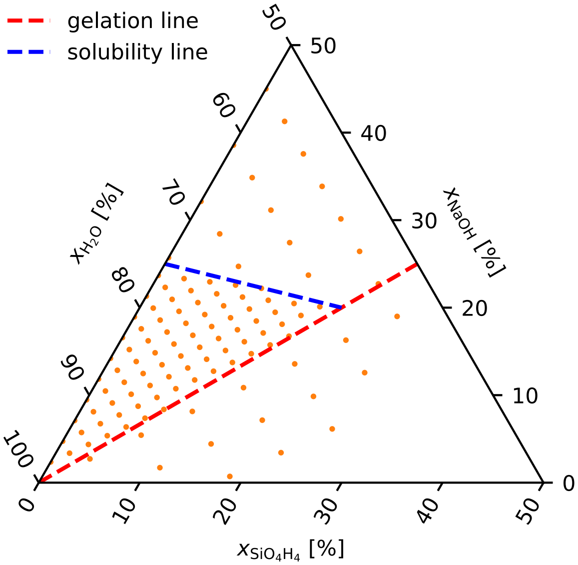

Figure 5.2: Ternary phase diagram of the SiO_4H_4-NaOH-H_2O system. The red dashed line represents the gelation limit, beyond which silicate networks form viscous gels. The blue dashed line marks the solubility limit, where components become insoluble. The orange dots indicate the compositions that were sampled to build initial configurations for MD simulations. At all times, at least 75% of the samples are within the HSIL region. The sampled points densely cover the HSIL region and extend beyond its boundaries to probe phase stability and reactivity near the solubility and gelation thresholds. Figure created with mpltern.[274]

The HSIL region is determined by the NaOH solubility line and the gelation line, as found in experiments and NMR characterization. [80,88] The sampled points are depicted as orange markers and encompass the core HSIL region but also exceed both the solubility and gelation limits. Within the HSIL range, we ensured dense coverage to capture subtle trends in oligomerization, reactivity, and molecular structure. To diversify the training set some compositions beyond the HSIL domain are also included. At time scales accessible in MD simulations these states are expected to remain homogeneous without undergoing precipitation or gelation.

5.4 Graph-Based Oligomerization Tracking in Molecular Trajectories

To analyze the evolution of molecular species in reactive MD simulations, we employ a graph-based approach where atoms are treated as nodes and bonds as edges using NetworkX. [275] This methodology enables the detection of transient molecular species by tracking their connectivity over time.

Each frame of the MD trajectory is converted into an undirected graph $G = (V, E)$, where the nodes V correspond to atoms, each labeled by its chemical element. The edges E represent bonds, determined by interatomic distances. A bond between two atoms i and j is defined if:

$$d_{ij} < d_{\text{form}} \tag{5.13}$$

and broken when

$$d_{ij} > d_{\text{break}} \tag{5.14}$$

where d_{ij} is the interatomic distance, and d_{form} and d_{break} are element-pair-specific bonding and breaking thresholds, respectively. This approach introduces a hysteresis: for distances d_{ij} such that $d_{\text{form}} \leq d_{ij} \leq d_{\text{break}}$, a bond exists if and only if it had previously formed (by satisfying $d_{ij} < d_{\text{form}}$) and has not subsequently broken (by exceeding d_{break} since formation). If a bond has not formed, it remains absent in this intermediate distance range. This prevents rapid, artificial bond fluctuations due to minor distance variations around a single threshold.

Since our primary goal in this study is to accurately track Si-O connectivity for analyzing silicate oligomerization, the specific threshold values, particularly d_{break}, are chosen based on the characteristics of the Si-O interaction, although the general methodology described can be applied to other element pairs. Highly dynamic bonds, such as O-H involved in water formation or proton transfer, are deliberately excluded from the species definition to avoid obscuring the evolution of the more stable Si-O bonds. The criteria for bond formation and breaking are determined from the radial distribution function (RDF): For a pair of chemical elements, the RDF is computed over a representative segment of the trajectory. The first peak of the RDF is then considered to be the bonded species and is fitted with a Gaussian function $g(x; \mu, \sigma)$ using the full width at half height to determine the characteristic bond length distribution. The following thresholds are then defined as:

$$\begin{gathered} d_{\text{form}} = \mu + 4\sigma \\ d_{\text{break}} = 3 \end{gathered} \tag{5.15}$$

where μ and σ are the mean and standard deviation of the fitted Gaussian function, respectively. The value of 3 Å for the breaking threshold is fixed manually, as the RDF's between Si-O pairs showed a clear distinction between the first and second peak at that distance, as shown in Figure 5.7.

Once the molecular graph is constructed, species are identified by extracting connected components of the graph. Species detection is repeated at each trajectory snapshot to record the time evolution of species populations in the simulation. To categorize silicons based on their connectivity in oligomerization processes, we classify each Si atom according to the number of Si-O-Si bridges it has with neighbouring Si atoms. A silicon atom is then assigned a connectivity class: Q_0 indicates Si-O-Si bridges, Q_1 one Si-O-Si bridge, Q_2 two Si-O-Si bridges, and so on. These classifications are stored for each snapshot as well as their temporal evolution to track structural changes over time.

5.4.1 Workflow Management and Cluster Parallellization

Parman is an in-house developed, open-source workflow manager designed to extend Python's *concurrent.futures* with additional features that facilitate efficient and transparent workflow execution. It introduces WaitFuture and ScheduledFuture to enable managing task dependencies without blocking execution. Parman also provides Runner classes, which function similarly to executors but allow closures with embedded metadata, enabling features such as dry-run validation and automatic dependency resolution. Unlike traditional workflow managers, Parman integrates seamlessly into Python scripts and therefore allows efficient job scheduling with minimal overhead. Its lightweight design without third-party dependencies makes it a practical tool for managing complex computational workflows with ease.

5.4.2 Computational Infrastructure

All calculations were performed using the Flemish Tier-1 HPC infrastructure, *Hortense*. DFT reference calculations were executed on AMD Epyc 7763 CPUs, utilizing 16 cores per calculation. ReaxFF calculations were executed on AMD Epyc 7763 CPUs, utilizing 16, 32 or 64 (*i.e.* a full node) cores per calculation. MLP calculations (training and inference) were conducted on NVIDIA A100 GPUs with either 40 GB or 80 GB of memory. For performance benchmarking, only NVIDIA A100 GPUs with 80 GB of memory were used.

5.5 Results and Discussion

5.5.1 Active Learning Loop and Node Energy Density as Selection Criterion

The active learning procedure was run for seven iterations before the loop terminated due to convergence. In the first loop, MD simulations were run for 100 steps, corresponding to 120 fs at a temperature of 0 K. As described in Section 5.3, both the number of time steps and the temperature were increased across successive iterations—exponentially for the former and linearly for the latter. Consequently, in the final loop, MD simulations extended up to 40,000 time steps (48 ps) at a temperature of 540 K, well above experimental synthesis conditions to promote sampling of high-energy configurations. An overview of the active learning evolution for both the model with a $r_{\max}$ cutoff of 4 Å and 5 Å is shown in Table 5.2.

Both models ultimately developed training sets of approximately 250 configurations. Over half of these configurations originated from the initial bootstrap loop, which included cell relaxation data. The first active learning loop (100 steps, 0 K) then contributed another significant batch of configurations, underscoring the importance of MD data for capturing configurations not accessible via cell relaxation alone. While later loops added fewer configurations overall, they remained crucial. Notably, these subsequent loops selected configurations from simulations at higher temperatures and longer timescales, indicating the capture of activated events containing essential

	Simulations	steps	T [K]	Snapshots	OOD		Training Set	
Cutoff					4	5	4	5
Bootstrap	24	5	/	144	144	144	136	140
Loop 1	112	100	0	2800	80	112	207	244
Loop 2	112	271	90	2800	4	7	208	246
Loop 3	112	738	180	2800	37	5	245	247
Loop 4	51	2008	270	1275	1	10	245	256
Loop 5	20	5459	360	500	2	2	245	257
Loop 6	7	14841	450	175	0	0	245	257
Loop 7	2	40342	540	50	0	0	245	257

Table 5.2: Summary of active learning simulation parameters and results per loop. “Simulations” indicates the quantity of simulations initiated at the beginning of each loop. “Steps” refers to the number of ionic steps performed within each simulation during that loop. “T” denotes the target temperature for the simulations in that loop. “Snapshots” represents the combined number of configurations generated across all simulations within a loop. “OOD” shows the number of configurations selected (e.g., identified as OOD) for potential inclusion in the training set. “Training Set” indicates the total size of the training set after the update in that loop. For “Snapshots” and “Training Set”, values are presented separately for the dedicated potential with a cutoff of 4 Å and 5 Å.

information for model accuracy. It is important to note, however, that not every configuration identified as OOD is ultimately included in the training set due to the exclusion of structures exhibiting a maximum atomic force greater than 10 $\mathrm{eV\AA^{-1}}$ when computed using DFT. For both workflows, the active learning process terminated after 7 loops; no additional samples were selected in the final two loops (loops 6 and 7), during which simulation temperatures exceeded 500 K, a point well above experimental conditions. This lack of new selections at high temperatures does not necessarily preclude the occurrence of new reactive, coordinated events, but may suggest that any such events, if present, are composed of local atomic interactions already learned by the model, as the selection method inherently focuses on these local environments.

We continue to analyze how the model evolves during active learning by evaluating how it perceives the same dataset at different stages. Figure 5.3 shows the node energy distributions per element (H, Na, O, Si) for the final training set (loop 7), evaluated using the model from loop 0 and 7. The distributions differ significantly between the two evaluations.

In the initial model, the node energy distributions are narrower, indicating that the model assigns a more limited range of energies to atomic species. In contrast, the final model assigns broader distributions, capturing a wider variety of chemical environments. This suggests that the initial model oversimplifies the atomic contributions to the total energy—an expected outcome, given that it was only trained on cell relaxation data, with little to no exposure to reactive events or thermally induced distortions.

We now demonstrate the benefits of using node energy density as a selection criterion within the active learning algorithm. In the initial iterations, MD trajectories are short and conducted at low temperatures, where no significant reactivity is expected. As a result, atomic configurations are highly correlated, and KDE-based selection typically identifies the last frame of each trajectory as (the most) OOD due to the limited diversity. At this early stage, different selection criteria—such as uniform or KDE-based sampling—are likely to yield similar results.

However, the strength of our method becomes evident in later iterations, as both the temperature and trajectory length increase. At iteration 4, for example, 20 MD simulations were launched. Yet, the KDE-based selection identified only two configurations as truly OOD—both at high silicate concentration. This indicates that most trajectories no longer contribute new information to the training set, and blindly including all of them would introduce redundancy and needless computational cost for reference calculations and training.

A comparison between the KDE-selected samples and the naive last-frame selection at iteration 4 reveals key differences, as shown in Figure 5.4. The node energy distributions for oxygen and hydrogen appear similar in both cases, though the last-frame method results in higher counts due to the larger number of included configurations. However, this added volume offers no new insight. Oxygen and hydrogen exhibit limited reactive pathways in

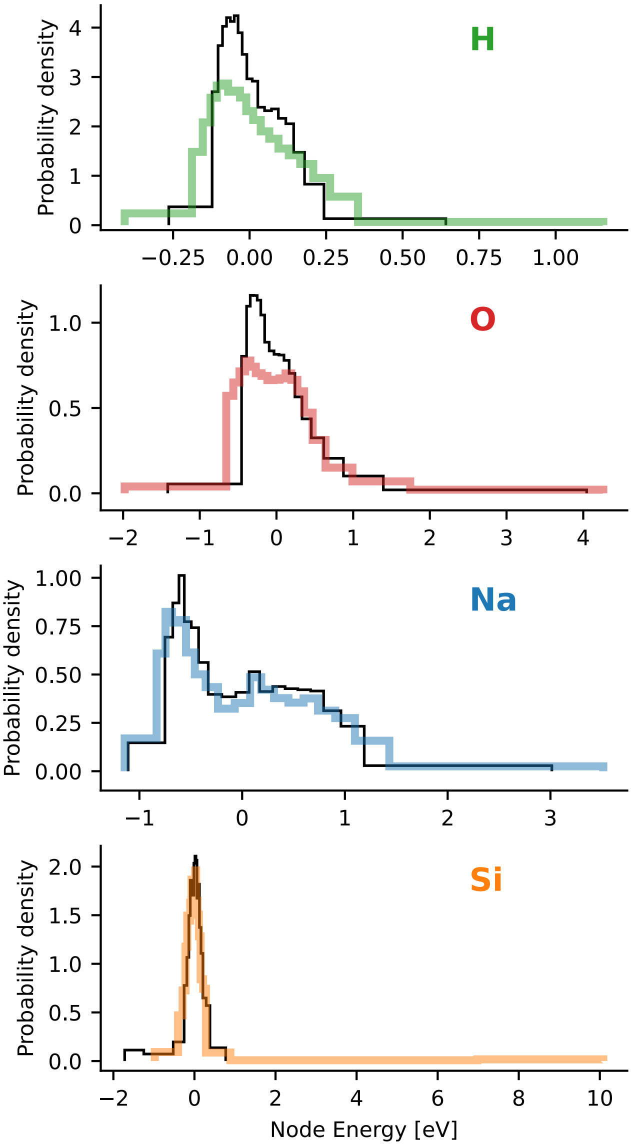

Figure 5.3: Histograms of node energies for the training set at loop 7, evaluated using the model from loop 0 (black lines) and loop 7 (colored lines). The median was subtracted from each set of energies before plotting the histogram.

HSILs and are learned early in the training process, making them unlikely drivers for selection. The distinction becomes more apparent for sodium and silicon. Even though only two samples are selected by KDE-based sampling, the range of node energies covered is comparable to that of the naive approach.

This illustrates a clear advantage of node energy density sampling as a data reduction strategy, yielding high-value configurations with minimal computational cost. The method efficiently identifies OOD samples and offers insight into dataset diversity, making it a practical and scalable component of the active learning workflow.

5.5.2 Model Benchmark

To assess the effectiveness of the active learning procedure, we evaluate the root mean squared error (RMSE) of residual energy and forces, along with the maximum force error, across different stages of model training and for the universal model. Additionally, the MD configurations from the training set for the dedicated model with a 5 Å cutoff, used exclusively for the computational benchmarking in Section 5.5.4, is treated as an independent validation set. This training set remains entirely unseen by both the universal and dedicated models discussed below. The validation set contains a total of 117 MD configurations, as shown previously in Table 5.2.

Since the universal model employs a different energy normalization in its training set, directly comparing absolute energy predictions would be unfair. To ensure a consistent comparison, we instead analyze residual energies, which represent the portion of the total energy not accounted for by a sum of constant atomic reference energies. Residual energies are computed by first determining atomic reference energies via least squares regression. The residual energy is then defined as the difference between the total energy and the sum of atomic reference energies. This approach ensures that the models are compared using chemically relevant energies. The results, presented in Table 5.3 - Table 5.5, highlight the progressive improvement in accuracy as additional OOD configurations are incorporated into the training set.

The RMSE values in Table 5.3 and Table 5.4 show a clear improvement across active learning iterations. Initially, the model trained only on cell relaxation data (loop 0) exhibits significantly higher errors when evaluated on MD-generated datasets. However, as new training points are added, the errors systematically decrease and therefore demonstrate the effectiveness of active learning in refining the model. By loop 7, the RMSE for both energy

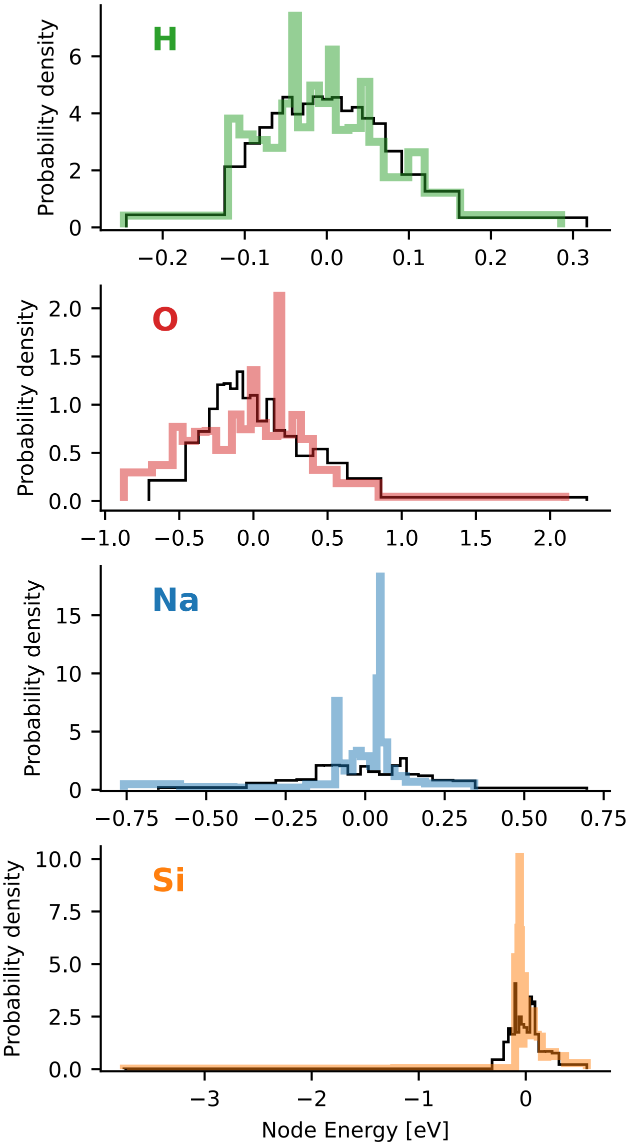

Figure 5.4: Node energy distributions for atomic species at iteration 4, evaluated using the model from the same iteration, for either the last frame of each trajectory (black lines) or KDE-based selection (colored lines). The median was subtracted from each set of energies before plotting the histogram.

and forces stabilizes, suggesting that the training set adequately represents the relevant atomic environments.

Dataset	loop 0	loop 1	loop 7	validation
Model				
trained on loop 0	0.91	9.91	3.10	7.02
trained on loop 1	0.84	3.52	2.61	4.04
trained on loop 7	0.78	1.88	1.20	2.78
universal	1.81	2.14	2.68	1.51

Table 5.3: RMSE Residual energy error (in eV per configuration) for models obtained at different stages of the active learning scheme, and the universal model. Each model is evaluated on four datasets: (i) the training set from loop 0, which consists solely of cell-relaxation, (ii) short MD-generated configurations selected during loop 1, (iii) long MD-generated configurations selected during loop 7, and (iv) a validation set containing unseen Ab Initio MD configurations. The dataset names correspond to the training sets used at each respective loop.

A notable observation is that models trained in later active learning iterations perform better on the initial training set than the first model in terms of residual energy RMSE, despite the fact that the first model was trained

Dataset	loop 0	loop 1	loop 7	validation
Model				
trained on loop 0	0.16	0.47	0.46	0.78
trained on loop 1	0.18	0.21	0.37	0.32
trained on loop 7	0.18	0.20	0.18	0.24
universal	0.44	0.35	0.40	0.18

Table 5.4: RMSE per-atom force error (in eVÅ^{-1} per atom) for models obtained at different stages of the active learning scheme, and the universal model. Each model is evaluated on four datasets: (i) the training set from loop 0, which consists solely of cell relaxation data, (ii) short MD-generated configurations selected during loop 1, (iii) long MD-generated configurations selected during loop 7, and (iv) a validation set containing unseen Ab Initio MD configurations. The dataset names correspond to the training sets used at each respective loop.

exclusively on that data. Intuitively, since all models have the same number of parameters and therefore the same degree of flexibility, one might expect that introducing more data would reduce accuracy on this specific subset, as the model must now generalize over a broader potential energy surface. However, our results show the opposite—a model trained on a more diverse dataset outperforms the one trained solely on the initial data, suggesting that exposure to a wider range of configurations enhances the model's ability to capture underlying energy trends more effectively.

The universal model performs competitively in terms of residual energy errors and force predictions. It performs worse on loop 0 data compared to the dedicated models, reflecting a lack of fine-tuning for cell relaxation data. On the validation set, the universal model slightly outperforms the dedicated model. This is likely because the validation set consists exclusively of MD data, unlike the loop 0 training data which includes cell relaxation structures where the universal model shows to be less optimized. Nevertheless, the higher errors observed for training sets primarily composed of MD structures, exemplified by Loop 7, indicate that some MD configurations are still not accurately captured by the universal model. This suggest that there are still parts of the system-specific potential energy surface that are not well reproduced by the universal model.

While RMSE is a commonly used benchmark, the maximum force error (Table 5.5) provides important additional insight as it plays a critical role in simulation stability. The selection criterion used in the active learning loop is inherently local, so when a configuration is selected for DFT reference calculations, it is often due to a single atom with OOD behavior. However, the majority of atoms in that configuration do not show such behavior. As a result, these rare but important deviations can easily be obscured in RMSE calculations, which average errors over all atoms.

By analyzing maximum per-atom force errors, we observe that active learning effectively reduces extreme outliers. After just two iterations, the maximum force error on the validation set is already slightly lower than in the final model when evaluated against the validation set. This suggests that the most severe errors are addressed early in the learning process. The maximum force errors are still orders of magnitude higher than the RMSE force errors, especially when evaluated on training set issued at loop 7. This

Dataset	loop 0	loop 1	loop 7	validation
Model				
trained on loop 0	1.30	8.38	5.62	5.16
trained on loop 1	1.21	2.90	8.10	3.97
trained on loop 7	1.56	3.23	2.19	4.10
universal	4.49	5.48	20.89	1.47

Table 5.5: Maximum per-atom force error (in $\mathrm{eV\AA^{-1}}$ per atom) for models obtained at different stages of the active learning scheme, and the universal model. Each model is evaluated on four datasets: (i) the training set from loop 0, which consists solely of cell relaxation data, (ii) short MD-generated configurations selected during loop 1, (iii) long MD-generated configurations selected during loop 7, and (iv) a validation set containing unseen MD configurations. The dataset names correspond to the training sets used at each respective loop.

indicates that certain atomic environments remain challenging, even after training.

The maximum per-atom force errors show again that the universal model performs well on the validation set. However, on the training set from loop 7, it exhibits large force errors. This indicates that certain regions of the potential energy surface remain poorly represented by the universal model. This reinforces the limitation that, despite its general applicability, the universal model fails to accurately capture all relevant atomic environments.

The results confirm that active learning is an effective strategy for systematically improving MLPs while minimizing human intervention. By targeting rare but important atomic environments, the method ensures that the model remains well-conditioned and does not extrapolate unpredictably.

5.5.3 Time step Stability

The stability of MD simulations using MLPs was evaluated by analyzing the energy drift in NVE simulations at different time steps. The universal MLP and the dedicated MLP, as detailed in Table 5.1 were tested with increasing time step sizes to determine the maximum stable time step for each model. Stability was assessed by monitoring the conservation of the total energy, as shown in Figure 5.5 and Figure 5.6.

The universal MLP exhibited significant energy drift for time steps above 0.6 fs (Figure 5.5). While the energy remained relatively stable at 0.6 fs, increasing the time step to 0.8 fs or beyond led to noticeable deviations, with substantial energy drift observed at 1.0 fs, 1.2 fs, and 1.5 fs. The rapid divergence in total energy observed at larger time steps suggests numerical instabilities compromise the stability of atomic force integration. These observations of reduced time step stability are consistent with the higher force errors reported for the universal model in Section 5.5.2. One hypothesis relates to the smoothness of the potential energy surface learned by the universal model. Localized roughness or high-frequency components in the potential energy surface could translate into noisy force predictions, hindering stable integration with larger time steps. Another possibility is that the universal model extrapolates less reliably when atoms are in close proximity, particularly if such configurations are underrepresented in the universal training dataset. Finally, the large number of parameters and computational operations inherent to the universal model might introduce subtle numerical noise that accumulates during integration. Distinguishing between these potential causes and confirming their precise impact would require more thorough investigation, potentially including detailed analysis of the potential energy surface smoothness, vibrational spectra, and model behavior in specific close-contact scenarios.

In contrast, the dedicated MLP with a cutoff of 4 Å remained stable up to a time step of 1.2 fs (Figure 5.6). The total energy drift was minimal throughout the simulations at 1.0 fs and 1.2 fs, indicating consistent force predictions and good numerical stability. However, at 1.5 fs, energy drift became evident, suggesting that at this time step, the model's predicted forces are no longer sufficiently accurate to support stable MD integration. This result implies that the dedicated model enables significantly larger time steps than the universal model while maintaining energy conservation.

5.5.4 Computational Benchmarking

The computational efficiency of the universal MLP, the dedicated MLP and the ReaxFF model were evaluated by benchmarking their performance in MD simulations on systems of approximately 1500 atoms over a total time of 72 hours for each run. Two performance metrics were assessed: simulation

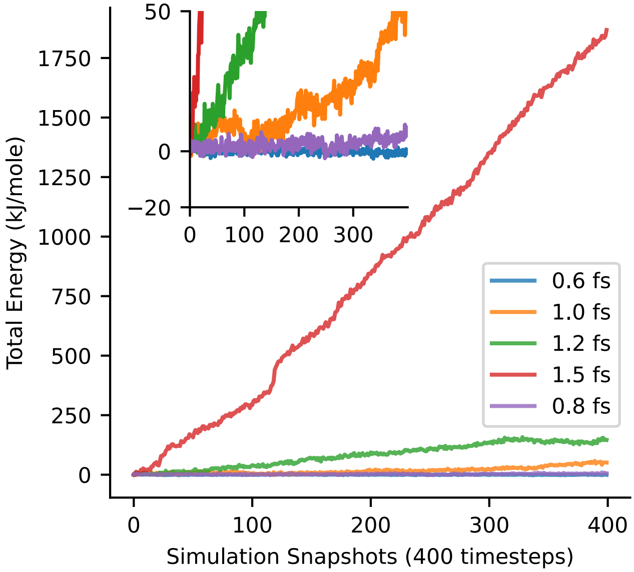

Figure 5.5: Total energy evolution in an NVE simulation with the universal model using different time steps. Total energy values are offset by their initial value to highlight drift over time. The inset zooms in on values between −20 and 50 kJmol^{-1} for clarity.

speed using the maximum stable time step (in nsd^{-1}), and GPU memory consumption if applicable. The results are summarized in Table 5.6.

ReaxFF simulations were performed using 16, 32, and 64 CPU cores with a fixed time step of 0.2 fs, commonly used in ReaxFF simulations with light elements. [266,276] The simulation speed increased from 0.46 nsd^{-1} (16 cores) to 0.86 nsd^{-1} (64 cores), demonstrating reasonable scalability with increasing computational resources. However, despite this improvement, the maximum time step strongly restricts the overall efficiency of the method.

As shown in Section 5.5.3, the universal model was constrained to a 0.6 fs time step. The universal MLP achieved a simulation speed of 0.46 nsd^{-1} with a high GPU memory requirement of 42.8 GB, which severely impacts the usability on larger systems. In contrast, the dedicated MLPs exhibited significantly improved efficiency. The dedicated MLP with a cut-off of 4 Å was stable at 1.2 fs and could therefore reach simulation speeds of 4.18 ns d^{-1};, nearly 10 times faster than the universal model and 5 times faster ReaxFF while maintaining stability. Moreover, its GPU memory footprint was only 3.4 GB, making it computationally efficient for extended simulations.

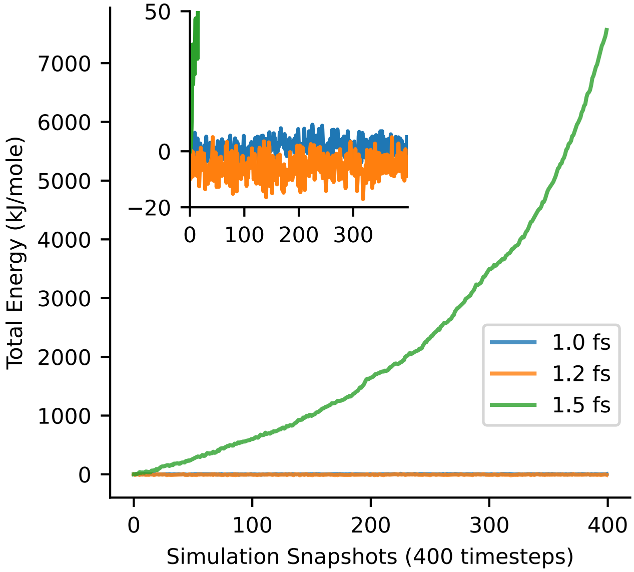

Figure 5.6: Total energy evolution in an NVE simulation with the dedicated model using different time steps. Total energy values are offset by their initial value to highlight drift over time. The inset zooms in on values between −20 and 50 kJmol^{-1} for clarity.

Increasing rmax to 5 Å resulted in higher memory consumption (17.9 GB) and a reduced speed (2.88 nsd^{-1}) which illustrates the trade-off between computational cost and accuracy.

Model	Time Step (fs) (fs)	Simulation Speed (nsd^{-1})	GPU Memory (GB)
Reaxff (16 cores)	0.20	0.46	N/A
Reaxff (32 cores)	0.20	0.75	N/A
Reaxff (64 cores)	0.20	0.86	N/A
Mace (universal)	0.60	0.46	42.76
Mace (ded; rmax4)	1.20	4.18	3.40
Mace (ded; rmax5)	1.20	2.88	17.94

Table 5.6: Computational Benchmark of ReaxFF on 16, 32 and 64 cores compared with a universal MLP, a dedicated MACE model with a cutoff of 4 Å and a dedicated MACE model with a cutoff of 5 Å.

The results highlight the substantial advantages of dedicated MLP models in terms of simulation speed and efficiency. The ability to use larger time steps while maintaining accuracy allows dedicated MLP to surpass ReaxFF and universal MLPs in both performance and scalability. This efficiency advantage extends to the environmental impact; The dedicated model produces over ten times less estimated CO_2 per simulated nanosecond compared to the universal model, as detailed in the Supplementary Information.[277] The reduced memory footprint, performance and scalability of the dedicated models compared to the universal MLP model suggests that tailoring MLPs to specific chemical systems is a viable strategy for optimizing efficiency.

5.5.5 Composition Evolution

To analyze the formation and stability of silicate species in HSILs, we investigated the RDF of Si-O pairs to determine bonding and breaking thresholds to count oligomeric species. Based on these thresholds to detect reactive events, the evolution of different silicate species over time was tracked during MD simulations at 425 K using the dedicated MLP. The results are presented in Figure 5.7 and Figure 5.8.

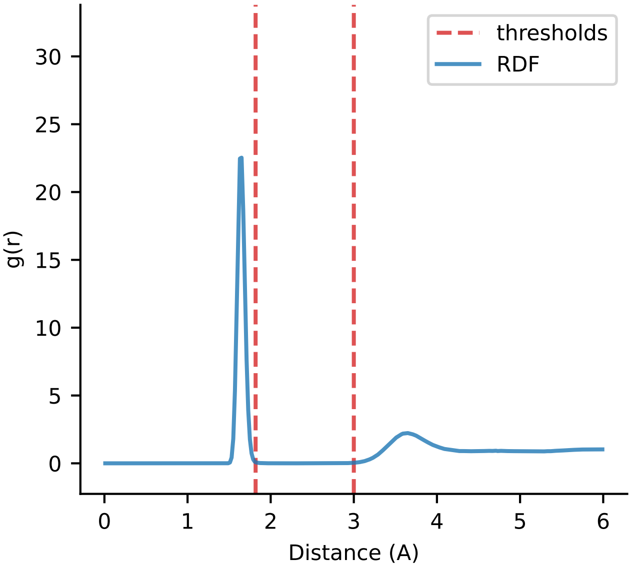

Figure 5.7: Radial Distribution Function of Si-O pairs. The right threshold corresponds with the bond-breaking threshold, while the left threshold corresponds with the bond formation.

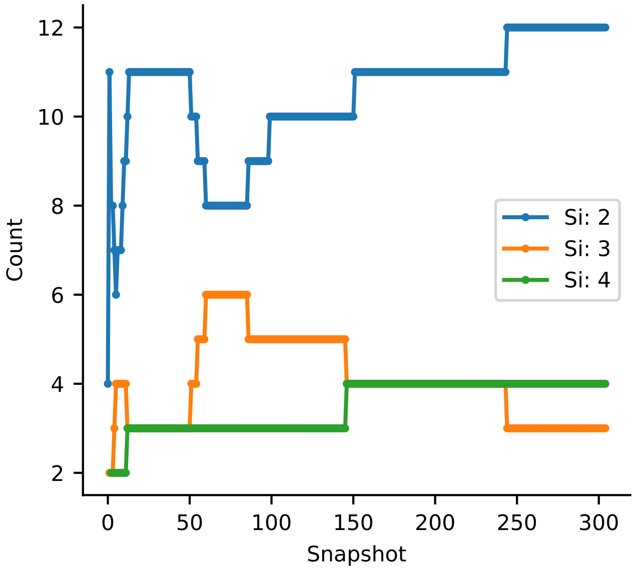

Figure 5.8: Concentration of different oligomers as a function of simulation time at composition (Si/OH=1; H_2O/OH=2). Si: 2 corresponds with dimers (blue), Si: 3 with trimers (orange) and Si: 4 with tetramers (green). Simulations are run at 425 K at fixed volume with the dedicated model.

The Si-O RDF (Figure 5.7) exhibits a pronounced first peak at approximately 1.6 Å, corresponding to the typical Si-O bond length. To distinguish bonded from non-bonded species, we define the bonding threshold as the first peak position plus 2 standard deviations. As mentioned in Section 5.3, a fixed value of 3 Å is used as the bond breaking criteria. This value is set manually as this marks the start of the second peak in the Si-O RDF, as seen in Figure 5.7. These RDF-derived thresholds were used to classify oligomeric species based on connectivity.

Over the total simulation time of 10 ns, the oligomer concentrations, shown in Figure 5.8, reveal the dynamic behavior of silicate oligomers at a Si/OH ratio of 1 and H_2O/OH ratio of 2.2, the highest concentration achievable in an experimental setting. Initially, dimers are the most abundant species, forming rapidly and reaching a stable count early in the simulation. Trimers and tetramers emerge as polymerization progresses, though their populations fluctuate.

Interestingly, a slight decrease in silicate concentration results in a significant drop in reactivity, with almost no bond formation events observed under these conditions. This indicates that the high-concentration regime plays an essential role in the reactivity of HSILs, likely by maintaining a sufficiently high density of reactive species and facilitating bond formation due to the coordination of reactive species within the constrained ionic liquid environment.

The observed trends suggest that oligomer formation stabilizes within the first few nanoseconds of the simulation, with dimers and trimers being the predominant species under these conditions. The threshold-based classification highlights the stability of short-chain silicate species, which may play a critical role in the nucleation and growth of larger silicate networks. These findings align with experimental NMR studies, which also observed the formation of small oligomers in HSIL solutions.[80] However, in their study, higher ratios of trimers and tetramers were reported compared to those observed in our simulations. This discrepancy is likely due to the limited timescales accessible in MD simulations, where the system has not yet reached full chemical equilibrium within the simulated timeframe. To capture longer-timescale reactivity and nucleation mechanisms, enhanced sampling techniques such as metadynamics should be explored.[278,279] These methods could provide a more comprehensive view of the oligomerization landscape and improve compositional accuracy compared with experimental results. Nonetheless, this analysis demonstrates the strength of MLPs in tracking molecular interactions with high fidelity and offers valuable insights into oligomer formation. Future studies exploring different compositions and reaction conditions could refine our understanding of oligomer evolution in HSILs and aid in the design of tailored synthesis protocols.

These results suggest that dedicated MLPs can effectively capture key structural features of HSILs. Further refinements, particularly in high-concentration regimes, may be necessary to improve the balance between short-range and long-range interactions. MLP's are inherently short-ranged, as they rely on local atomic environments within a finite interaction cutoff. This can limit their ability to fully capture long-range electrostatic and polarization effects that play a crucial role in HSILs. A possible avenue for improvement could be combining machine learned short-range forces physical models for long-range interactions, such as charge equilibration or multipole expansions or explicit electrons.[280–282] This hybrid approach may help refine density predictions and improve overall accuracy in modeling complex liquid-phase systems like HSILs.

5.5.6 Comparison of Experimental Density with Computational Models

The accuracy of the different models in predicting the density of HSILs was assessed by comparing their results to experimental values across a range of silicate concentrations. Figure 5.9 shows the predicted densities for the ReaxFF model, the universal MLP model, and the dedicated MLP model with cutoff distance 4 Å compared with the experimental results.

Among the tested models, the dedicated model is the only model that closely approaches the experimental density at diluted compositions. However, at higher concentrations, it overestimates the density, deviating from the experimental trend. In contrast, both ReaxFF and the universal MLP systematically underestimate the density across all concentrations. Despite these deviations, the magnitude of the error for all models remains comparable. This indicates that the dedicated MLP does not introduce additional bias but rather shifts the error trend from underestimation to overestimation.

Comparison of the compositional results, presented in Section 5.5.5, with composition measured in the experimental results through NMR suggest that equilibrium has not yet been reached. [80] Even though the simulations lasted up to 10 ns for the dedicated MLP, achieving full chemical equilibrium within this timeframe is indeed unlikely. This effect could further modulate density trends, particularly at higher concentrations where equilibration may require longer simulations.

5.6 Conclusions

This work details the development and successful implementation of an active learning workflow, centered on node energy densities, designed to autonomously generate dedicated MLPs tailored for complex chemical systems. We specifically showcase its utility by creating a dedicated MLP for HSILs and rigorously evaluating its performance, stability, and computational cost against both a conventional ReaxFF model and a universal MLP, thereby assessing its suitability for simulating intricate liquid-phase reactions.

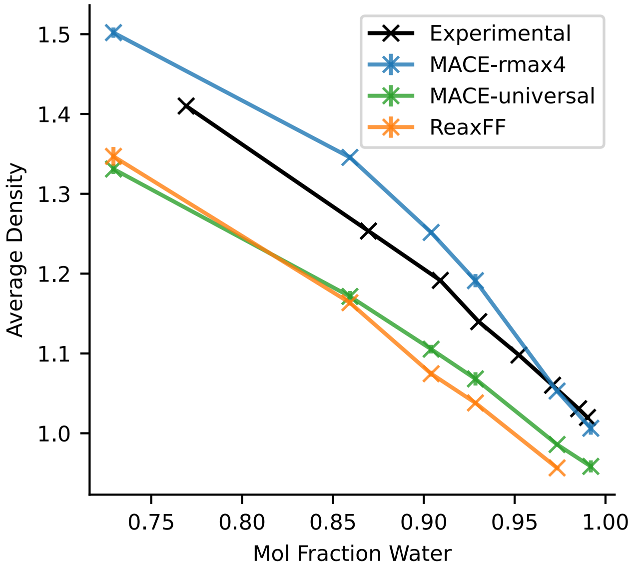

Figure 5.9: Density at different concentrations for compositions $M_2SiO_2(OH)_2.yH_2O$ using ReaxFF (green), universal MLP (orange) and dedicated MLP (blue). The experimental results (black) were taken from *Vandenabeele et al.* [80] Simulations are run at 300 K to be consistent with experiments.

In direct performance comparisons, the bespoke dedicated MLP significantly surpassed both ReaxFF and the universal model in terms of achievable simulation speed and the maximum stable integration time step. Our dedicated MLP consistently maintained energy conservation using a 1.2 fs time step, yielding a simulation throughput of 4.18 nsd^{-1}. This represents a nearly five-fold acceleration compared to ReaxFF (typically limited to 0.2 fs steps) and a ten-fold acceleration over the universal MLP (which struggled beyond 0.6 fs steps) for the same test system.

Further analysis focusing on predictive accuracy revealed that the dedicated model, trained via our active learning workflow, was uniquely capable of correctly reproducing system densities at low water concentrations. While density prediction proved challenging for all models at higher water concentrations—with ReaxFF and the universal model underestimating, and the dedicated model slightly overestimating—the magnitude of the error was smallest for our dedicated MLP.

Beyond structural properties, the simulations investigated chemical reactions, specifically silicate oligomerization dynamics. These simulations showed that at high concentrations, silicate oligomers formed rapidly and reached a nearly stationary regime within the initial simulation phase. Conversely, a slight decrease in concentration led to a dramatic drop in reactivity, with minimal bond formation observed. This underscores the critical role of high concentration regimes in promoting HSIL reactivity, likely driven by the increased probability of encounters between reactive species. Achieving true chemical equilibrium, however, remains beyond the reach of these simulation of 10 ns, suggesting that enhanced sampling techniques would be beneficial for a comprehensive understanding of oligomerization dynamics.

These findings suggest advantages for using dedicated MLPs in simulating specific complex systems compared to a universal model and ReaxFF. The observed balance of computational efficiency, accuracy, and stability results from both the dedicated model architecture and the data selection strategy employing KDE within the active learning workflow. Operating on node energies allows the KDE method to efficiently identify OOD configurations without requiring high-dimensional feature analysis or ensemble approaches. This process guides the expansion of the training set to incorporate relevant structural diversity while mitigating data redundancy and the computational expense of reference calculations. While some specific parameters guiding this data selection proved effective for the HSIL system, they were determined empirically. Future work should focus on generalizing these parameter choices to enhance the workflow's autonomy and reduce the need for system-specific tuning. Overall, this work demonstrates a method for generating dedicated MLPs suitable for large-scale reactive MD production runs of complex processes. Furthermore, this KDE-based data selection approach could potentially be applied to reduce dataset redundancy when training large universal MLPs.

6

Conclusions & Perspectives

"One never notices what has been done;
one can only see what remains to be done."
–Marie Curie

6.1 Conclusions

This document marked the beginning of the end of nearly five years of PhD research. The ambition driving this work was to advance the methodologies for developing accurate and computationally efficient reactive all-atom potentials. These potentials serve as the engine for MD simulations, a powerful, physics-based computational microscope used to explore the intricate dynamic processes during chemical reactions. The core challenge confronted in this work is the creation and curation of high-quality training data for these models, a task often hampered by the significant expense and relative scarcity of such data for reactive chemistry. This work proposes and validates innovative "Model-informed Training Data Curation" techniques, in this context aiming to render the development of both Reactive Force Fields (ReaxFF) and MLIPs more systematic, efficient, and insightful. By applying these advanced methodologies to chemically relevant and complex systems, specifically alumina and Hydrated Silicate Ionic Liquids, the research demonstrates effective pathways toward constructing superior reactive models capable of handling large-scale MD simulations. The primary contributions emerge from the studies presented in Chapters 4 and 5, that offer complementary solutions targeting different classes of reactive all-atomic potentials.

Chapter 4 addressed the prevalent challenge of balancing diverse data types in ReaxFF parameterization. The introduction and validation of the "Balanced Loss" function provided a novel solution, replacing subjective weighting with physically interpretable tolerances. This method streamlines the optimization, dynamically focusing on the most critical data discrepancies using a Log-Sum-Exp formulation. Optimization using Balanced Loss not only yields an improved force field for simulating water adsorption on alumina, but also provided crucial and interpretable feedback on the intrinsic capabilities and limitations of the ReaxFF model itself relative to defined physical expectations.

Subsequently, Chapter 5 tackled the efficient generation of training data for bespoke, high-accuracy MLIPs, aiming to create models suitable for large-scale MD where universal models might be too computationally demanding. An innovative active learning framework was introduced, leveraging Kernel Density Estimation on predicted "atomic" node energies to autonomously identify the most valuable out-of-distribution (OOD) structures needed for training refinement. Applied to generate a MACE potential for the complex HSIL system, this framework demonstrated a practical path towards developing tailored MLIPs with minimal computational overhead. Although this study only used node energies, any atomic predicted property could hypothetically be used to select OOD structures. The study showed that such dedicated MLIPs can achieve computational throughput for large, reactive MD simulations exceeding established empirical potentials like ReaxFF and universal MLIPs.

Rather than viewing data generation as a brute-force prerequisite, the methodologies developed here embed model awareness into the data handling process itself. The Balanced Loss provides a systematic way to manage expectations and leverage diverse datasets, while the node-energy active learning provides an efficient engine for dataset growth. Both techniques contribute significantly to making the development of potentials capable of driving large-scale, physically sound MD simulations more tractable and reliable. Moreover, they can be readily implemented in existing workflows, and therefore also be combined together in a single workflow for developing reactive potentials.

This brings this work to its final conclusion. I believe this research adds value to the field of computational modeling by focusing on a crucial, yet often overlooked, aspect: the intelligent handling of training data. While many studies concentrate on developing new modeling methodologies, improving or transferring refinement/optimization algorithms and workflows, this work emphasizes the interface between the model and the data used to build it. The true value of the presented techniques lies not only in enabling more efficient ways to weight or generate training data but also in making the potential development process more comprehensible. In a scientific landscape where computational tools increasingly resemble ‘black boxes’, the ability to systematically understand and validate the choices made during model training, as facilitated by these methods, is a contribution that cannot be overlooked.

6.2 Perspectives

Concluding nearly five years of PhD research invites critical reflection on the work accomplished and its place within the evolving landscape of molecular modeling. A primary question emerging from this period concerns the coexistence and future interplay between reactive force fields, like ReaxFF, and the rapidly advancing MLIPs. When this research began in 2020, the focus was primarily on using machine learning for ReaxFF optimization. Although MLIPs existed, they required vast datasets. However, the field witnessed an explosion in MLIP data efficiency within months, leading to widespread discussion, even among prominent developers like Prof. Dr. Adri Van Duin, about the potential obsolescence of traditional reactive force fields. This work also underscored the contrast in development processes, showing that MLIP training is often notably more direct than ReaxFF parameterization, requiring less manual intervention for tasks like parameter selection, setting parameter boundaries and data curation.

While the progress in MLIPs is undeniable, the relationship between these methodologies need not be entirely polarized. Force field approaches may still offer unique value. Despite their complexity, traditional force fields are built upon empirically defined functional forms, with parameters ideally motivated by physical principles or, at minimum, rationalized human insight. This structure can, hypothetically, make interpreting their fidelity more straightforward than assessing MLIPs, which often function more like ‘black boxes’. Furthermore, based on the fundamental number of operations involved, force fields should inherently possess the potential for greater computational speed than MLIPs. While this holds true for many non-reactive force fields, this work highlights that current ReaxFF implementations do not always realize this theoretical advantage. Lastly, the process of developing and parameterizing force fields retains significant didactic value, offering scientists opportunities to learn deeply about the connections between model parameters and the reproduction of chemical, material, or biological phenomena; an insight less directly accessible when training MLIPs, which leans more heavily on computer science approaches and are ‘agnostic’ to any phyiscal laws.

However, realizing these potential advantages hinges on a crucial caveat: the understandability of the force field itself. If the core functional form and parameter interactions within a model like ReaxFF become opaque, its capacity for physics-informed improvement diminishes, and improved implementations become disproportionately difficult. Recent efforts, such as Jax-ReaxFF and PuReMD-GPU, aim to address performance limitations through GPU acceleration, yet they do not fundamentally alter the complexity that has made ReaxFF itself resemble a ‘black box’.[283,284] To remain competitive and relevant, reactive force fields may need to emphasize their physics-informed nature and interpretability, rather than solely competing on computational efficiency, though efficiency gains from clearer implementations would certainly help.

Perhaps the most promising direction lies not in strict competition, but in synergy. Physics-informed MLIPs, which combine the learned representations of MLIPs with physically motivated terms, represent an exciting frontier. Current MLIPs are often inherently short-range; incorporating such terms could explicitly add crucial non-local interactions, such as the long-range electrostatics and polarization recently explored to enhance (universal) MLIPs. Such hybrid models could truly represent the next generation of interatomic potentials.

The other pertinent question concerns the enduring relevance of this thesis’s specific contribution in a future possibly dominated by universal MLIPs and vastly increased computational power. Even in such a scenario, these methodologies retain significance. Firstly, the Balanced Loss concept is not intrinsically tied to ReaxFF; its principles are applicable to any multivariate optimization problem where training data can be meaningfully categorized. It could, therefore, readily be adapted for optimizing MLIPs, offering the same benefits of improved insight and systematic handling of diverse training objectives. Secondly, while universal models aim for broad applicability, the datasets they require are enormous and likely contain significant redundancies. Additionally, the considerable energy cost associated with training and frequently deploying these large universal models raises sustainability concerns often overlooked in the pursuit of broad applicability. The computationally inexpensive node-energy-based active learning strategy presented here could prove valuable for intelligently curating and reducing the size of these massive datasets, ensuring efficiency without sacrificing essential information, thus remaining relevant even as the landscape shifts towards large, pre-trained models.

Ultimately, the path forward in computational chemistry likely lies not in definitively choosing between established empirical potentials and emergent machine learning techniques, but in intelligently combining their respective strengths. Regardless of the specific models that dominate future simulations, the underlying principles of systematic, insightful, and comprehensible model development explored throughout this research will remain crucial to propelling chemical science forward through the power of simulation.

Part II

Appendices and Conference Contributions

A

List of Publications and Pre-prints

1 **Managing Expectations and Imbalanced Training Data in Reactive Force Field Development: An Application to Water Adsorption on Alumina**
Loïc Dumortier, Céline Chizallet, Benoit Creton, Theodorus de Bruin and Toon Verstraelen
Journal of Chemical Theory and Computation, **2024**, *20*, 3779-3797

2 **Scalable Reactive Molecular Dynamics with Machine Learning Potentials: Active Learning with Node Energy Density Estimation**
Loïc Dumortier, Jelle Vekeman, Benoit Creton, Theodorus de Bruin and Toon Verstraelen
To Be Submitted in Journal of Chemical Theory and Computation, **2025**

B

List of Conference Contributions, Workshops and Awards

B.1 Conference Contributions

1 **A reactive force field for alumina systems**
Loïc Dumortier, Benoit Creton, Theodorus de Bruin and Toon Verstraelen
AutoCheMo International Reactive Force Field Workshop
December, 2021
Ghent, Belgium

2 **A Reactive Force Field for Alumina Systems**
Loïc Dumortier, Benoit Creton, Theodorus de Bruin and Toon Verstraelen
CRF-Chemcys 2022
October, 2022
Blankenberge, Belgium

3 **Balanced Loss: a Novel Cost-Function for ReaxFF Optimization**
Loïc Dumortier, Benoit Creton, Theodorus de Bruin and Toon Verstraelen
COST Action: CA18234 - Computational materials sciences for efficient water splitting with nanocrystals from abundant elements
Februari, 2023
Brussels, Belgium

4 **Balanced Loss: a novel cost function for ReaxFF optimization to omit manual weighting coefficients**
Loïc Dumortier, Benoit Creton, Theodorus de Bruin and Toon Verstraelen
ACS Fall Meeting 2023
August, 2023
San Francisco, United States of America

5 **Studying Alumina Species Interacting with Zeolitic Frameworks using ReaxFF (*under review*)**
Loïc Dumortier, Laureline Treps, Benoît Creton and Toon Verstraelen, Céline Chizallet, Theodorus de Bruin
WATOC 2023
June, 2025
Oslo, Norway

6 **Machine Learning Potentials at Force Field Efficiency (*under review*)**
Loïc Dumortier, Jelle Vekeman, Benoît Creton, Theodorus de Bruin and Toon Verstraelen
WATOC 2023
June, 2025
Oslo, Norway

B.2 Workshops

1 **MolSim-2022**
January, 2022
Amsterdam, Netherlands

B.3 Awards

1 **SCM ParAMS Challenge Winner**
June, 2022

C Appendices

C.1 Used Software

- **VASP:** The Vienna Ab Initio Software Package. A sophisticated software package for performing ab initio quantum mechanical calculations. It primarily uses Density Functional Theory, plane-wave basis sets, and pseudopotentials to compute the electronic structure and properties of materials, molecules, and surfaces. [261,262]
- **MACE:** Multi-Atomic Cluster Expansion, the main Machine learning interatomic potential used in this work. [64] MACE uses atomic cluster expansion and to create equivariant neural networks for energy predictions based on molecular structures.
- **Atomic Simulation Environment:** A Python library designed for setting up, manipulating, running, and analyzing atomistic simulations. [190] ASE provides a powerful interface to numerous simulation codes (calculators) and includes tools for structure manipulation, visualization, and workflow automation. ASE is used extensively for data wrangling and prototyping.
- **AMS:** The Amsterdam Modeling suite integrates various simulation methods, including Density Functional Theory (ADF engine), Density Functional Tight Binding (DFTB), classical force fields (e.g., ReaxFF), and continuum solvation models (COSMO-RS), often accompanied by a graphical user interface (GUI).[285,286] This work used AMS extensively to parametrize ReaxFF force fields and run MD simulations.
- **OpenMM:** A high-performance toolkit specifically designed for molecular simulation, particularly classical Molecular Dynamics. [287–289] It functions as a library that can be integrated into other software or used standalone, notable for its flexibility, extensibility, and efficient utilization of hardware accelerators like GPUs.
- **Parman**: An in-house developed workflow manager. Purely written in Python and extending the Futures-class, it allows to distribute workloads efficiently and parallellize all the way from computers to multi-node HPC calculations. [290]

- **StepUp**: An in-house developed build tool, that is extendeble as a workflow manager. It is used to produce this work and papers in a reproducible and transparant way.
- **Typst**: This thesis is written in Typst. [291,292] Typst is a modern, markup-based typesetting system used for producing scientific documents, including articles, presentations, and theses. It aims to provide a more streamlined syntax and powerful scripting capabilities compared to traditional systems like LaTeX, facilitating the creation of complex layouts and reproducible documents.[293]

C.2 Supporting Information Paper I

Supporting Information: Managing Expectations and Imbalanced Training Data in Reactive Force Field Development: An Application to Water Adsorption on Alumina

Loïc Dumortier, Céline Chizallet, Benoit Creton, Theodorus de Bruin and Toon Verstraelen

Journal of Chemical Theory and Computation, **2024**, *20*, 3779-3797

Supporting Information for

Managing Expectations and Imbalanced Training Data in Reactive Force Field Development: an Application to Water Adsorption on Alumina

Loïc Dumortier,[†,‡] Céline Chizallet,[¶] Benoit Creton,[†] Theodorus de Bruin,[†] and Toon Verstraelen[*,‡]

[†]*IFP Energies nouvelles, 1 et 4 Avenue de Bois-Préau, 92852 Rueil-Malmaison, France*
[‡]*Center for Molecular Modeling (CMM), Ghent University, Technologiepark-Zwijnaarde 46, B-9052, Zwijnaarde, Belgium*
[¶]*IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP3, 69360 Solaize, France*

E-mail: toon.verstraelen@ugent.be

Contents

S1 Additional Display Items

Table S1: Structures in the training set and contributions to each category of data. Bonds are represented by pairs of chemical elements, angles by triplets. The symbol ··· denotes a hydrogen bond. Internal coordinates with oxygen not bound to aluminum are discarded. For hydrated alumina surfaces and edges, and for boehmite surfaces, the internal coordinates without hydrogen are not considered.

Structure	Chem. Form.	O–H	O···H	Al–O	Al–Al	O–Al–O	Al–O–Al	Al–O–H	Al–O···H	H–O···H	H···O···H
alpha_bulk	$Al_{12}O_{18}$			72	84	180	108				
boehm_bulk	$Al_{32}O_{64}H_{32}$	32	32	192	128	480	224	64	64	32	
boehm_surf-001_00w	$Al_{48}O_{96}H_{48}$	48	32	272	160	640	288	96	64	32	
boehm_surf-001_08w	$Al_{48}O_{112}H_{80}$	80	40		160			160	80	40	
boehm_surf-010_00w	$Al_{72}O_{144}H_{72}$	72	48	432	288	1080	504	144	96	48	
boehm_surf-100_00w	$Al_{48}O_{96}H_{48}$	48	48	256	160	576	272	80	80	48	
boehm_surf-100_12w	$Al_{48}O_{120}H_{96}$	96	80		160			128	112	96	24
boehm_surf-100_16w	$Al_{48}O_{128}H_{112}$	112	89		160			144	137	98	24
boehm_surf-101_00w	$Al_{48}O_{96}H_{48}$	49	33	254	168	560	250	100	68	18	6
boehm_surf-101_12w	$Al_{48}O_{120}H_{96}$	96	66		156			156	102	72	18
gamma_bulk	$Al_{16}O_{24}$			88	88	204	120				
gamma_edge-100-110_00w	$Al_{96}O_{144}$			481	424	1003	598				
gamma_edge-100-110_01w	$Al_{96}O_{145}H_{2}$	2			425			3			
gamma_edge-100-110_02w	$Al_{96}O_{146}H_{4}$	4	1		426			6	2		
gamma_edge-100-110_03w	$Al_{96}O_{147}H_{6}$	6	2		422			10	4		
gamma_edge-100-110_04w	$Al_{96}O_{148}H_{8}$	8	3		426			13	6		1
gamma_edge-100-110_05w	$Al_{96}O_{149}H_{10}$	10	4		424			17	8		2
gamma_edge-100-110_07w	$Al_{96}O_{151}H_{14}$	14	7		424			23	13	2	2
gamma_surf-100_00w	$Al_{64}O_{96}$			320	280	656	384				
gamma_surf-100_01w	$Al_{64}O_{100}H_{8}$	8	4		284			16	4	4	
gamma_surf-100_02w	$Al_{64}O_{104}H_{16}$	16	8		282			24	8	8	4
gamma_surf-100_03w	$Al_{64}O_{108}H_{24}$	24	16		280			40	16	16	8
gamma_surf-100_04w	$Al_{64}O_{112}H_{32}$	32	20		280			52	28	20	8
gamma_surf-110_00w	$Al_{64}O_{96}$			304	240	600	368				
gamma_surf-110_01w	$Al_{64}O_{100}H_{8}$	8	4		232			16	8		
gamma_surf-110_02w	$Al_{64}O_{104}H_{16}$	16	8		240			24	16		
gamma_surf-110_03w	$Al_{64}O_{108}H_{24}$	24	12		240			44	16	8	4
gamma_surf-110_04w	$Al_{64}O_{112}H_{32}$	32	24		240			52	44	12	8
gamma_surf-110_05w	$Al_{64}O_{116}H_{40}$	40	28		240			60	52	12	12
gamma_surf-110_06w	$Al_{64}O_{120}H_{48}$	48	32		236			88	32	32	12
gamma_surf-111_00w	$Al_{40}O_{60}$			212	190	480	286				
gamma_surf-111_04w	$Al_{32}O_{56}H_{16}$	16	8		144			30	14	6	
gamma_surf-111_05w	$Al_{32}O_{58}H_{20}$	20	10		146			36	14	10	2
gamma_surf-111_06w	$Al_{32}O_{60}H_{24}$	24	12		146			44	16	12	2
monomer	$AlO_{4}H_{5}$	5		4		6		5			
water	OH_{2}										
total		990	671	2887	8383	6465	3402	1675	1104	626	137

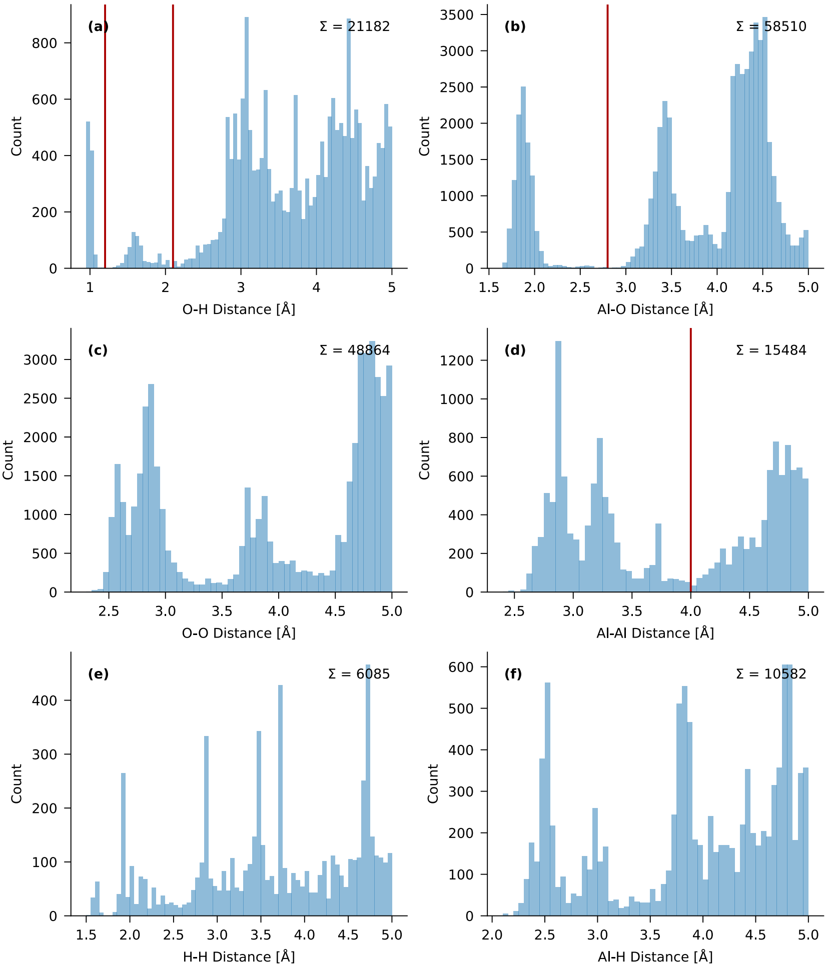

Figure S1: Histograms of all interatomic distances in the training set up to 5 Å, grouped per pair of chemical elements. Cutoffs for OH and AlO pairs depicted as vertical red lines: 1.2 Å for O−H bonds, 2.1 Å for hydrogen bonds and 2.8 Å for Al−O bonds. See main text for a more detailed description.

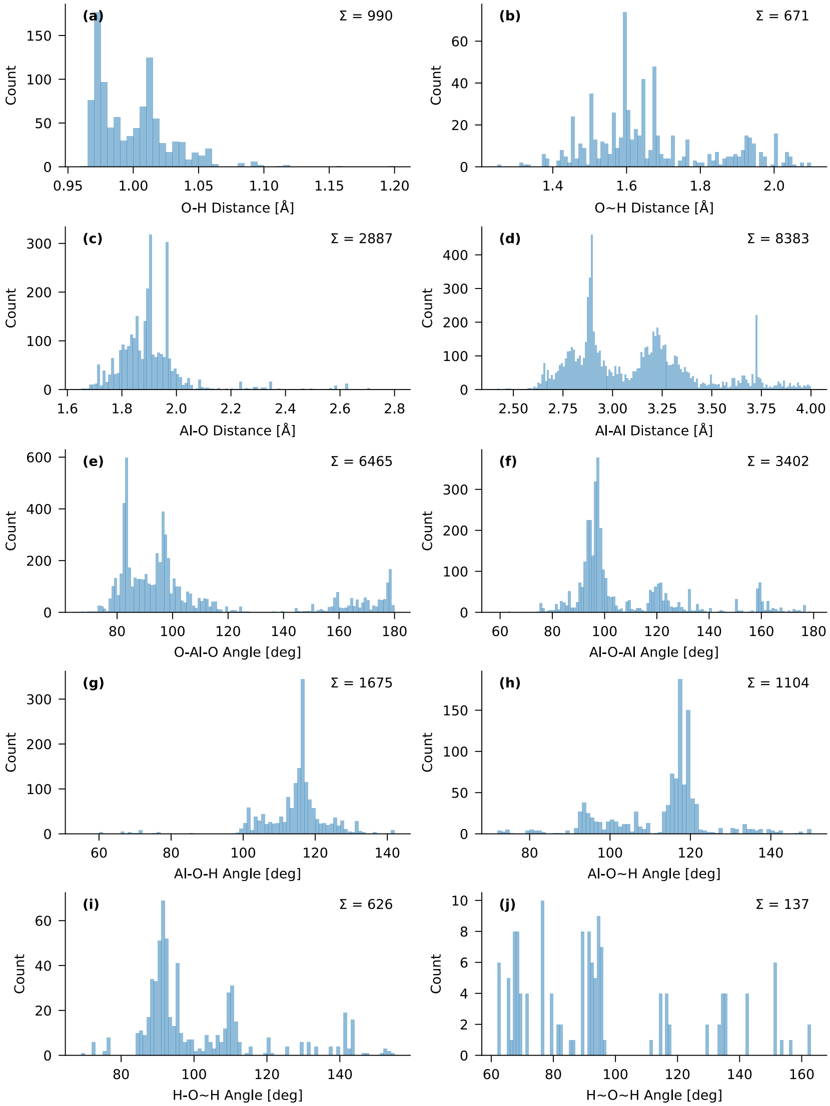

Figure S2: Histograms of internal coordinates in the training set. In the labels on the horizontal axis, a dash represents a regular bond and a tilde represents a hydrogen bond.

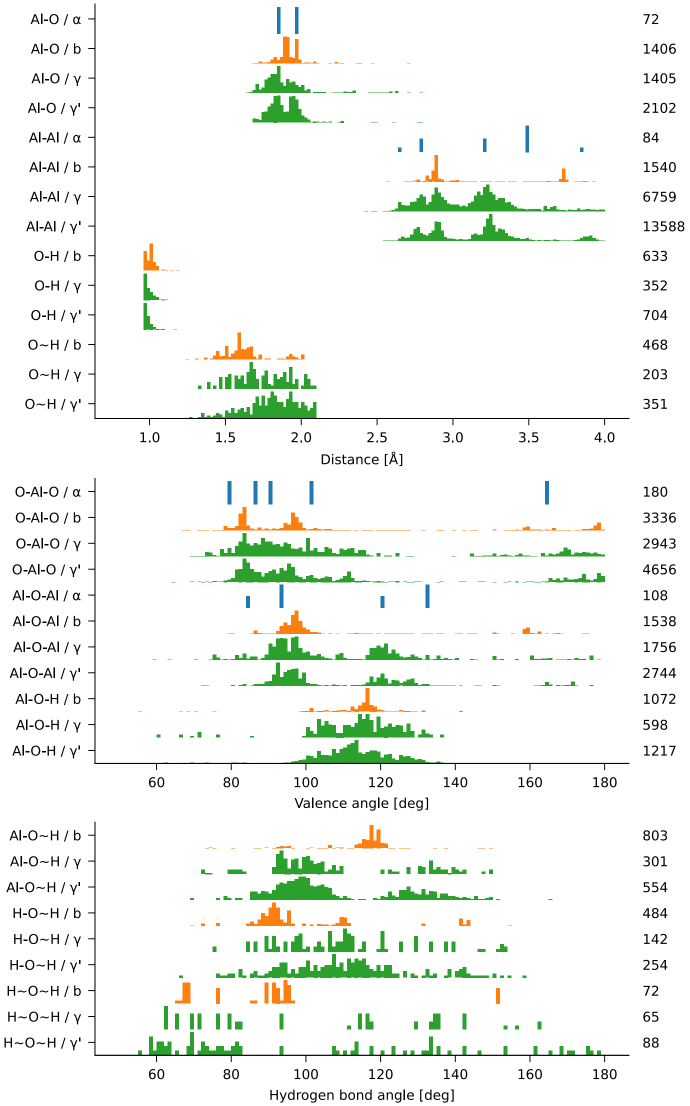

Figure S3: Histograms of internal coordinates, using the notation of Figure S2, per material. Histograms for the training set: $\alpha = \alpha$-Al_2O_3 (blue), b = boehmite (orange), $\gamma = \gamma$-Al_2O_3 (green). Histograms for the validation set: $\gamma' = \gamma$-Al_2O_3 (green). The number of internal coordinates within each class is shown to the right of the corresponding histogram.

Table S2: Overview of all chemical equations in the training set. Reactants are given negative coefficients. For each reaction, three reaction energies are in $\mathrm{kcal\,mol^{-1}}$**: the reference DFT result (R), the prediction with the Joshi force field (J) and the prediction with the new force field in this work (B). Water adsorption energies are normalized to the number of water molecules. All other reaction energies are normalized on the number of Al atoms. The categories are defined in the main text.**

Category	Coeff.	Structure	Chem. Form.	Reaction energy	
BSH	−1/16×	`boehm_surf-001_00w`	$Al_{48}O_{96}H_{48}$	R	−24.7
	−1×	`water`	OH_2	J	28.0
	1/16×	`boehm_surf-001_08w`	$Al_{48}O_{112}H_{80}$	B	−21.9
BSH	−1/24×	`boehm_surf-100_00w`	$Al_{48}O_{96}H_{48}$	R	−46.0
	−1×	`water`	OH_2	J	−1.7
	1/24×	`boehm_surf-100_12w`	$Al_{48}O_{120}H_{96}$	B	−38.0
BSH	−1/32×	`boehm_surf-100_00w`	$Al_{48}O_{96}H_{48}$	R	−38.0
	−1×	`water`	OH_2	J	−4.2
	1/32×	`boehm_surf-100_16w`	$Al_{48}O_{128}H_{112}$	B	−32.2
BSH	−1/8×	`boehm_surf-100_12w`	$Al_{48}O_{120}H_{96}$	R	−14.0
	−1×	`water`	OH_2	J	−11.9
	1/8×	`boehm_surf-100_16w`	$Al_{48}O_{128}H_{112}$	B	−15.1
BSH	−1/24×	`boehm_surf-101_00w`	$Al_{48}O_{96}H_{48}$	R	−41.0
	−1×	`water`	OH_2	J	−4.3
	1/24×	`boehm_surf-101_12w`	$Al_{48}O_{120}H_{96}$	B	−39.3
GEH	−1×	`gamma_edge-100-110_00w`	$Al_{96}O_{144}$	R	−104.1
	−1×	`water`	OH_2	J	−101.6
	1×	`gamma_edge-100-110_01w`	$Al_{96}O_{145}H_2$	B	−103.0
GEH	−1/2×	`gamma_edge-100-110_00w`	$Al_{96}O_{144}$	R	−81.2
	−1×	`water`	OH_2	J	−86.6
	1/2×	`gamma_edge-100-110_02w`	$Al_{96}O_{146}H_4$	B	−80.3
GEH	−1×	`gamma_edge-100-110_01w`	$Al_{96}O_{145}H_2$	R	−58.3
	−1×	`water`	OH_2	J	−71.5
	1×	`gamma_edge-100-110_02w`	$Al_{96}O_{146}H_4$	B	−57.5
GEH	−1/3×	`gamma_edge-100-110_00w`	$Al_{96}O_{144}$	R	−72.1
	−1×	`water`	OH_2	J	−18.0
	1/3×	`gamma_edge-100-110_03w`	$Al_{96}O_{147}H_6$	B	−71.2
GEH	−1/2×	`gamma_edge-100-110_01w`	$Al_{96}O_{145}H_2$	R	−56.1
	−1×	`water`	OH_2	J	23.8
	1/2×	`gamma_edge-100-110_03w`	$Al_{96}O_{147}H_6$	B	−55.3
GEH	−1×	`gamma_edge-100-110_02w`	$Al_{96}O_{146}H_4$	R	−53.9
	−1×	`water`	OH_2	J	119.1
	1×	`gamma_edge-100-110_03w`	$Al_{96}O_{147}H_6$	B	−53.0
GEH	−1/4×	`gamma_edge-100-110_00w`	$Al_{96}O_{144}$	R	−67.5
	−1×	`water`	OH_2	J	−71.0

	1/4×	`gamma_edge-100-110_04w`	$Al_{96}O_{148}H_8$	B	−67.0
GEH	−1/3×	`gamma_edge-100-110_01w`	$Al_{96}O_{145}H_2$	R	−55.3
	−1×	`water`	OH_2	J	−60.8
	1/3×	`gamma_edge-100-110_04w`	$Al_{96}O_{148}H_8$	B	−55.0
GEH	−1/2×	`gamma_edge-100-110_02w`	$Al_{96}O_{146}H_4$	R	−53.9
	−1×	`water`	OH_2	J	−55.4
	1/2×	`gamma_edge-100-110_04w`	$Al_{96}O_{148}H_8$	B	−53.7
GEH	−1×	`gamma_edge-100-110_03w`	$Al_{96}O_{147}H_6$	R	−53.8
	−1×	`water`	OH_2	J	−230.0
	1×	`gamma_edge-100-110_04w`	$Al_{96}O_{148}H_8$	B	−54.5
GEH	−1/5×	`gamma_edge-100-110_00w`	$Al_{96}O_{144}$	R	−63.3
	−1×	`water`	OH_2	J	−22.9
	1/5×	`gamma_edge-100-110_05w`	$Al_{96}O_{149}H_{10}$	B	−63.8
GEH	−1/4×	`gamma_edge-100-110_01w`	$Al_{96}O_{145}H_2$	R	−53.1
	−1×	`water`	OH_2	J	−3.2
	1/4×	`gamma_edge-100-110_05w`	$Al_{96}O_{149}H_{10}$	B	−54.0
GEH	−1/3×	`gamma_edge-100-110_02w`	$Al_{96}O_{146}H_4$	R	−51.4
	−1×	`water`	OH_2	J	19.5
	1/3×	`gamma_edge-100-110_05w`	$Al_{96}O_{149}H_{10}$	B	−52.8
GEH	−1/2×	`gamma_edge-100-110_03w`	$Al_{96}O_{147}H_6$	R	−50.2
	−1×	`water`	OH_2	J	−30.3
	1/2×	`gamma_edge-100-110_05w`	$Al_{96}O_{149}H_{10}$	B	−52.7
GEH	−1×	`gamma_edge-100-110_04w`	$Al_{96}O_{148}H_8$	R	−46.6
	−1×	`water`	OH_2	J	169.4
	1×	`gamma_edge-100-110_05w`	$Al_{96}O_{149}H_{10}$	B	−51.0
GEH	−1/7×	`gamma_edge-100-110_00w`	$Al_{96}O_{144}$	R	−55.8
	−1×	`water`	OH_2	J	−12.6
	1/7×	`gamma_edge-100-110_07w`	$Al_{96}O_{151}H_{14}$	B	−57.1
GEH	−1/6×	`gamma_edge-100-110_01w`	$Al_{96}O_{145}H_2$	R	−47.8
	−1×	`water`	OH_2	J	2.2
	1/6×	`gamma_edge-100-110_07w`	$Al_{96}O_{151}H_{14}$	B	−49.5
GEH	−1/5×	`gamma_edge-100-110_02w`	$Al_{96}O_{146}H_4$	R	−45.7
	−1×	`water`	OH_2	J	16.9
	1/5×	`gamma_edge-100-110_07w`	$Al_{96}O_{151}H_{14}$	B	−47.9
GEH	−1/4×	`gamma_edge-100-110_03w`	$Al_{96}O_{147}H_6$	R	−43.6
	−1×	`water`	OH_2	J	−8.6
	1/4×	`gamma_edge-100-110_07w`	$Al_{96}O_{151}H_{14}$	B	−46.6
GEH	−1/3×	`gamma_edge-100-110_04w`	$Al_{96}O_{148}H_8$	R	−40.2
	−1×	`water`	OH_2	J	65.1
	1/3×	`gamma_edge-100-110_07w`	$Al_{96}O_{151}H_{14}$	B	−43.9
GEH	−1/2×	`gamma_edge-100-110_05w`	$Al_{96}O_{149}H_{10}$	R	−37.0
	−1×	`water`	OH_2	J	13.0
	1/2×	`gamma_edge-100-110_07w`	$Al_{96}O_{151}H_{14}$	B	−40.4
GSH	−1/4×	`gamma_surf-100_00w`	$Al_{64}O_{96}$	R	−24.0

	−1×	`water`	$O H_2$	J	−2.0
	1/4×	`gamma_surf-100_01w`	$Al_{64} O_{100} H_8$	B	−23.8
GSH	−1/8×	`gamma_surf-100_00w`	$Al_{64} O_{96}$	R	−25.0
	−1×	`water`	$O H_2$	J	−10.7
	1/8×	`gamma_surf-100_02w`	$Al_{64} O_{104} H_{16}$	B	−26.0
GSH	−1/4×	`gamma_surf-100_01w`	$Al_{64} O_{100} H_8$	R	−26.0
	−1×	`water`	$O H_2$	J	−19.5
	1/4×	`gamma_surf-100_02w`	$Al_{64} O_{104} H_{16}$	B	−28.2
GSH	−1/12×	`gamma_surf-100_00w`	$Al_{64} O_{96}$	R	−24.3
	−1×	`water`	$O H_2$	J	−6.8
	1/12×	`gamma_surf-100_03w`	$Al_{64} O_{108} H_{24}$	B	−26.8
GSH	−1/8×	`gamma_surf-100_01w`	$Al_{64} O_{100} H_8$	R	−24.4
	−1×	`water`	$O H_2$	J	−9.2
	1/8×	`gamma_surf-100_03w`	$Al_{64} O_{108} H_{24}$	B	−28.3
GSH	−1/4×	`gamma_surf-100_02w`	$Al_{64} O_{104} H_{16}$	R	−22.8
	−1×	`water`	$O H_2$	J	1.1
	1/4×	`gamma_surf-100_03w`	$Al_{64} O_{108} H_{24}$	B	−28.3
GSH	−1/16×	`gamma_surf-100_00w`	$Al_{64} O_{96}$	R	−21.7
	−1×	`water`	$O H_2$	J	−3.6
	1/16×	`gamma_surf-100_04w`	$Al_{64} O_{112} H_{32}$	B	−22.4
GSH	−1/12×	`gamma_surf-100_01w`	$Al_{64} O_{100} H_8$	R	−21.0
	−1×	`water`	$O H_2$	J	−4.1
	1/12×	`gamma_surf-100_04w`	$Al_{64} O_{112} H_{32}$	B	−22.0
GSH	−1/8×	`gamma_surf-100_02w`	$Al_{64} O_{104} H_{16}$	R	−18.5
	−1×	`water`	$O H_2$	J	3.5
	1/8×	`gamma_surf-100_04w`	$Al_{64} O_{112} H_{32}$	B	−18.8
GSH	−1/4×	`gamma_surf-100_03w`	$Al_{64} O_{108} H_{24}$	R	−14.1
	−1×	`water`	$O H_2$	J	6.0
	1/4×	`gamma_surf-100_04w`	$Al_{64} O_{112} H_{32}$	B	−9.3
GSH	−1/4×	`gamma_surf-110_00w`	$Al_{64} O_{96}$	R	−92.4
	−1×	`water`	$O H_2$	J	50.4
	1/4×	`gamma_surf-110_01w`	$Al_{64} O_{100} H_8$	B	−108.6
GSH	−1/8×	`gamma_surf-110_00w`	$Al_{64} O_{96}$	R	−50.2
	−1×	`water`	$O H_2$	J	7.3
	1/8×	`gamma_surf-110_02w`	$Al_{64} O_{104} H_{16}$	B	−61.8
GSH	−1/4×	`gamma_surf-110_01w`	$Al_{64} O_{100} H_8$	R	−8.0
	−1×	`water`	$O H_2$	J	−35.8
	1/4×	`gamma_surf-110_02w`	$Al_{64} O_{104} H_{16}$	B	−15.0
GSH	−1/12×	`gamma_surf-110_00w`	$Al_{64} O_{96}$	R	−42.5
	−1×	`water`	$O H_2$	J	42.3
	1/12×	`gamma_surf-110_03w`	$Al_{64} O_{108} H_{24}$	B	−50.1
GSH	−1/8×	`gamma_surf-110_01w`	$Al_{64} O_{100} H_8$	R	−17.6
	−1×	`water`	$O H_2$	J	38.2
	1/8×	`gamma_surf-110_03w`	$Al_{64} O_{108} H_{24}$	B	−20.8

GSH	−1/4×	`gamma_surf-110_02w`	$Al_{64}O_{104}H_{16}$	R	−27.2
	−1×	`water`	OH_2	J	112.3
	1/4×	`gamma_surf-110_03w`	$Al_{64}O_{108}H_{24}$	B	−26.7
GSH	−1/16×	`gamma_surf-110_00w`	$Al_{64}O_{96}$	R	−39.6
	−1×	`water`	OH_2	J	28.9
	1/16×	`gamma_surf-110_04w`	$Al_{64}O_{112}H_{32}$	B	−44.9
GSH	−1/12×	`gamma_surf-110_01w`	$Al_{64}O_{100}H_8$	R	−22.1
	−1×	`water`	OH_2	J	21.7
	1/12×	`gamma_surf-110_04w`	$Al_{64}O_{112}H_{32}$	B	−23.7
GSH	−1/8×	`gamma_surf-110_02w`	$Al_{64}O_{104}H_{16}$	R	−29.1
	−1×	`water`	OH_2	J	50.5
	1/8×	`gamma_surf-110_04w`	$Al_{64}O_{112}H_{32}$	B	−28.1
GSH	−1/4×	`gamma_surf-110_03w`	$Al_{64}O_{108}H_{24}$	R	−31.0
	−1×	`water`	OH_2	J	−11.3
	1/4×	`gamma_surf-110_04w`	$Al_{64}O_{112}H_{32}$	B	−29.5
GSH	−1/20×	`gamma_surf-110_00w`	$Al_{64}O_{96}$	R	−35.2
	−1×	`water`	OH_2	J	16.0
	1/20×	`gamma_surf-110_05w`	$Al_{64}O_{116}H_{40}$	B	−39.4
GSH	−1/16×	`gamma_surf-110_01w`	$Al_{64}O_{100}H_8$	R	−20.9
	−1×	`water`	OH_2	J	7.4
	1/16×	`gamma_surf-110_05w`	$Al_{64}O_{116}H_{40}$	B	−22.1
GSH	−1/12×	`gamma_surf-110_02w`	$Al_{64}O_{104}H_{16}$	R	−25.1
	−1×	`water`	OH_2	J	21.9
	1/12×	`gamma_surf-110_05w`	$Al_{64}O_{116}H_{40}$	B	−24.5
GSH	−1/8×	`gamma_surf-110_03w`	$Al_{64}O_{108}H_{24}$	R	−24.1
	−1×	`water`	OH_2	J	−23.4
	1/8×	`gamma_surf-110_05w`	$Al_{64}O_{116}H_{40}$	B	−23.4
GSH	−1/4×	`gamma_surf-110_04w`	$Al_{64}O_{112}H_{32}$	R	−17.3
	−1×	`water`	OH_2	J	−35.4
	1/4×	`gamma_surf-110_05w`	$Al_{64}O_{116}H_{40}$	B	−17.2
GSH	−1/24×	`gamma_surf-110_00w`	$Al_{64}O_{96}$	R	−33.5
	−1×	`water`	OH_2	J	21.5
	1/24×	`gamma_surf-110_06w`	$Al_{64}O_{120}H_{48}$	B	−37.5
GSH	−1/20×	`gamma_surf-110_01w`	$Al_{64}O_{100}H_8$	R	−21.7
	−1×	`water`	OH_2	J	15.8
	1/20×	`gamma_surf-110_06w`	$Al_{64}O_{120}H_{48}$	B	−23.3
GSH	−1/16×	`gamma_surf-110_02w`	$Al_{64}O_{104}H_{16}$	R	−25.1
	−1×	`water`	OH_2	J	28.7
	1/16×	`gamma_surf-110_06w`	$Al_{64}O_{120}H_{48}$	B	−25.3
GSH	−1/12×	`gamma_surf-110_03w`	$Al_{64}O_{108}H_{24}$	R	−24.4
	−1×	`water`	OH_2	J	0.8
	1/12×	`gamma_surf-110_06w`	$Al_{64}O_{120}H_{48}$	B	−24.9
GSH	−1/8×	`gamma_surf-110_04w`	$Al_{64}O_{112}H_{32}$	R	−21.1
	−1×	`water`	OH_2	J	6.8

	1/8×	`gamma_surf-110_06w`	$Al_{64}O_{120}H_{48}$	B	−22.6
GSH	−1/4×	`gamma_surf-110_05w`	$Al_{64}O_{116}H_{40}$	R	−24.9
	−1×	`water`	OH_2	J	49.1
	1/4×	`gamma_surf-110_06w`	$Al_{64}O_{120}H_{48}$	B	−27.9
GSH	−1/10×	`gamma_surf-111_00w`	$Al_{40}O_{60}$	R	−42.1
	−1×	`water`	OH_2	J	7.2
	1/8×	`gamma_surf-111_04w`	$Al_{32}O_{56}H_{16}$	B	−31.1
GSH	−2/25×	`gamma_surf-111_00w`	$Al_{40}O_{60}$	R	−45.8
	−1×	`water`	OH_2	J	5.6
	1/10×	`gamma_surf-111_05w`	$Al_{32}O_{58}H_{20}$	B	−32.8
GSH	−1/2×	`gamma_surf-111_04w`	$Al_{32}O_{56}H_{16}$	R	−60.5
	−1×	`water`	OH_2	J	−0.7
	1/2×	`gamma_surf-111_05w`	$Al_{32}O_{58}H_{20}$	B	−39.8
GSH	−1/15×	`gamma_surf-111_00w`	$Al_{40}O_{60}$	R	−43.1
	−1×	`water`	OH_2	J	13.5
	1/12×	`gamma_surf-111_06w`	$Al_{32}O_{60}H_{24}$	B	−32.5
GSH	−1/4×	`gamma_surf-111_04w`	$Al_{32}O_{56}H_{16}$	R	−45.0
	−1×	`water`	OH_2	J	26.2
	1/4×	`gamma_surf-111_06w`	$Al_{32}O_{60}H_{24}$	B	−35.3
GSH	−1/2×	`gamma_surf-111_05w`	$Al_{32}O_{58}H_{20}$	R	−29.4
	−1×	`water`	OH_2	J	53.2
	1/2×	`gamma_surf-111_06w`	$Al_{32}O_{60}H_{24}$	B	−30.9
SUR	−1/32×	`boehm_bulk`	$Al_{32}O_{64}H_{32}$	R	11.2
	1/48×	`boehm_surf-001_00w`	$Al_{48}O_{96}H_{48}$	J	5.5
				B	9.0
SUR	−1/32×	`boehm_bulk`	$Al_{32}O_{64}H_{32}$	R	2.9
	1/72×	`boehm_surf-010_00w`	$Al_{72}O_{144}H_{72}$	J	2.1
				B	3.1
SUR	−1/32×	`boehm_bulk`	$Al_{32}O_{64}H_{32}$	R	23.0
	1/48×	`boehm_surf-100_00w`	$Al_{48}O_{96}H_{48}$	J	−0.4
				B	19.1
SUR	−1/32×	`boehm_bulk`	$Al_{32}O_{64}H_{32}$	R	18.6
	1/48×	`boehm_surf-101_00w`	$Al_{48}O_{96}H_{48}$	J	7.5
				B	16.4
SUR	−1/16×	`gamma_bulk`	$Al_{16}O_{24}$	R	13.5
	1/96×	`gamma_edge-100-110_00w`	$Al_{96}O_{144}$	J	7.0
				B	14.0
SUR	−1/16×	`gamma_bulk`	$Al_{16}O_{24}$	R	10.4
	1/64×	`gamma_surf-100_00w`	$Al_{64}O_{96}$	J	6.6
				B	13.0
SUR	−1/16×	`gamma_bulk`	$Al_{16}O_{24}$	R	21.2
	1/64×	`gamma_surf-110_00w`	$Al_{64}O_{96}$	J	7.4
				B	26.6
SUR	−1/16×	`gamma_bulk`	$Al_{16}O_{24}$	R	13.5

	$1/40\times$	`gamma_surf-111_00w`	$Al_{40}O_{60}$	J	5.6
				B	11.1
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−27.8
	$1/12\times$	`alpha_bulk`	$Al_{12}O_{18}$	J	−44.9
	$5/2\times$	`water`	OH_2	B	−31.9
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−38.8
	$1/32\times$	`boehm_bulk`	$Al_{32}O_{64}H_{32}$	J	−38.6
	$2\times$	`water`	OH_2	B	−36.5
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−27.6
	$1/48\times$	`boehm_surf-001_00w`	$Al_{48}O_{96}H_{48}$	J	−33.1
	$2\times$	`water`	OH_2	B	−27.4
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−35.9
	$1/72\times$	`boehm_surf-010_00w`	$Al_{72}O_{144}H_{72}$	J	−36.5
	$2\times$	`water`	OH_2	B	−33.4
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−15.8
	$1/48\times$	`boehm_surf-100_00w`	$Al_{48}O_{96}H_{48}$	J	−39.0
	$2\times$	`water`	OH_2	B	−17.4
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−20.2
	$1/48\times$	`boehm_surf-101_00w`	$Al_{48}O_{96}H_{48}$	J	−31.1
	$2\times$	`water`	OH_2	B	−20.1
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−25.3
	$1/16\times$	`gamma_bulk`	$Al_{16}O_{24}$	J	−48.6
	$5/2\times$	`water`	OH_2	B	−26.4
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−11.7
	$1/96\times$	`gamma_edge-100-110_00w`	$Al_{96}O_{144}$	J	−41.6
	$5/2\times$	`water`	OH_2	B	−12.4
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−14.8
	$1/64\times$	`gamma_surf-100_00w`	$Al_{64}O_{96}$	J	−42.0
	$5/2\times$	`water`	OH_2	B	−13.4
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−4.1
	$1/64\times$	`gamma_surf-110_00w`	$Al_{64}O_{96}$	J	−41.2
	$5/2\times$	`water`	OH_2	B	0.2
FOR	$-1\times$	`monomer`	AlO_4H_5	R	−11.7
	$1/40\times$	`gamma_surf-111_00w`	$Al_{40}O_{60}$	J	−43.0
	$5/2\times$	`water`	OH_2	B	−15.3

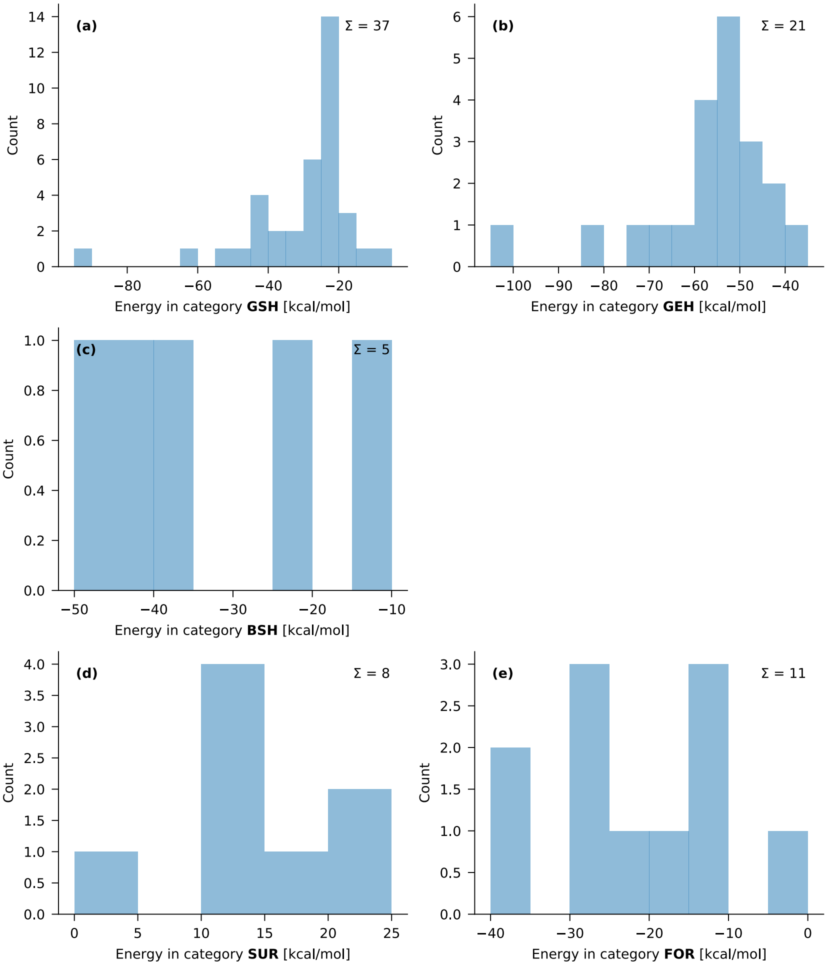

Figure S4: Histograms of reaction energies in the training set.

Table S3: List of activated parameters and their corresponding block, initial values from 1 and bounds.

Name	Unit	Atoms	Block	Joshi 2014	Lower bound	Upper bound	This work
p_val3	1	Al	ATM	1.5000	1.2000	3.0000	2.9993
p_val5	1	Al	ATM	2.5791	2.0633	3.0949	2.4980
D_e^sigma	$\mathrm{kcal\,mol^{-1}}$	Al-H	BND	92.8579	0.8579	122.7844	77.1779
p_be1	1	Al-H	BND	−0.6528	−0.7834	1.0000	−0.7823
p_ovun1	1	Al-H	BND	0.1551	0.0100	0.5000	0.3089
p_be2	1	Al-H	BND	10.0663	0.2281	13.0000	2.0985
p_bo1	1	Al-H	BND	−0.0842	−0.3320	−0.0674	−0.1398
p_bo2	1	Al-H	BND	7.1758	5.0015	15.0000	11.7930
D_e^sigma	$\mathrm{kcal\,mol^{-1}}$	Al-O	BND	182.0654	118.9203	232.7313	166.4329
p_be1	1	Al-O	BND	−0.0920	−1.0000	−0.0736	−0.0963
p_ovun1	1	Al-O	BND	0.1688	0.0100	0.4562	0.1457
p_be2	1	Al-O	BND	0.0010	0.0008	1.5477	0.7025
p_bo1	1	Al-O	BND	−0.1959	−0.2351	−0.0740	−0.2053
p_bo2	1	Al-O	BND	6.0894	4.6533	7.3073	7.1635
D_e^sigma	$\mathrm{kcal\,mol^{-1}}$	Al-Al	BND	34.0777	27.2622	65.7742	31.9434
r_0^sigma	Å	Al-H	OFD	1.7276	1.3821	2.0731	1.4893
r_0^sigma	Å	Al-O	OFD	1.5646	1.2517	1.8775	1.6172
p_val1	1	Al-H-O	ANG	4.2750	3.4200	20.0000	19.7283
p_val2	1	Al-H-O	ANG	1.0250	0.8200	4.8339	3.9208
p_val4	1	Al-H-O	ANG	1.4750	1.0100	1.7700	1.0931
Theta_0,0	deg	Al-O-H	ANG	88.6163	64.6197	106.3396	66.0975
p_val1	1	Al-O-H	ANG	10.1310	4.2037	19.7491	16.7556
p_val2	1	Al-O-H	ANG	1.6896	1.3517	10.0000	2.4332
p_val4	1	Al-O-H	ANG	1.0000	0.8000	3.0000	1.4778
Theta_0,0	deg	Al-O-Al	ANG	13.8580	5.2474	64.5513	64.5486
p_val1	1	Al-O-Al	ANG	12.3669	9.8935	40.0000	12.8201
p_val2	1	Al-O-Al	ANG	4.4355	0.5527	9.9945	7.0734
p_val4	1	Al-O-Al	ANG	1.1908	0.9526	3.0000	2.9786
Theta_0,0	deg	H-Al-O	ANG	41.8108	0.0000	64.8437	9.7233
p_val1	1	H-Al-O	ANG	17.3800	5.4547	30.9495	23.4162
p_val2	1	H-Al-O	ANG	2.6618	0.9702	3.1942	3.0390
p_val4	1	H-Al-O	ANG	1.0100	0.8080	3.0000	1.0218
Theta_0,0	deg	O-Al-O	ANG	55.4358	43.7395	84.7469	61.5592
p_val1	1	O-Al-O	ANG	22.1089	7.3926	40.0000	24.2337
p_val2	1	O-Al-O	ANG	3.7402	1.2450	4.4882	1.2622
p_val4	1	O-Al-O	ANG	2.2064	1.0123	3.0000	2.9975

Table S4: Structures in the validation set and contributions to each category of data. Bonds are represented by pairs of chemical elements, angles by triplets. The symbol ··· denotes a hydrogen bond. Internal coordinates with oxygen not bound to aluminum are discarded. For hydrated alumina surfaces, the internal coordinates without hydrogen are not considered.

Structure	Chem. Form.	O−H	O···H	Al−O	Al−Al	O−Al−O	Al−O−Al	Al−O−H	Al−O···H	H−O···H	H···O···H
`gamma_bulk`	$Al_{32}O_{48}$			176	176	408	240				
`gamma_surf-001_00w`	$Al_{64}O_{96}$			336	316	736	432				
`gamma_surf-001_01w`	$Al_{64}O_{98}H_{4}$	4	2		316			8	2	2	
`gamma_surf-001_02w`	$Al_{64}O_{100}H_{8}$	8	4		318			12	4	4	2
`gamma_surf-001_03w`	$Al_{64}O_{102}H_{12}$	12	8		316			20	8	8	4
`gamma_surf-001_04w`	$Al_{64}O_{104}H_{16}$	16	10		316			26	14	10	4
`gamma_surf-110b_from_boehm_00w`	$Al_{80}O_{120}$			414	416	902	532				
`gamma_surf-110b_from_boehm_02w`	$Al_{80}O_{124}H_{8}$	8			436			16			
`gamma_surf-110b_from_boehm_03w`	$Al_{80}O_{126}H_{12}$	12			436			22			
`gamma_surf-110b_from_boehm_04w`	$Al_{80}O_{128}H_{16}$	16	8		436			24	8	8	
`gamma_surf-110b_from_bulk_00w`	$Al_{64}O_{96}$			336	324	744	440				
`gamma_surf-110b_from_bulk_02w`	$Al_{64}O_{98}H_{4}$	4	2		324			6	4		
`gamma_surf-110b_from_bulk_04w`	$Al_{64}O_{100}H_{8}$	8	4		324			12	4	4	
`gamma_surf-110b_from_bulk_06w`	$Al_{64}O_{102}H_{12}$	12	6		324			16	10	4	
`gamma_surf-110b_from_bulk_08w`	$Al_{64}O_{104}H_{16}$	16	10		324			20	18	6	2
`gamma_surf-110l_A1_00w`	$Al_{48}O_{72}$			250	238	552	324				
`gamma_surf-110l_A1_01w`	$Al_{48}O_{74}H_{4}$	4	2		234			8	4		
`gamma_surf-110l_A1_02w`	$Al_{48}O_{76}H_{8}$	8	4		240			14	6	2	
`gamma_surf-110l_A1_03w`	$Al_{48}O_{78}H_{12}$	12	4		240			22	6	4	
`gamma_surf-110l_A1_04w`	$Al_{48}O_{80}H_{16}$	16	12		236			26	22	6	4
`gamma_surf-110l_A1_05w`	$Al_{48}O_{82}H_{20}$	20	14		236			30	26	8	4
`gamma_surf-110l_A1_06w`	$Al_{48}O_{84}H_{24}$	24	17		236			30	37	4	9
`gamma_surf-110l_A2_00w`	$Al_{56}O_{84}$			296	276	660	392				
`gamma_surf-110l_A2_01w`	$Al_{56}O_{86}H_{4}$	4	2		274			8	4		
`gamma_surf-110l_A2_02w`	$Al_{56}O_{88}H_{8}$	8	4		282			14	8		2
`gamma_surf-110l_A2_03w`	$Al_{56}O_{90}H_{12}$	12	7		284			20	11	5	1
`gamma_surf-110l_A2_04w`	$Al_{56}O_{92}H_{16}$	16	10		284			24	18	8	
`gamma_surf-110l_A2_05w`	$Al_{56}O_{94}H_{20}$	20	14		282			26	30	4	6
`gamma_surf-110l_A2_06w`	$Al_{56}O_{96}H_{24}$	24	18		282			30	38	6	8
`gamma_surf-110l_L1_04w`	$Al_{48}O_{80}H_{16}$	16	6		240			32	12	4	
`gamma_surf-110l_L2_00w`	$Al_{56}O_{84}$			294	286	654	384				
`gamma_surf-110l_L2_01w`	$Al_{56}O_{86}H_{4}$	4	2		276			8	4		
`gamma_surf-110l_L2_02w`	$Al_{56}O_{88}H_{8}$	8	4		282			14	8		2
`gamma_surf-110l_L2_03w`	$Al_{56}O_{90}H_{12}$	12	4		282			22	10	2	
`gamma_surf-110l_L2_04w`	$Al_{56}O_{92}H_{16}$	16	6		282			32	16	2	
`gamma_surf-111_D1_03w`	$Al_{32}O_{54}H_{12}$	12	5		144			22	6	4	2
`gamma_surf-111_D1_04w`	$Al_{32}O_{56}H_{16}$	16	5		146			30	6	5	2
`gamma_surf-111_D1_05w`	$Al_{32}O_{58}H_{20}$	20	10		146			36	14	10	2
`gamma_surf-111_D1_06w`	$Al_{32}O_{60}H_{24}$	24	14		146			44	17	14	5
`gamma_surf-111_D2_03w`	$Al_{32}O_{54}H_{12}$	12	8		144			22	12	4	2
`gamma_surf-111_D2_04w`	$Al_{32}O_{56}H_{16}$	16	9		146			30	14	7	2

`gamma_surf-111_D2_05w`	$Al_{32}O_{58}H_{20}$	20	12		146			36	15	12	6
`gamma_surf-111_D2_06w`	$Al_{32}O_{60}H_{24}$	24	15		146			44	18	15	6
`gamma_surf-111_P1_1_05w`	$Al_{40}O_{70}H_{20}$	20	3		184			43	6	3	
`gamma_surf-111_P1_1_06w`	$Al_{40}O_{72}H_{24}$	24	4		178			54	8	4	
`gamma_surf-111_P1_2_03w`	$Al_{40}O_{66}H_{12}$	12	2		178			24	4	2	
`gamma_surf-111_P1_2_04w`	$Al_{40}O_{68}H_{16}$	16	4		178			34	8	4	
`gamma_surf-111_P1_2_05w`	$Al_{40}O_{70}H_{20}$	20	5		178			44	10	5	
`gamma_surf-111_P1_2_06w`	$Al_{40}O_{72}H_{24}$	24	2		182			54	4	2	
`gamma_surf-111_P2_1_05w`	$Al_{40}O_{70}H_{20}$	20	14		196			27	15	13	4
`gamma_surf-111_P2_1_06w`	$Al_{40}O_{72}H_{24}$	24	16		200			38	16	16	4
`gamma_surf-111_P2_2_04w`	$Al_{40}O_{68}H_{16}$	16	10		194			26	16	6	
`gamma_surf-111_P2_2_05w`	$Al_{40}O_{70}H_{20}$	20	14		190			32	18	12	3
`gamma_surf-111_P2_2_06w`	$Al_{40}O_{72}H_{24}$	24	15		192			35	15	15	2
`water`	OH_2										
`total`		704	351	2102	13 588	4656	2744	1217	554	254	88

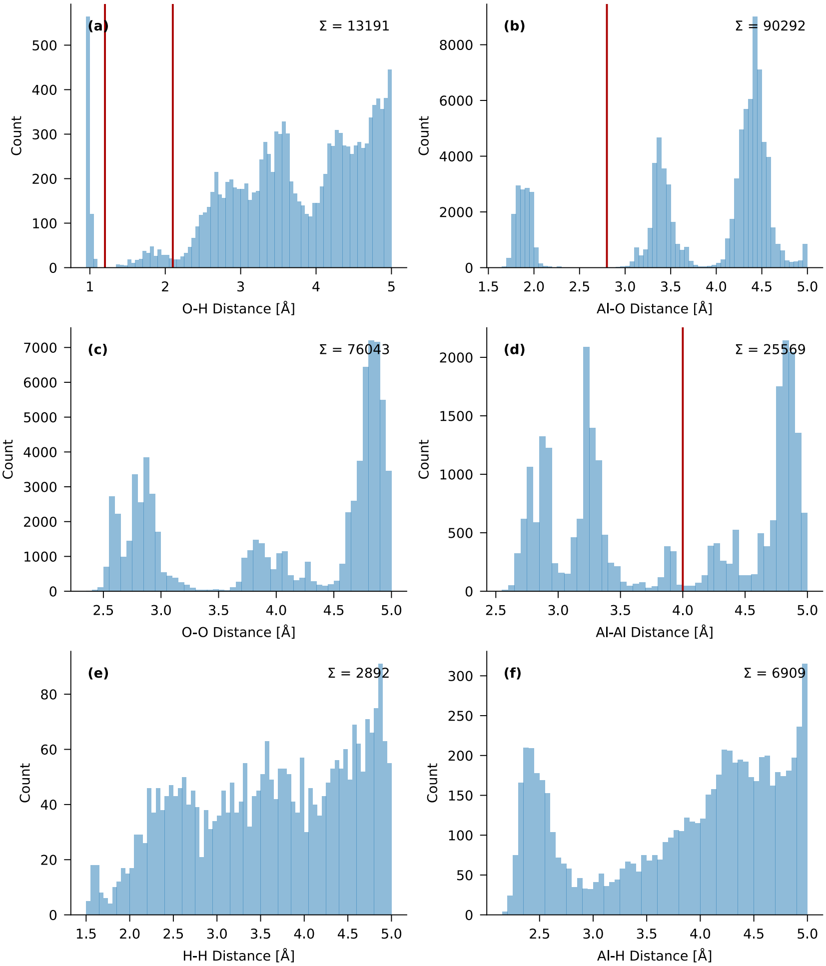

Figure S5: Histograms of all interatomic distances in the validation set up to 5 Å, grouped per pair of chemical elements. Cutoffs for OH and AlO pairs depicted as vertical red lines: 1.2 Å for O−H bonds, 2.1 Å for hydrogen bonds and 2.8 Å for Al−O bonds. See main text for a more detailed description.

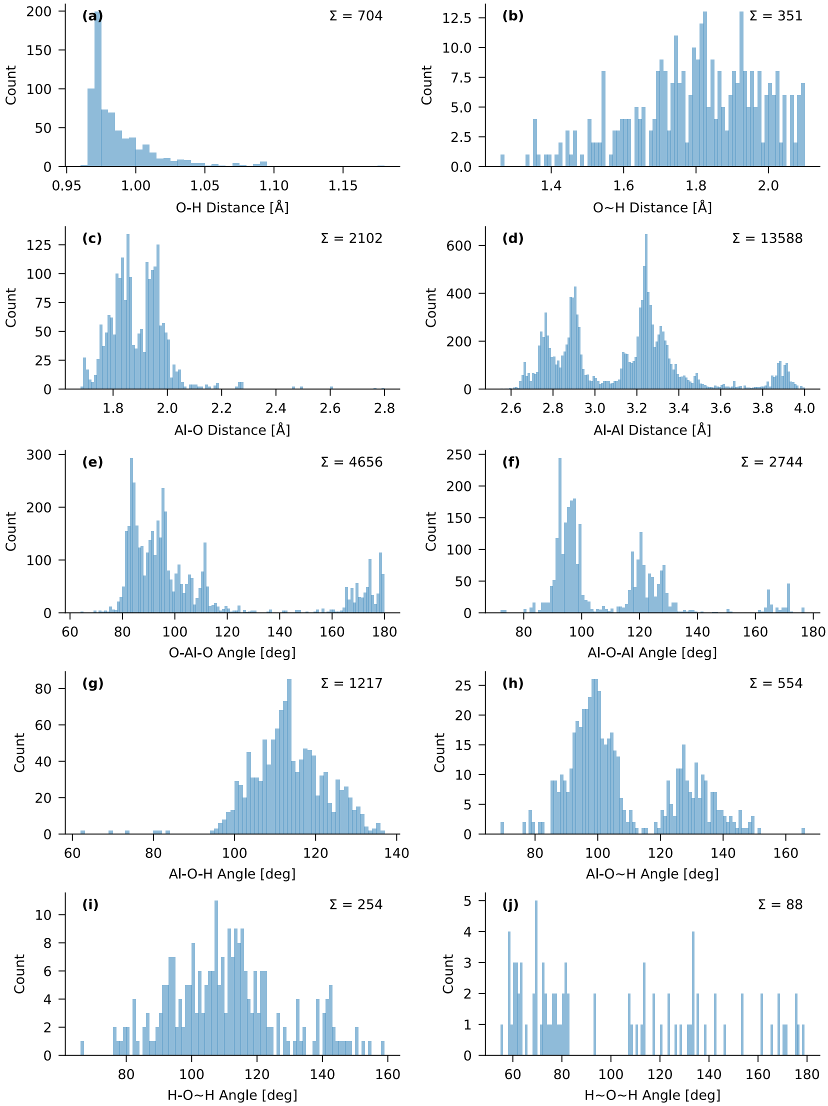

Figure S6: Histograms of internal coordinates in the validation set. In the labels on the horizontal axis, a dash represents a regular bond and a tilde represents a hydrogen bond.

Table S5: Overview of all chemical equations in the validation set. Reactants are given negative coefficients. For each reaction, three reaction energies are in $\mathrm{kcal\,mol^{-1}}$: the reference DFT result (R), the prediction with the Joshi force field (J) and the prediction with the new force field in this work (B). Water adsorption energies are normalized to the number of water molecules. All other reaction energies are normalized on the number of Al atoms. The categories are defined in the main text.

Category	Coeff.	Structure	Chem. Form.	Reaction energy	
GSH	$-1/2\times$	`gamma_surf-001_00w`	$Al_{64}\,O_{96}$	R	-22.5
	$-1\times$	`water`	$O\,H_2$	J	-14.3
	$1/2\times$	`gamma_surf-001_01w`	$Al_{64}\,O_{98}\,H_4$	B	-31.6
GSH	$-1/4\times$	`gamma_surf-001_00w`	$Al_{64}\,O_{96}$	R	-24.4
	$-1\times$	`water`	$O\,H_2$	J	-5.0
	$1/4\times$	`gamma_surf-001_02w`	$Al_{64}\,O_{100}\,H_8$	B	-27.5
GSH	$-1/2\times$	`gamma_surf-001_01w`	$Al_{64}\,O_{98}\,H_4$	R	-26.3
	$-1\times$	`water`	$O\,H_2$	J	4.2
	$1/2\times$	`gamma_surf-001_02w`	$Al_{64}\,O_{100}\,H_8$	B	-23.4
GSH	$-1/6\times$	`gamma_surf-001_00w`	$Al_{64}\,O_{96}$	R	-23.5
	$-1\times$	`water`	$O\,H_2$	J	0.2
	$1/6\times$	`gamma_surf-001_03w`	$Al_{64}\,O_{102}\,H_{12}$	B	-28.4
GSH	$-1/4\times$	`gamma_surf-001_01w`	$Al_{64}\,O_{98}\,H_4$	R	-24.0
	$-1\times$	`water`	$O\,H_2$	J	7.4
	$1/4\times$	`gamma_surf-001_03w`	$Al_{64}\,O_{102}\,H_{12}$	B	-26.8
GSH	$-1/2\times$	`gamma_surf-001_02w`	$Al_{64}\,O_{100}\,H_8$	R	-21.6
	$-1\times$	`water`	$O\,H_2$	J	10.6
	$1/2\times$	`gamma_surf-001_03w`	$Al_{64}\,O_{102}\,H_{12}$	B	-30.1
GSH	$-1/8\times$	`gamma_surf-001_00w`	$Al_{64}\,O_{96}$	R	-20.8
	$-1\times$	`water`	$O\,H_2$	J	-12.3
	$1/8\times$	`gamma_surf-001_04w`	$Al_{64}\,O_{104}\,H_{16}$	B	-24.1
GSH	$-1/6\times$	`gamma_surf-001_01w`	$Al_{64}\,O_{98}\,H_4$	R	-20.2
	$-1\times$	`water`	$O\,H_2$	J	-11.6
	$1/6\times$	`gamma_surf-001_04w`	$Al_{64}\,O_{104}\,H_{16}$	B	-21.6
GSH	$-1/4\times$	`gamma_surf-001_02w`	$Al_{64}\,O_{100}\,H_8$	R	-17.1
	$-1\times$	`water`	$O\,H_2$	J	-19.5
	$1/4\times$	`gamma_surf-001_04w`	$Al_{64}\,O_{104}\,H_{16}$	B	-20.7
GSH	$-1/2\times$	`gamma_surf-001_03w`	$Al_{64}\,O_{102}\,H_{12}$	R	-12.6
	$-1\times$	`water`	$O\,H_2$	J	-49.7
	$1/2\times$	`gamma_surf-001_04w`	$Al_{64}\,O_{104}\,H_{16}$	B	-11.3
GSH	$-1/4\times$	`gamma_surf-110b_from_boehm_00w`	$Al_{80}\,O_{120}$	R	-60.7
	$-1\times$	`water`	$O\,H_2$	J	5.5
	$1/4\times$	`gamma_surf-110b_from_boehm_02w`	$Al_{80}\,O_{124}\,H_8$	B	-50.1
GSH	$-1/6\times$	`gamma_surf-110b_from_boehm_00w`	$Al_{80}\,O_{120}$	R	-58.9
	$-1\times$	`water`	$O\,H_2$	J	10.3

	1/6×	`gamma_surf-110b_from_boehm_03w`	$Al_{80}O_{126}H_{12}$	B	−56.7
GSH	−1/2×	`gamma_surf-110b_from_boehm_02w`	$Al_{80}O_{124}H_{8}$	R	−55.1
	−1×	`water`	OH_{2}	J	19.7
	1/2×	`gamma_surf-110b_from_boehm_03w`	$Al_{80}O_{126}H_{12}$	B	−70.0
GSH	−1/8×	`gamma_surf-110b_from_boehm_00w`	$Al_{80}O_{120}$	R	−58.2
	−1×	`water`	OH_{2}	J	−7.7
	1/8×	`gamma_surf-110b_from_boehm_04w`	$Al_{80}O_{128}H_{16}$	B	−53.6
GSH	−1/4×	`gamma_surf-110b_from_boehm_02w`	$Al_{80}O_{124}H_{8}$	R	−55.7
	−1×	`water`	OH_{2}	J	−21.0
	1/4×	`gamma_surf-110b_from_boehm_04w`	$Al_{80}O_{128}H_{16}$	B	−57.2
GSH	−1/2×	`gamma_surf-110b_from_boehm_03w`	$Al_{80}O_{126}H_{12}$	R	−56.3
	−1×	`water`	OH_{2}	J	−61.7
	1/2×	`gamma_surf-110b_from_boehm_04w`	$Al_{80}O_{128}H_{16}$	B	−44.4
GSH	−1/2×	`gamma_surf-110b_from_bulk_00w`	$Al_{64}O_{96}$	R	−43.6
	−1×	`water`	OH_{2}	J	−26.1
	1/2×	`gamma_surf-110b_from_bulk_02w`	$Al_{64}O_{98}H_{4}$	B	−49.6
GSH	−1/4×	`gamma_surf-110b_from_bulk_00w`	$Al_{64}O_{96}$	R	−45.8
	−1×	`water`	OH_{2}	J	−43.6
	1/4×	`gamma_surf-110b_from_bulk_04w`	$Al_{64}O_{100}H_{8}$	B	−50.9
GSH	−1/2×	`gamma_surf-110b_from_bulk_02w`	$Al_{64}O_{98}H_{4}$	R	−48.0
	−1×	`water`	OH_{2}	J	−61.0
	1/2×	`gamma_surf-110b_from_bulk_04w`	$Al_{64}O_{100}H_{8}$	B	−52.1
GSH	−1/6×	`gamma_surf-110b_from_bulk_00w`	$Al_{64}O_{96}$	R	−39.1
	−1×	`water`	OH_{2}	J	−32.6
	1/6×	`gamma_surf-110b_from_bulk_06w`	$Al_{64}O_{102}H_{12}$	B	−41.3
GSH	−1/4×	`gamma_surf-110b_from_bulk_02w`	$Al_{64}O_{98}H_{4}$	R	−36.8
	−1×	`water`	OH_{2}	J	−35.9
	1/4×	`gamma_surf-110b_from_bulk_06w`	$Al_{64}O_{102}H_{12}$	B	−37.2
GSH	−1/2×	`gamma_surf-110b_from_bulk_04w`	$Al_{64}O_{100}H_{8}$	R	−25.6
	−1×	`water`	OH_{2}	J	−10.7
	1/2×	`gamma_surf-110b_from_bulk_06w`	$Al_{64}O_{102}H_{12}$	B	−22.2
GSH	−1/8×	`gamma_surf-110b_from_bulk_00w`	$Al_{64}O_{96}$	R	−36.4
	−1×	`water`	OH_{2}	J	−26.9
	1/8×	`gamma_surf-110b_from_bulk_08w`	$Al_{64}O_{104}H_{16}$	B	−38.1
GSH	−1/6×	`gamma_surf-110b_from_bulk_02w`	$Al_{64}O_{98}H_{4}$	R	−34.0
	−1×	`water`	OH_{2}	J	−27.1
	1/6×	`gamma_surf-110b_from_bulk_08w`	$Al_{64}O_{104}H_{16}$	B	−34.3
GSH	−1/4×	`gamma_surf-110b_from_bulk_04w`	$Al_{64}O_{100}H_{8}$	R	−26.9
	−1×	`water`	OH_{2}	J	−10.2
	1/4×	`gamma_surf-110b_from_bulk_08w`	$Al_{64}O_{104}H_{16}$	B	−25.4
GSH	−1/2×	`gamma_surf-110b_from_bulk_06w`	$Al_{64}O_{102}H_{12}$	R	−28.2
	−1×	`water`	OH_{2}	J	−9.7
	1/2×	`gamma_surf-110b_from_bulk_08w`	$Al_{64}O_{104}H_{16}$	B	−28.6
GSH	−1/2×	`gamma_surf-110l_A1_00w`	$Al_{48}O_{72}$	R	−86.5

	−1×	`water`	$O\,H_2$	J	61.0
	1/2×	`gamma_surf-110l_A1_01w`	$Al_{48}\,O_{74}\,H_4$	B	−83.3
GSH	−1/4×	`gamma_surf-110l_A1_00w`	$Al_{48}\,O_{72}$	R	−58.5
	−1×	`water`	$O\,H_2$	J	4.1
	1/4×	`gamma_surf-110l_A1_02w`	$Al_{48}\,O_{76}\,H_8$	B	−65.9
GSH	−1/2×	`gamma_surf-110l_A1_01w`	$Al_{48}\,O_{74}\,H_4$	R	−30.5
	−1×	`water`	$O\,H_2$	J	−52.8
	1/2×	`gamma_surf-110l_A1_02w`	$Al_{48}\,O_{76}\,H_8$	B	−48.5
GSH	−1/6×	`gamma_surf-110l_A1_00w`	$Al_{48}\,O_{72}$	R	−65.3
	−1×	`water`	$O\,H_2$	J	21.2
	1/6×	`gamma_surf-110l_A1_03w`	$Al_{48}\,O_{78}\,H_{12}$	B	−67.2
GSH	−1/4×	`gamma_surf-110l_A1_01w`	$Al_{48}\,O_{74}\,H_4$	R	−54.6
	−1×	`water`	$O\,H_2$	J	1.4
	1/4×	`gamma_surf-110l_A1_03w`	$Al_{48}\,O_{78}\,H_{12}$	B	−59.2
GSH	−1/2×	`gamma_surf-110l_A1_02w`	$Al_{48}\,O_{76}\,H_8$	R	−78.8
	−1×	`water`	$O\,H_2$	J	55.5
	1/2×	`gamma_surf-110l_A1_03w`	$Al_{48}\,O_{78}\,H_{12}$	B	−69.8
GSH	−1/8×	`gamma_surf-110l_A1_00w`	$Al_{48}\,O_{72}$	R	−57.0
	−1×	`water`	$O\,H_2$	J	15.1
	1/8×	`gamma_surf-110l_A1_04w`	$Al_{48}\,O_{80}\,H_{16}$	B	−55.8
GSH	−1/6×	`gamma_surf-110l_A1_01w`	$Al_{48}\,O_{74}\,H_4$	R	−47.1
	−1×	`water`	$O\,H_2$	J	−0.2
	1/6×	`gamma_surf-110l_A1_04w`	$Al_{48}\,O_{80}\,H_{16}$	B	−46.6
GSH	−1/4×	`gamma_surf-110l_A1_02w`	$Al_{48}\,O_{76}\,H_8$	R	−55.4
	−1×	`water`	$O\,H_2$	J	26.1
	1/4×	`gamma_surf-110l_A1_04w`	$Al_{48}\,O_{80}\,H_{16}$	B	−45.6
GSH	−1/2×	`gamma_surf-110l_A1_03w`	$Al_{48}\,O_{78}\,H_{12}$	R	−32.0
	−1×	`water`	$O\,H_2$	J	−3.4
	1/2×	`gamma_surf-110l_A1_04w`	$Al_{48}\,O_{80}\,H_{16}$	B	−21.4
GSH	−1/10×	`gamma_surf-110l_A1_00w`	$Al_{48}\,O_{72}$	R	−49.8
	−1×	`water`	$O\,H_2$	J	5.5
	1/10×	`gamma_surf-110l_A1_05w`	$Al_{48}\,O_{82}\,H_{20}$	B	−48.5
GSH	−1/8×	`gamma_surf-110l_A1_01w`	$Al_{48}\,O_{74}\,H_4$	R	−40.7
	−1×	`water`	$O\,H_2$	J	−8.4
	1/8×	`gamma_surf-110l_A1_05w`	$Al_{48}\,O_{82}\,H_{20}$	B	−39.8
GSH	−1/6×	`gamma_surf-110l_A1_02w`	$Al_{48}\,O_{76}\,H_8$	R	−44.1
	−1×	`water`	$O\,H_2$	J	6.4
	1/6×	`gamma_surf-110l_A1_05w`	$Al_{48}\,O_{82}\,H_{20}$	B	−36.9
GSH	−1/4×	`gamma_surf-110l_A1_03w`	$Al_{48}\,O_{78}\,H_{12}$	R	−26.7
	−1×	`water`	$O\,H_2$	J	−18.1
	1/4×	`gamma_surf-110l_A1_05w`	$Al_{48}\,O_{82}\,H_{20}$	B	−20.4
GSH	−1/2×	`gamma_surf-110l_A1_04w`	$Al_{48}\,O_{80}\,H_{16}$	R	−21.4
	−1×	`water`	$O\,H_2$	J	−32.8
	1/2×	`gamma_surf-110l_A1_05w`	$Al_{48}\,O_{82}\,H_{20}$	B	−19.4

GSH	−1/12×	`gamma_surf-110l_A1_00w`	$Al_{48}O_{72}$	R	−43.6
	−1×	`water`	OH_2	J	−1.2
	1/12×	`gamma_surf-110l_A1_06w`	$Al_{48}O_{84}H_{24}$	B	−42.5
GSH	−1/10×	`gamma_surf-110l_A1_01w`	$Al_{48}O_{74}H_4$	R	−35.0
	−1×	`water`	OH_2	J	−13.6
	1/10×	`gamma_surf-110l_A1_06w`	$Al_{48}O_{84}H_{24}$	B	−34.4
GSH	−1/8×	`gamma_surf-110l_A1_02w`	$Al_{48}O_{76}H_8$	R	−36.2
	−1×	`water`	OH_2	J	−3.8
	1/8×	`gamma_surf-110l_A1_06w`	$Al_{48}O_{84}H_{24}$	B	−30.8
GSH	−1/6×	`gamma_surf-110l_A1_03w`	$Al_{48}O_{78}H_{12}$	R	−22.0
	−1×	`water`	OH_2	J	−23.6
	1/6×	`gamma_surf-110l_A1_06w`	$Al_{48}O_{84}H_{24}$	B	−17.8
GSH	−1/4×	`gamma_surf-110l_A1_04w`	$Al_{48}O_{80}H_{16}$	R	−17.0
	−1×	`water`	OH_2	J	−33.7
	1/4×	`gamma_surf-110l_A1_06w`	$Al_{48}O_{84}H_{24}$	B	−16.0
GSH	−1/2×	`gamma_surf-110l_A1_05w`	$Al_{48}O_{82}H_{20}$	R	−12.5
	−1×	`water`	OH_2	J	−34.6
	1/2×	`gamma_surf-110l_A1_06w`	$Al_{48}O_{84}H_{24}$	B	−12.7
GSH	−1/2×	`gamma_surf-110l_A2_00w`	$Al_{56}O_{84}$	R	−82.6
	−1×	`water`	OH_2	J	−40.3
	1/2×	`gamma_surf-110l_A2_01w`	$Al_{56}O_{86}H_4$	B	−89.7
GSH	−1/4×	`gamma_surf-110l_A2_00w`	$Al_{56}O_{84}$	R	−66.4
	−1×	`water`	OH_2	J	−41.8
	1/4×	`gamma_surf-110l_A2_02w`	$Al_{56}O_{88}H_8$	B	−75.6
GSH	−1/2×	`gamma_surf-110l_A2_01w`	$Al_{56}O_{86}H_4$	R	−50.2
	−1×	`water`	OH_2	J	−43.2
	1/2×	`gamma_surf-110l_A2_02w`	$Al_{56}O_{88}H_8$	B	−61.5
GSH	−1/6×	`gamma_surf-110l_A2_00w`	$Al_{56}O_{84}$	R	−60.3
	−1×	`water`	OH_2	J	−35.4
	1/6×	`gamma_surf-110l_A2_03w`	$Al_{56}O_{90}H_{12}$	B	−65.2
GSH	−1/4×	`gamma_surf-110l_A2_01w`	$Al_{56}O_{86}H_4$	R	−49.1
	−1×	`water`	OH_2	J	−32.9
	1/4×	`gamma_surf-110l_A2_03w`	$Al_{56}O_{90}H_{12}$	B	−52.9
GSH	−1/2×	`gamma_surf-110l_A2_02w`	$Al_{56}O_{88}H_8$	R	−48.1
	−1×	`water`	OH_2	J	−22.5
	1/2×	`gamma_surf-110l_A2_03w`	$Al_{56}O_{90}H_{12}$	B	−44.3
GSH	−1/8×	`gamma_surf-110l_A2_00w`	$Al_{56}O_{84}$	R	−52.7
	−1×	`water`	OH_2	J	−10.4
	1/8×	`gamma_surf-110l_A2_04w`	$Al_{56}O_{92}H_{16}$	B	−56.2
GSH	−1/6×	`gamma_surf-110l_A2_01w`	$Al_{56}O_{86}H_4$	R	−42.7
	−1×	`water`	OH_2	J	−0.4
	1/6×	`gamma_surf-110l_A2_04w`	$Al_{56}O_{92}H_{16}$	B	−45.1
GSH	−1/4×	`gamma_surf-110l_A2_02w`	$Al_{56}O_{88}H_8$	R	−38.9
	−1×	`water`	OH_2	J	21.0

	1/4×	`gamma_surf-110l_A2_04w`	$Al_{56}O_{92}H_{16}$	B	−36.8
GSH	−1/2×	`gamma_surf-110l_A2_03w`	$Al_{56}O_{90}H_{12}$	R	−29.8
	−1×	`water`	OH_2	J	64.5
	1/2×	`gamma_surf-110l_A2_04w`	$Al_{56}O_{92}H_{16}$	B	−29.3
GSH	−1/10×	`gamma_surf-110l_A2_00w`	$Al_{56}O_{84}$	R	−46.1
	−1×	`water`	OH_2	J	−5.3
	1/10×	`gamma_surf-110l_A2_05w`	$Al_{56}O_{94}H_{20}$	B	−50.1
GSH	−1/8×	`gamma_surf-110l_A2_01w`	$Al_{56}O_{86}H_4$	R	−37.0
	−1×	`water`	OH_2	J	3.5
	1/8×	`gamma_surf-110l_A2_05w`	$Al_{56}O_{94}H_{20}$	B	−40.2
GSH	−1/6×	`gamma_surf-110l_A2_02w`	$Al_{56}O_{88}H_8$	R	−32.6
	−1×	`water`	OH_2	J	19.0
	1/6×	`gamma_surf-110l_A2_05w`	$Al_{56}O_{94}H_{20}$	B	−33.0
GSH	−1/4×	`gamma_surf-110l_A2_03w`	$Al_{56}O_{90}H_{12}$	R	−24.9
	−1×	`water`	OH_2	J	39.8
	1/4×	`gamma_surf-110l_A2_05w`	$Al_{56}O_{94}H_{20}$	B	−27.4
GSH	−1/2×	`gamma_surf-110l_A2_04w`	$Al_{56}O_{92}H_{16}$	R	−20.0
	−1×	`water`	OH_2	J	15.2
	1/2×	`gamma_surf-110l_A2_05w`	$Al_{56}O_{94}H_{20}$	B	−25.5
GSH	−1/12×	`gamma_surf-110l_A2_00w`	$Al_{56}O_{84}$	R	−42.0
	−1×	`water`	OH_2	J	−17.4
	1/12×	`gamma_surf-110l_A2_06w`	$Al_{56}O_{96}H_{24}$	B	−47.1
GSH	−1/10×	`gamma_surf-110l_A2_01w`	$Al_{56}O_{86}H_4$	R	−33.9
	−1×	`water`	OH_2	J	−12.8
	1/10×	`gamma_surf-110l_A2_06w`	$Al_{56}O_{96}H_{24}$	B	−38.5
GSH	−1/8×	`gamma_surf-110l_A2_02w`	$Al_{56}O_{88}H_8$	R	−29.8
	−1×	`water`	OH_2	J	−5.1
	1/8×	`gamma_surf-110l_A2_06w`	$Al_{56}O_{96}H_{24}$	B	−32.8
GSH	−1/6×	`gamma_surf-110l_A2_03w`	$Al_{56}O_{90}H_{12}$	R	−23.7
	−1×	`water`	OH_2	J	0.6
	1/6×	`gamma_surf-110l_A2_06w`	$Al_{56}O_{96}H_{24}$	B	−29.0
GSH	−1/4×	`gamma_surf-110l_A2_04w`	$Al_{56}O_{92}H_{16}$	R	−20.7
	−1×	`water`	OH_2	J	−31.3
	1/4×	`gamma_surf-110l_A2_06w`	$Al_{56}O_{96}H_{24}$	B	−28.8
GSH	−1/2×	`gamma_surf-110l_A2_05w`	$Al_{56}O_{94}H_{20}$	R	−21.4
	−1×	`water`	OH_2	J	−77.7
	1/2×	`gamma_surf-110l_A2_06w`	$Al_{56}O_{96}H_{24}$	B	−32.1
GSH	−1/2×	`gamma_surf-110l_L2_00w`	$Al_{56}O_{84}$	R	−60.8
	−1×	`water`	OH_2	J	−15.8
	1/2×	`gamma_surf-110l_L2_01w`	$Al_{56}O_{86}H_4$	B	−58.3
GSH	−1/4×	`gamma_surf-110l_L2_00w`	$Al_{56}O_{84}$	R	−44.6
	−1×	`water`	OH_2	J	10.5
	1/4×	`gamma_surf-110l_L2_02w`	$Al_{56}O_{88}H_8$	B	−51.9
GSH	−1/2×	`gamma_surf-110l_L2_01w`	$Al_{56}O_{86}H_4$	R	−28.4

	$-1\times$	`water`	OH_2	J	36.8
	$1/2\times$	`gamma_surf-110l_L2_02w`	$Al_{56}O_{88}H_8$	B	-45.5
GSH	$-1/6\times$	`gamma_surf-110l_L2_00w`	$Al_{56}O_{84}$	R	-50.5
	$-1\times$	`water`	OH_2	J	25.8
	$1/6\times$	`gamma_surf-110l_L2_03w`	$Al_{56}O_{90}H_{12}$	B	-53.2
GSH	$-1/4\times$	`gamma_surf-110l_L2_01w`	$Al_{56}O_{86}H_4$	R	-45.3
	$-1\times$	`water`	OH_2	J	46.6
	$1/4\times$	`gamma_surf-110l_L2_03w`	$Al_{56}O_{90}H_{12}$	B	-50.6
GSH	$-1/2\times$	`gamma_surf-110l_L2_02w`	$Al_{56}O_{88}H_8$	R	-62.3
	$-1\times$	`water`	OH_2	J	56.5
	$1/2\times$	`gamma_surf-110l_L2_03w`	$Al_{56}O_{90}H_{12}$	B	-55.7
GSH	$-1/8\times$	`gamma_surf-110l_L2_00w`	$Al_{56}O_{84}$	R	-53.0
	$-1\times$	`water`	OH_2	J	24.9
	$1/8\times$	`gamma_surf-110l_L2_04w`	$Al_{56}O_{92}H_{16}$	B	-49.3
GSH	$-1/6\times$	`gamma_surf-110l_L2_01w`	$Al_{56}O_{86}H_4$	R	-50.4
	$-1\times$	`water`	OH_2	J	38.5
	$1/6\times$	`gamma_surf-110l_L2_04w`	$Al_{56}O_{92}H_{16}$	B	-46.2
GSH	$-1/4\times$	`gamma_surf-110l_L2_02w`	$Al_{56}O_{88}H_8$	R	-61.5
	$-1\times$	`water`	OH_2	J	39.3
	$1/4\times$	`gamma_surf-110l_L2_04w`	$Al_{56}O_{92}H_{16}$	B	-46.6
GSH	$-1/2\times$	`gamma_surf-110l_L2_03w`	$Al_{56}O_{90}H_{12}$	R	-60.7
	$-1\times$	`water`	OH_2	J	22.1
	$1/2\times$	`gamma_surf-110l_L2_04w`	$Al_{56}O_{92}H_{16}$	B	-37.5
GSH	$-1/2\times$	`gamma_surf-111_D1_03w`	$Al_{32}O_{54}H_{12}$	R	-60.7
	$-1\times$	`water`	OH_2	J	-5.4
	$1/2\times$	`gamma_surf-111_D1_04w`	$Al_{32}O_{56}H_{16}$	B	-49.0
GSH	$-1/4\times$	`gamma_surf-111_D1_03w`	$Al_{32}O_{54}H_{12}$	R	-49.5
	$-1\times$	`water`	OH_2	J	1.5
	$1/4\times$	`gamma_surf-111_D1_05w`	$Al_{32}O_{58}H_{20}$	B	-38.6
GSH	$-1/2\times$	`gamma_surf-111_D1_04w`	$Al_{32}O_{56}H_{16}$	R	-38.4
	$-1\times$	`water`	OH_2	J	8.5
	$1/2\times$	`gamma_surf-111_D1_05w`	$Al_{32}O_{58}H_{20}$	B	-28.3
GSH	$-1/6\times$	`gamma_surf-111_D1_03w`	$Al_{32}O_{54}H_{12}$	R	-42.9
	$-1\times$	`water`	OH_2	J	15.4
	$1/6\times$	`gamma_surf-111_D1_06w`	$Al_{32}O_{60}H_{24}$	B	-34.8
GSH	$-1/4\times$	`gamma_surf-111_D1_04w`	$Al_{32}O_{56}H_{16}$	R	-34.0
	$-1\times$	`water`	OH_2	J	25.8
	$1/4\times$	`gamma_surf-111_D1_06w`	$Al_{32}O_{60}H_{24}$	B	-27.7
GSH	$-1/2\times$	`gamma_surf-111_D1_05w`	$Al_{32}O_{58}H_{20}$	R	-29.6
	$-1\times$	`water`	OH_2	J	43.2
	$1/2\times$	`gamma_surf-111_D1_06w`	$Al_{32}O_{60}H_{24}$	B	-27.1
GSH	$-1/2\times$	`gamma_surf-111_D2_03w`	$Al_{32}O_{54}H_{12}$	R	-50.3
	$-1\times$	`water`	OH_2	J	46.1
	$1/2\times$	`gamma_surf-111_D2_04w`	$Al_{32}O_{56}H_{16}$	B	-27.4

GSH	−1/4×	`gamma_surf-111_D2_03w`	$Al_{32}O_{54}H_{12}$	R	−44.4
	−1×	`water`	OH_2	J	39.7
	1/4×	`gamma_surf-111_D2_05w`	$Al_{32}O_{58}H_{20}$	B	−34.8
GSH	−1/2×	`gamma_surf-111_D2_04w`	$Al_{32}O_{56}H_{16}$	R	−38.5
	−1×	`water`	OH_2	J	33.3
	1/2×	`gamma_surf-111_D2_05w`	$Al_{32}O_{58}H_{20}$	B	−42.1
GSH	−1/6×	`gamma_surf-111_D2_03w`	$Al_{32}O_{54}H_{12}$	R	−39.4
	−1×	`water`	OH_2	J	24.6
	1/6×	`gamma_surf-111_D2_06w`	$Al_{32}O_{60}H_{24}$	B	−34.0
GSH	−1/4×	`gamma_surf-111_D2_04w`	$Al_{32}O_{56}H_{16}$	R	−33.9
	−1×	`water`	OH_2	J	13.8
	1/4×	`gamma_surf-111_D2_06w`	$Al_{32}O_{60}H_{24}$	B	−37.3
GSH	−1/2×	`gamma_surf-111_D2_05w`	$Al_{32}O_{58}H_{20}$	R	−29.4
	−1×	`water`	OH_2	J	−5.6
	1/2×	`gamma_surf-111_D2_06w`	$Al_{32}O_{60}H_{24}$	B	−32.4
GSH	−1/2×	`gamma_surf-111_P1_1_05w`	$Al_{40}O_{70}H_{20}$	R	−21.2
	−1×	`water`	OH_2	J	1.7
	1/2×	`gamma_surf-111_P1_1_06w`	$Al_{40}O_{72}H_{24}$	B	−22.9
GSH	−1/2×	`gamma_surf-111_P1_2_03w`	$Al_{40}O_{66}H_{12}$	R	−51.2
	−1×	`water`	OH_2	J	47.5
	1/2×	`gamma_surf-111_P1_2_04w`	$Al_{40}O_{68}H_{16}$	B	−36.4
GSH	−1/4×	`gamma_surf-111_P1_2_03w`	$Al_{40}O_{66}H_{12}$	R	−41.3
	−1×	`water`	OH_2	J	34.9
	1/4×	`gamma_surf-111_P1_2_05w`	$Al_{40}O_{70}H_{20}$	B	−35.9
GSH	−1/2×	`gamma_surf-111_P1_2_04w`	$Al_{40}O_{68}H_{16}$	R	−31.4
	−1×	`water`	OH_2	J	22.3
	1/2×	`gamma_surf-111_P1_2_05w`	$Al_{40}O_{70}H_{20}$	B	−35.4
GSH	−1/6×	`gamma_surf-111_P1_2_03w`	$Al_{40}O_{66}H_{12}$	R	−35.0
	−1×	`water`	OH_2	J	24.3
	1/6×	`gamma_surf-111_P1_2_06w`	$Al_{40}O_{72}H_{24}$	B	−33.6
GSH	−1/4×	`gamma_surf-111_P1_2_04w`	$Al_{40}O_{68}H_{16}$	R	−26.9
	−1×	`water`	OH_2	J	12.7
	1/4×	`gamma_surf-111_P1_2_06w`	$Al_{40}O_{72}H_{24}$	B	−32.2
GSH	−1/2×	`gamma_surf-111_P1_2_05w`	$Al_{40}O_{70}H_{20}$	R	−22.3
	−1×	`water`	OH_2	J	3.1
	1/2×	`gamma_surf-111_P1_2_06w`	$Al_{40}O_{72}H_{24}$	B	−29.1
GSH	−1/2×	`gamma_surf-111_P2_1_05w`	$Al_{40}O_{70}H_{20}$	R	−23.0
	−1×	`water`	OH_2	J	60.2
	1/2×	`gamma_surf-111_P2_1_06w`	$Al_{40}O_{72}H_{24}$	B	−15.2
GSH	−1/2×	`gamma_surf-111_P2_2_04w`	$Al_{40}O_{68}H_{16}$	R	−46.0
	−1×	`water`	OH_2	J	29.6
	1/2×	`gamma_surf-111_P2_2_05w`	$Al_{40}O_{70}H_{20}$	B	−20.8
GSH	−1/4×	`gamma_surf-111_P2_2_04w`	$Al_{40}O_{68}H_{16}$	R	−17.4
	−1×	`water`	OH_2	J	3.1

	1/4×	`gamma_surf-111_P2_2_06w`	$Al_{40}O_{72}H_{24}$	B	−21.7
GSH	−1/2×	`gamma_surf-111_P2_2_05w`	$Al_{40}O_{70}H_{20}$	R	11.2
	−1×	`water`	OH_2	J	−23.5
	1/2×	`gamma_surf-111_P2_2_06w`	$Al_{40}O_{72}H_{24}$	B	−22.6
SUR	−1/32×	`gamma_bulk`	$Al_{32}O_{48}$	R	5.1
	1/64×	`gamma_surf-001_00w`	$Al_{64}O_{96}$	J	2.5
				B	4.9
SUR	−1/32×	`gamma_bulk`	$Al_{32}O_{48}$	R	5.7
	1/80×	`gamma_surf-110b_from_boehm_00w`	$Al_{80}O_{120}$	J	2.0
				B	3.6
SUR	−1/32×	`gamma_bulk`	$Al_{32}O_{48}$	R	7.7
	1/64×	`gamma_surf-110b_from_bulk_00w`	$Al_{64}O_{96}$	J	5.7
				B	7.3
SUR	−1/32×	`gamma_bulk`	$Al_{32}O_{48}$	R	9.3
	1/48×	`gamma_surf-110l_A1_00w`	$Al_{48}O_{72}$	J	1.4
				B	7.9
SUR	−1/32×	`gamma_bulk`	$Al_{32}O_{48}$	R	7.4
	1/56×	`gamma_surf-110l_A2_00w`	$Al_{56}O_{84}$	J	4.3
				B	6.6
SUR	−1/32×	`gamma_bulk`	$Al_{32}O_{48}$	R	7.2
	1/56×	`gamma_surf-110l_L2_00w`	$Al_{56}O_{84}$	J	1.1
				B	5.2

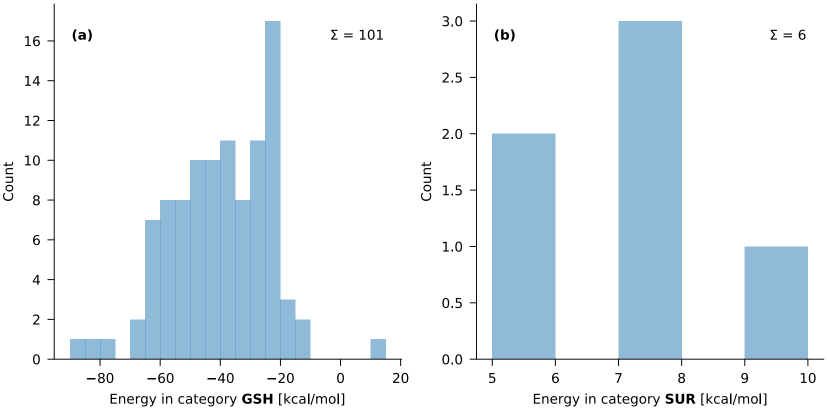

Figure S7: Histograms of reaction energies in the validation set.

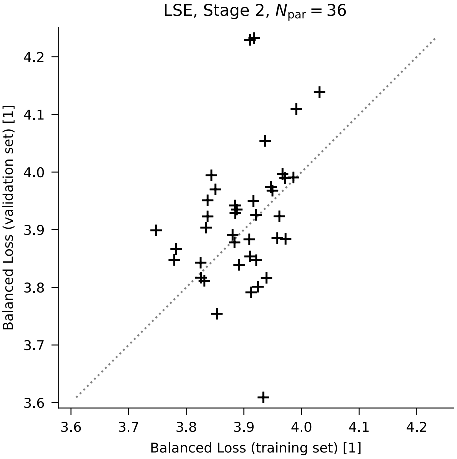

Figure S8: Parity plot comparing the value of the Balanced Loss function, for the training and validation sets, for the best solution from all 40 CMA runs (LSE form, optimization stage 2, see main text for details). Despite that the 40 optimized parameter vectors different significantly, their performance for the training and validation sets is similar.

Table S6: Comparison of root-mean-square-errors (RMSEs) of the initial force field by Joshi *et al.*[1] and the force field optimized in this work (BL). The RMSEs are computed for categories of training data and structures for which there is no counterpart in the validation set, i.e. not related to γ-Al_2O_3 surfaces. For reference, the standard deviation (SD) on the reference data is included.

Category	Unit	Ref. SD	Joshi RMSE	BL RMSE	#
Al−O	Å	0.08	0.35	0.07	1482
Al−Al	Å	0.37	0.33	0.07	1624
O−H	Å	0.03	0.26	0.04	638
O···H	Å	0.15	0.42	0.12	468
Al−O−Al	deg	22.0	9.3	3.3	1646
Al−O−H	deg	7.6	14.2	5.1	1077
O−Al−O	deg	32.2	11.0	3.2	3522
Al−O···H	deg	8.6	12.8	4.9	803
H−O···H	deg	15.8	18.2	8.1	484
H···O···H	deg	21.5	15.1	7.1	72
BSH	kcal mol^{-1}	11.8	38.1	4.7	5
GEH	kcal mol^{-1}	14.6	84.5	1.9	21
FOR	kcal mol^{-1}	10.3	22.4	2.4	11

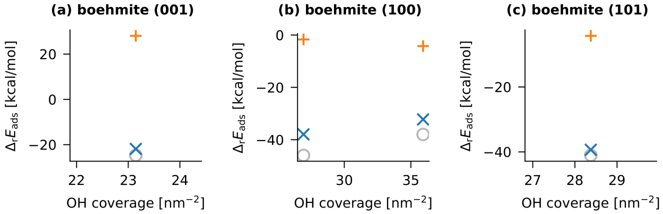

Figure S9: Adsorption energies, as defined in Eq. (3) in the main text, on boehmite surfaces as a function of the OH coverage, computed with different models: DFT (gray circle), Joshi (orange plus), this work (blue cross). All structures are derived from the training set.

S2 Comparison of Non-Equilibrium ReaxFF and DFT energies

A hydrated alumina slab (structure `gamma_surf-110l_A1_06w` from the validation set) was used as a starting point for a constant-temperature molecular dynamics (MD) run with VASP using the same level of theory as the training set. A Nosé-Hoover thermostat with a temperature of 1000 K and a relaxation time of 40 fs was used to stimulate the desorption of water from the surface. The simulation ran for 500 steps of 1 fs, and snapshots were taken every 10 steps for further analysis. Two desorption events occurs during the first 200 fs, after which no additional water molecules desorb.

The total single-point energies, E_{total}, of the selected snapshots are compared in Fig. S10(a). The DFT energies are shown in black and ReaxFF results with the new parameters in blue. ReaxFF energies obtained with the parameters of Joshi *et al.*[1] are depicted in red. The average is subtracted from both time series because these energies are only comparable up to a constant. While the fluctuations in DFT and ReaxFF energy are correlated, they also show significant deviations. This is expected, since the majority of the atoms are in the slab, not the water, and our training set emphasizes hydration reactions, not the vibrations in the alumina slab.

To show that our ReaxFF parameters can reproduce the relevant contribution to the DFT reference energy, additional single-point energies were calculated on the same snapshots from which some atoms were removed:

- E_{slab} is the energy of the slab and the water molecules that remain adsorbed, but without the atoms of the two water molecules that desorb during the simulation.
- $E_{2\text{H}_2\text{O}}$ is the energy computed for only the atoms of the two desorbing water molecules.

The energies of these two complementary subsystems are shown in Fig. S10(b) and Fig. S10(c), respectively. (Since these are also absolute energies, the average is again subtracted in both plots.) Finally, using these data, also the energy difference $(E_{\text{total}} - E_{\text{slab}} - E_{2\text{H}_2\text{O}})/2$ was computed and is shown in Fig. S10(d). In this case, no average was subtracted because the difference in energy has a chemically meaningful reference.

Fig. S10(b) shows deviations between ReaxFF and DFT energies that are very similar to those in Fig. S10(a), confirming that these deviations are due to the internal energy of the alumina slab. The results in Fig. S10(c) and Fig. S10(d) show a fair agreement between the DFT and ReaxFF energies obtained with our new parameters. Mainly Fig. S10(d) is of interest, because it is closely related to the hydration energies in the training set: This energy difference is calculated similarly to energy training data (Eq. (3) in the main text) but is now evaluated using non-equilibrium snapshots instead of optimized geometries. Because of this similarity, it is reasonable to expect a correspondence of the energies in Fig. S10(d). The average of the relative error over the first 200 fs is 25 % for the Balanced Loss parameters, which is comparable to relative errors on adsorption energies in the training set. For Fig. S10(a) and Fig. S10(b), however, a similar agreement would have been coincidental, since no related data were used for training. Also note that the new ReaxFF parameters show a clear improvement in Fig. S10(d) with respect to the parameters of Joshi *et al.*,[1] for which the relative error is 79 %. In Fig. S10(c) both ReaxFF parameterizations yield the same results because the parameters for water were not refitted.

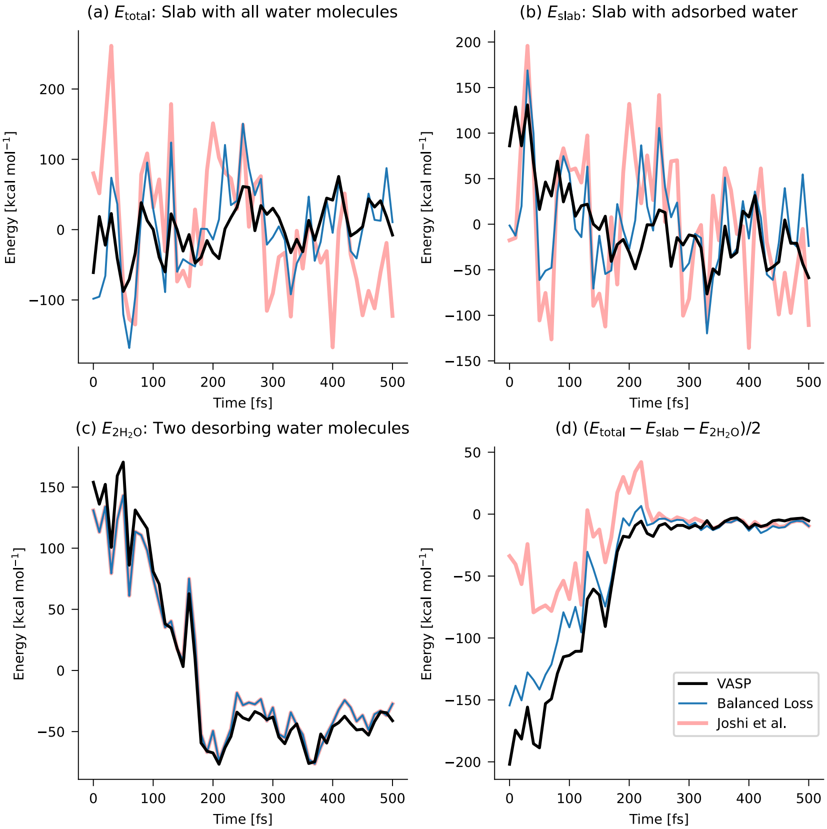

Figure S10: Comparison of DFT and ReaxFF energies for snapshots taken from a 1000 K DFT molecular dynamics simulation. (black: DFT, blue: ReaxFF with Balanced Loss parameters, red: ReaxFF with Joshi *et al.* parameters[1]). Panel (a) contains the total energy of the system, panel (b) the energy of the slab and the water molecules that remain adsorbed, and panel (c) the energy of only the desorbing water molecules. Panel (d) depicts the instantaneous hydration energy as computed with Eq. (3) in the main text.

References

(1) Joshi, K. L.; Psofogiannakis, G.; Duin, A. C. T. v.; Raman, S. Reactive molecular simulations of protonation of water clusters and depletion of acidity in H-ZSM-5 zeolite. *Phys. Chem. Chem. Phys.* **2014**, *16*, 18433–18441.

C.3 Supporting Information Paper II

Supporting Information: Scalable Reactive Molecular Dynamics with Machine Learning Potentials: Active Learning with Node Energy Density Estimation

Loïc Dumortier, Jelle Vekeman, Benoit Creton, Theodorus de Bruin and Toon Verstraelen

To Be Submitted in Journal of Chemical Theory and Computation, **2024**, *20*, 3779-3797

Reprinted with permission.

Supporting Information for:

Scalable Reactive Molecular Dynamics with Machine Learning Potentials: Active Learning with Node Energy Density Estimation

Loïc Dumortier,[†] Jelle Vekeman,[†] Benoit Creton,[‡] Theodorus de Bruin,[‡] and Toon Verstraelen[*,†]

[†]*Center for Molecular Modeling (CMM), Ghent University, Technologiepark-Zwijnaarde 46, B-9052, Zwijnaarde, Belgium*

[‡]*IFP Energies nouvelles, 1 et 4 Avenue de Bois-Préau, 92852 Rueil-Malmaison, France*

E-mail: toon.verstraelen@ugent.be

S1 Volume estimation for Packmol

To approximate the molecular volume to pack the structures, we employ a statistical estimation based on the atomic coordinates. Given a molecular structure, we define the set of atomic positions $r_i = (x_i, y_i, z_i)$ for each atom i, stored as:

$$R = \{r_1, r_2, ...r_N\} \in IR \tag{S1}$$

where N is the number of atoms in that molecular structure. The radius vector r_i^{vdW} is then obtained by by assigning the van der Waals radius of each atom. Subsequently, we sample n_{rep} random unit vectors u^k from a uniform distribution on the unit sphere. Each atomic position is then projected onto u^k. The molecular range on vector u^k is then expressed as:

$$t^k_{\min} = \min_i (r_i \cdot u^k - r_i^{\mathrm{vdW}}) \tag{S2}$$

$$t^k_{\max} = \max_i (r_i \cdot u^k + r_i^{\mathrm{vdW}}) \tag{S3}$$

the radius according to direction u^k is then:

$$d^k = \frac{t^k_{\max} - t^k_{\min}}{2} \tag{S4}$$

An estimate for the molecular volume can then be found by averaging out the sampled unit vectors:

$$V^k = \frac{1}{n_{\mathrm{rep}}} \sum_{k=1}^{n_{\mathrm{rep}}} \frac{4}{3} \pi {d^k}^3 \tag{S5}$$

Given a set of molecular species j with estimated volumes V_j and quantities N_j, the total system volume is then:

$$V_{\mathrm{sys}} = f \sum_j N_j V_n \tag{S6}$$

where f is an empirical adjustment factor and can be adjusted iteratively until packing the molecules is successful. Typical values are between 1.0 and 1.6.

S2 Carbon footprint

Table S1: Energy consumption and CO2 emissions for dedicated versus universal models simulating ∼5000 atom systems for 10,000 steps. Simulations were performed on two configurations representing different conditions: a dilute system (ratio 1 SiO_4H_4:7 NaOH:50 H_2O ; 4992 atoms) and a highly concentrated system (ratio 1 SiO_4H_4:2 NaOH:5 H_2O ; 4995 atoms). Note the differing timesteps: 1.2 fs (dedicated) and 0.6 fs (universal). Emissions (total and per nanosecond simulated) were estimated using the `code-carbon` package based on the projected 2025 Belgian energy mix, which is representative for the location of the high-performance clusters used in this work.

model	Energy (kWh)	Atom count	emissions (kg CO2)	timestep (fs)	kg CO2 per ns
universal	0.5181	4995	0.0715	0.6000	11.9247
universal	0.5321	4992	0.0735	0.6000	12.2487
dedicated	0.0935	4995	0.0129	1.2000	1.0758
dedicated	0.0916	4992	0.0126	1.2000	1.0539

D Bibliography

[1] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, Coarse-Grained Protein Models and Their Applications, Chemical Reviews **116**, 7898 (2016).

[2] T. Schlick, Molecular Dynamics: Basics, Molecular Modeling and Simulation: An Interdisciplinary Guide 383 (2002).

[3] T. Schlick, Molecular Dynamics: Further Topics, Molecular Modeling and Simulation: An Interdisciplinary Guide 419 (2002).

[4] P. Lykos and G. W. Pratt, Discussion on The Hartree-Fock Approximation, Reviews of Modern Physics **35**, 496 (1963).

[5] J. C. Slater, A Simplification of the Hartree-Fock Method, Physical Review **81**, 385 (1951).

[6] R. J. Bartlett and M. Musiał, Coupled-Cluster Theory in Quantum Chemistry, Reviews of Modern Physics **79**, 291 (2007).

[7] R. F. Bishop, An Overview of Coupled Cluster Theory and Its Applications in Physics, Theoretica Chimica Acta **80**, 95 (1991).

[8] P. Hohenberg and W. Kohn, Inhomogeneous Electron Gas, Physical Review **136**, B864 (1964).

[9] E. Teller, On the Stability of Molecules in the Thomas-Fermi Theory, Rev. Mod. Phys. **34**, 627 (1962).

[10] W. Kohn and L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, Physical Review **140**, A1133 (1965).

[11] M. Levy, Density-Functional Exchange Correlation through Coordinate Scaling in Adiabatic Connection and Correlation Hole, Physical Review a **43**, 4637 (1991).

[12] E. G. Lewars, Density Functional Calculations, Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics 445 (2011).

[13] D. M. Ceperley and B. J. Alder, Ground State of the Electron Gas by a Stochastic Method, Physical Review Letters **45**, 566 (1980).

[14] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Physical Review Letters **77**, 3865 (1996).

[15] J. P. Perdew and W. Yue, Accurate and Simple Density Functional for the Electronic Exchange Energy: Generalized Gradient Approximation, Physical Review B **33**, 8800 (1986).

[16] F. Tran, J. Stelzl, and P. Blaha, Rungs 1 to 4 of DFT Jacob's Ladder: Extensive Test on the Lattice Constant, Bulk Modulus, and Cohesive Energy of Solids, The Journal of Chemical Physics **144**, 204120 (2016).

[17] J. Moellmann and S. Grimme, DFT-D3 Study of Some Molecular Crystals, The Journal of Physical Chemistry C **118**, 7615 (2014).

[18] D. G. A. Smith, L. A. Burns, K. Patkowski, and C. D. Sherrill, Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory, The Journal of Physical Chemistry Letters **7**, 2197 (2016).

[19] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu, The Journal of Chemical Physics **132**, 154104 (2010).

[20] S. Grimme, S. Ehrlich, and L. Goerigk, Effect of the Damping Function in Dispersion Corrected Density Functional Theory, Journal of Computational Chemistry **32**, 1456 (2011).

[21] A. D. Becke and E. R. Johnson, Exchange-hole dipole moment and the dispersion interaction, The Journal of Chemical Physics **122**, 154104 (2005).

[22] S. N. Steinmann and C. Corminboeuf, Comprehensive Benchmarking of a Density-Dependent Dispersion Correction, Journal of Chemical Theory and Computation **7**, 3567 (2011).

[23] J. Thijssen, *Computational Physics*, 2nd ed. (Cambridge University Press, Cambridge, 2007).

[24] J. E. Jones and S. Chapman, On the Determination of Molecular Fields. —I. From the Variation of the Viscosity of a Gas with Temperature, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **106**, 441 (1997).

[25] B. R. Brooks et al., CHARMM: The Biomolecular Simulation Program, Journal of Computational Chemistry **30**, 1545 (2009).

[26] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, Journal of Computational Chemistry **4**, 187 (1983).

[27] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers, Softwarex 19 (2015).

[28] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Development and Testing of a General Amber Force Field, Journal of Computational Chemistry **25**, 1157 (2004).

[29] J. Tersoff, Modeling Solid-State Chemistry: Interatomic Potentials for Multicomponent Systems, Physical Review B **39**, 5566 (1989).

[30] J. Tersoff, New Empirical Approach for the Structure and Energy of Covalent Systems, Physical Review B **37**, 6991 (1988).

[31] D. W. Brenner, Empirical Potential for Hydrocarbons for Use in Simulating the Chemical Vapor Deposition of Diamond Films, Physical Review B **42**, 9458 (1990).

[32] A. C. T. van Duin, J. M. A. Baas, and B. van de Graaf, Delft Molecular Mechanics: A New Approach to Hydrocarbon Force Fields. Inclusion of a Geometry-Dependent Charge Calculation, Journal of the Chemical Society, Faraday Transactions **90**, 2881 (1994).

[33] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, ReaxFF: A Reactive Force Field for Hydrocarbons, The Journal of Physical Chemistry a **105**, 9396 (2001).

[34] M. F. Russo and A. C. T. van Duin, Atomistic-Scale Simulations of Chemical Reactions: Bridging from Quantum Chemistry to Engineering, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **269**, 1549 (2011).

[35] E. Iype, M. Hütter, A. P. J. Jansen, S. V. Nedea, and C. C. M. Rindt, Parameterization of a Reactive Force Field Using a Monte Carlo Algorithm, Journal of Computational Chemistry **34**, 1143 (2013).

[36] D. Furman and D. J. Wales, Transforming the Accuracy and Numerical Stability of ReaxFF Reactive Force Fields, The Journal of Physical Chemistry Letters **10**, 7215 (2019).

[37] D. Furman and D. J. Wales, A Well-Behaved Theoretical Framework for ReaxFF Reactive Force Fields, The Journal of Chemical Physics **153**, 21102 (2020).

[38] M. Freitas Gustavo and T. Verstraelen, GloMPO (Globally Managed Parallel Optimization): A Tool for Expensive, Black-Box Optimizations, Application to ReaxFF Reparameterizations, Journal of Cheminformatics **14**, 7 (2022).

[39] H. R. Larsson, A. C. van Duin, and B. Hartke, Global Optimization of Parameters in the Reactive Force Field ReaxFF for SiOH, Journal of Computational Chemistry **34**, 2178 (2013).

[40] P. Pahari and S. Chaturvedi, Determination of Best-Fit Potential Parameters for a Reactive Force Field Using a Genetic Algorithm, Journal of Molecular Modeling **18**, 1049 (2012).

[41] G. Shchygol, A. Yakovlev, V. D. Adri C. T., T. Tomas, and T. Verstraelen, ReaxFF Parameter Optimization with Monte-Carlo and Evolutionary Algorithms: Guidelines and Insights, (2019).

[42] J. Yeon, S. C. Chowdhury, C. M. Daksha, and J. W. J. Gillespie, Development of Mg/Al/Si/O ReaxFF Parameters for Magnesium Aluminosilicate Glass Using an Artificial Neural Network-Assisted Genetic Algorithm, The Journal of Physical Chemistry C **125**, 18380 (2021).

[43] Q. Sun, J. Zhong, P. Shi, H. Xu, and Y. Wang, An Improved Reactive Force Field Parameter Optimization Framework Based on Simulated Annealing and Particle Swarm Optimization Algorithms, Computational Materials Science **251**, 113776 (2025).

[44] G. Barcaro, S. Monti, L. Sementa, and V. Carravetta, Parametrization of a Reactive Force Field (ReaxFF) for Molecular Dynamics Simulations of Si Nanoparticles, Journal of Chemical Theory and Computation **13**, 3854 (2017).

[45] P. J. M. van Laarhoven and E. H. L. Aarts, Performance of the Simulated Annealing Algorithm, Simulated Annealing: Theory and Applications 77 (1987).

[46] P. J. M. van Laarhoven and E. H. L. Aarts, Simulated Annealing, Simulated Annealing: Theory and Applications 7 (1987).

[47] N. Hansen and A. Ostermeier, *Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adapta-*

tion, in *Proceedings of IEEE International Conference on Evolutionary Computation* (1996), pp. 312–317.

[48] N. Hansen, The CMA Evolution Strategy: A Comparing Review, Towards a New Evolutionary Computation. Advances on Estimation of Distribution Algorithms 75 (2006).

[49] N. Hansen, The CMA Evolution Strategy: A Tutorial, Arxiv:1604.00772 [Cs, Stat] (2016).

[50] N. Hansen, S. D. Müller, and P. Koumoutsakos, Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES), (2006).

[51] C. Igel, N. Hansen, and S. Roth, Covariance Matrix Adaptation for Multi-objective Optimization, Evolutionary Computation **15**, 1 (2007).

[52] J. Behler, Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials, The Journal of Chemical Physics **134**, 74106 (2011).

[53] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Physical Review Letters **98**, 146401 (2007).

[54] A. P. Bartók, R. Kondor, and G. Csányi, On Representing Chemical Environments, Physical Review B **87**, 184115 (2013).

[55] M. A. Caro, Optimizing Many-Body Atomic Descriptors for Enhanced Computational Performance of Machine Learning Based Interatomic Potentials, Physical Review B **100**, 24112 (2019).

[56] M. J. Willatt, F. Musil, and M. Ceriotti, Atom-Density Representations for Machine Learning, The Journal of Chemical Physics **150**, 154110 (2019).

[57] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural Message Passing for Quantum Chemistry, (2017).

[58] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Message Passing Neural Networks, Machine Learning Meets Quantum Physics 199 (2020).

[59] K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, *SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions*, in *Advances in Neural Information Processing Systems*, Vol. 30 (Curran Associates, Inc., 2017).

[60] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, SchNet – A Deep Learning Architecture for Molecules and Materials, The Journal of Chemical Physics **148**, 241722 (2018).

[61] K. Schütt, O. Unke, and M. Gastegger, *Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra*, in *Proceedings of the 38th International Conference on Machine Learning* (PMLR, 2021), pp. 9377–9388.

[62] M. Geiger and T. Smidt, E3nn: Euclidean Neural Networks, (2022).

[63] I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. C. Simm, R. Drautz, C. Ortner, B. Kozinsky, and G. Csányi, The Design Space of E(3)-Equivariant Atom-Centered Interatomic Potentials, (2022).

[64] I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner, and G. Csányi, MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields, (2023).

[65] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials, Nature Communications **13**, 2453 (2022).

[66] I. Levin and D. Brandon, Metastable Alumina Polymorphs: Crystal Structures and Transition Sequences, Journal of the American Ceramic Society **81**, 1995 (1998).

[67] M. Trueba and S. P. Trasatti, \gamma-Alumina as a Support for Catalysts: A Review of Fundamental Aspects, European Journal of Inorganic Chemistry **2005**, 3393 (2005).

[68] L. D. Hart and E. Lense, *Alumina Chemicals: Science and Technology Handbook* (John Wiley & Sons, 1990).

[69] J. R. Jain and C. N. Pillai, Catalytic Dehydration of Alcohols over Alumina: Mechanism of Ether Formation, Journal of Catalysis **9**, 322 (1967).

[70] A. Samimi, M. Zakeri, B. Maleki, and D. Mohebbi-Kalhori, Experimental and Statistical Assessments of the Mechanical Strength Reliability of Gamma Alumina Catalyst Supports, Particuology **21**, 74 (2015).

[71] F. M. Almulla, S. A. Ali, M. R. Aldossary, E. I. Alnaimi, V. I. Zholobenko, A. bin Jumah, and A. A. Garforth, On the Influence of Alumina as a Binder on the Performance of Pt-Beta Catalyst during the

Transalkylation of Toluene and 1,2,4-Trimethylbenzene, Microporous and Mesoporous Materials **320**, 111095 (2021).

[72] Z. Yang and Y. S. Lin, Sol-Gel Synthesis of Silicalite/\gamma-Alumina Granules, Industrial & Engineering Chemistry Research **39**, 4944 (2000).

[73] A. C. Vieira Coelho, G. A. Rocha, P. Souza Santos, H. Souza Santos, and P. K. Kiyohara, Specific Surface Area and Structures of Aluminas from Fibrillar Pseudoboehmite, Matéria (Rio De Janeiro) **13**, 329 (2008).

[74] H. Knözinger and P. and Ratnasamy, Catalytic Aluminas: Surface Models and Characterization of Surface Sites, Catalysis Reviews **17**, 31 (1978).

[75] N. M. Stuart and K. Sohlberg, The Microstructure of ?-Alumina, Energies **14**, 6472 (2021).

[76] G. Paglia, C. E. Buckley, A. L. Rohl, R. D. Hart, K. Winter, A. J. Studer, B. A. Hunter, and J. V. Hanna, Boehmite Derived Gamma-Alumina System. 1. Structural Evolution with Temperature, with the Identification and Structural Determination of a New Transition Phase, Gamma'-Alumina, Chemistry of Materials **16**, 220 (2004).

[77] I. Levin and D. Brandon, Metastable Alumina Polymorphs: Crystal Structures and Transition Sequences, Journal of the American Ceramic Society **81**, 1995 (1998).

[78] V. Jayaram and C. G. Levi, The Structure of \delta-Alumina Evolved from the Melt and the \emph{γ} \rightarrow \emph{δ} Transformation, Acta Metallurgica **37**, 569 (1989).

[79] M. Digne, P. Sautet, P. Raybaud, P. Euzen, and H. Toulhoat, Use of DFT to Achieve a Rational Understanding of Acid–Basic Properties of \gamma-Alumina Surfaces, Journal of Catalysis **226**, 54 (2004).

[80] D. Vandenabeele, N. Doppelhammer, S. Radhakrishnan, C. V. Chandran, W. Wangermez, A. Rais, E. Vandeurzen, B. Jakoby, C. Kirschhock, and E. Breynaert, Hydrated Silicate Ionic Liquids: Ionic Liquids for Silicate Material Synthesis, Journal of Molecular Liquids **417**, 126603 (2025).

[81] R. E. Morris, Ionothermal Synthesis—Ionic Liquids as Functional Solvents in the Preparation of Crystalline Materials, Chemical Communications 2990 (2009).

[82] E. R. Cooper, C. D. Andrews, P. S. Wheatley, P. B. Webb, P. Wormald, and R. E. Morris, Ionic Liquids and Eutectic Mixtures as Solvent and Template in Synthesis of Zeolite Analogues, Nature **430**, 1012 (2004).

[83] R. E. Morris, Ionic Liquids and Microwaves—Making Zeolites for Emerging Applications, Angewandte Chemie International Edition **47**, 442 (2008).

[84] M. Antonietti, D. Kuang, B. Smarsly, and Y. Zhou, Ionic Liquids for the Convenient Synthesis of Functional Nanoparticles and Other Inorganic Nanostructures, Angewandte Chemie International Edition **43**, 4988 (2004).

[85] D. Yuan, D. He, S. Xu, Z. Song, M. Zhang, Y. Wei, Y. He, S. Xu, Z. Liu, and Y. Xu, Imidazolium-Based Ionic Liquids as Novel Organic SDA to Synthesize High-Silica Y Zeolite, Microporous and Mesoporous Materials **204**, 1 (2015).

[86] L. van Tendeloo, M. Haouas, J. A. Martens, C. E. A. Kirschhock, E. Breynaert, and F. Taulelle, Zeolite Synthesis in Hydrated Silicate Ionic Liquids, Faraday Discussions **179**, 437 (2015).

[87] M. Houlleberghs, E. Breynaert, K. Asselman, E. Vaneeckhaute, S. Radhakrishnan, M. W. Anderson, F. Taulelle, M. Haouas, J. A. Martens, and C. E. A. Kirschhock, Evolution of the Crystal Growth Mechanism of Zeolite W (MER) with Temperature, Microporous and Mesoporous Materials **274**, 379 (2019).

[88] K. Asselman, C. Kirschhock, and E. Breynaert, Illuminating the Black Box: A Perspective on Zeolite Crystallization in Inorganic Media, Accounts of Chemical Research **56**, 2391 (2023).

[89] N. Pellens et al., Nucleation of Porous Crystals from Ion-Paired Prenucleation Clusters, Chemistry of Materials **34**, 7139 (2022).

[90] K. Asselman, S. Radhakrishnan, N. Pellens, C. V. Chandran, M. Houlleberghs, Y. Xu, J. A. Martens, S. P. Sree, C. E. Kirschhock, and E. Breynaert, HSIL-Based Synthesis of Ultracrystalline K,Na-JBW, a Zeolite Exhibiting Exceptional Framework Ordering and Flexibility, Chemistry of Materials **34**, 7159 (2022).

[91] K. Asselman, N. Pellens, B. Thijs, N. Doppelhammer, M. Haouas, F. Taulelle, J. A. Martens, E. Breynaert, and C. E. Kirschhock, Ion-Pairs in Aluminosilicate-Alkali Synthesis Liquids Determine the Aluminum

Content and Topology of Crystallizing Zeolites, Chemistry of Materials **34**, 7150 (2022).

[92] K. Asselman, D. Vandenabeele, N. Pellens, N. Doppelhammer, C. E. Kirschhock, and E. Breynaert, Structural Aspects Affecting Phase Selection in Inorganic Zeolite Synthesis, Chemistry of Materials **34**, 11081 (2022).

[93] M. Haouas, L. Lakiss, C. Martineau, J. El Fallah, V. Valtchev, and F. Taulelle, Silicate Ionic Liquid Synthesis of Zeolite Merlinoite: Crystal Size Control from Crystalline Nanoaggregates to Micron-Sized Single-Crystals, Microporous and Mesoporous Materials **198**, 35 (2014).

[94] J. Vekeman, D. Vandenabeele, N. Doppelhammer, E. Vandeurzen, E. Breynaert, C. E. A. Kirschhock, and T. Verstraelen, Simple Molecular Model for Hydrated Silicate Ionic Liquids, a Realistic Zeolite Precursor, Chemistry of Materials **36**, 3886 (2024).

[95] E. Brunk and U. Rothlisberger, Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States, Chemical Reviews **115**, 6217 (2015).

[96] M. Döntgen, M.-D. Przybylski-Freund, L. C. Kröger, W. A. Kopp, A. E. Ismail, and K. Leonhard, Automated Discovery of Reaction Pathways, Rate Constants, and Transition States Using Reactive Molecular Dynamics Simulations, Journal of Chemical Theory and Computation **11**, 2517 (2015).

[97] K. Chenoweth, A. C. T. van Duin, S. Dasgupta, and W. A. G. III, Initiation Mechanisms and Kinetics of Pyrolysis and Combustion of JP-10 Hydrocarbon Jet Fuel, Journal of Physical Chemistry a **113**, 1740 (2009).

[98] J. Ding, L. Zhang, Y. Zhang, and K.-L. Han, A Reactive Molecular Dynamics Study of n-Heptane Pyrolysis at High Temperature, Journal of Physical Chemistry a **117**, 3266 (2013).

[99] X. Li, M. Zheng, C. Ren, and L. Guo, ReaxFF Molecular Dynamics Simulations of Thermal Reactivity of Various Fuels in Pyrolysis and Combustion, Energy & Fuels **35**, 11707 (2021).

[100] J. C. Fogarty, H. M. Aktulga, A. Y. Grama, A. C. T. van Duin, and S. A. Pandit, A Reactive Molecular Dynamics Simulation of the Silica-Water Interface, Journal of Chemical Physics **132**, 174704 (2010).

[101] J. E. Mueller, A. C. T. van Duin, and W. A. Goddard, Application of the ReaxFF Reactive Force Field to Reactive Dynamics of Hydrocarbon Chemisorption and Decomposition, Journal of Physical Chemistry C **114**, 5675 (2010).

[102] J. Müller and B. Hartke, reaxFF Reactive Force Field for Disulfide Mechanochemistry, Fitted to Multireference Ab Initio Data, Journal of Chemical Theory and Computation **12**, 3913 (2016).

[103] J. M. Rimsza, R. E. Jones, and L. J. Criscenti, Crack Propagation in Silica from Reactive Classical Molecular Dynamics Simulations, Journal of the American Ceramic Society **101**, 1488 (2017).

[104] Q. Mao, A. C. van Duin, and K. Luo, Formation of Incipient Soot Particles from Polycyclic Aromatic Hydrocarbons: A ReaxFF Molecular Dynamics Study, Carbon **121**, 380 (2017).

[105] T. Lei, W. Guo, Q. Liu, H. Jiao, D.-B. Cao, B. Teng, Y.-W. Li, X. Liu, and X.-D. Wen, Mechanism of Graphene Formation via Detonation Synthesis: A DFTB Nanoreactor Approach, Journal of Chemical Theory and Computation **15**, 3654 (2019).

[106] A. C. T. van Duin, A. Strachan, S. Stewman, Q. Zhang, X. Xu, and W. A. Goddard, ReaxFFSiO Reactive Force Field for Silicon and Silicon Oxide Systems, Journal of Physical Chemistry a **107**, 3803 (2003).

[107] T. P. Senftle et al., The ReaxFF Reactive Force-Field: Development, Applications and Future Directions, Npj Computational Materials **2**, 15011 (2016).

[108] J. Tersoff, Empirical Interatomic Potential for Carbon, with Applications to Amorphous Carbon, Physical Review Letters **61**, 2879 (1988).

[109] T. C. O'Connor, J. Andzelm, and M. O. Robbins, AIREBO-M: A Reactive Model for Hydrocarbons at Extreme Pressures, Journal of Chemical Physics **142**, 24903 (2015).

[110] D. Zhang, A. F. Fonseca, T. Liang, S. R. Phillpot, and S. B. Sinnott, Dynamics of Graphene/Al Interfaces Using COMB3 Potentials, **3**, 114002 (2019).

[111] P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi, and A. Michaelides, An Accurate and Transferable Machine Learning Potential for Carbon, Journal of Chemical Physics **153**, 34702 (2020).

[112] L.-Y. Xue, F. Guo, Y.-S. Wen, S.-Q. Feng, X.-N. Huang, L. Guo, H.-S. Li, S.-X. Cui, G.-Q. Zhang, and Q.-L. Wang, ReaxFF-MPNN Machine

Learning Potential: A Combination of Reactive Force Field and Message Passing Neural Networks, Physical Chemistry Chemical Physics **23**, 19457 (2021).

[113] M. Schreiner, A. Bhowmik, T. Vegge, J. Busk, and O. Winther, Transition1x - a Dataset for Building Generalizable Reactive Machine Learning Potentials, Scientific Data **9**, 779 (2022).

[114] P. Pahari and S. Chaturvedi, Determination of Best-Fit Potential Parameters for a Reactive Force Field Using a Genetic Algorithm, Journal of Molecular Modeling **18**, 1049 (2011).

[115] J. D. Deetz and R. Faller, Parallel Optimization of a Reactive Force Field for Polycondensation of Alkoxysilanes, Journal of Physical Chemistry B **118**, 10966 (2014).

[116] A. Jaramillo-Botero, S. Naserifar, and W. A. Goddard, General Multiobjective Force Field Optimization Framework, with Application to Reactive Force Fields for Silicon Carbide, Journal of Chemical Theory and Computation **10**, 1426 (2014).

[117] M. Dittner, J. Müller, H. M. Aktulga, and B. Hartke, Efficient Global Optimization of Reactive Force-Field Parameters, Journal of Computational Chemistry **36**, 1550 (2015).

[118] T. Trnka, I. Tvarǒska, and J. Kǒca, Automated Training of ReaxFF Reactive Force Fields for Energetics of Enzymatic Reactions, Journal of Chemical Theory and Computation **14**, 291 (2017).

[119] D. Furman, B. Carmeli, Y. Zeiri, and R. Kosloff, Enhanced Particle Swarm Optimization Algorithm: Efficient Training of ReaxFF Reactive Force Fields, Journal of Chemical Theory and Computation **14**, 3100 (2018).

[120] M. Y. Sengul, Y. Song, N. Nayir, Y. Gao, Y. Hung, T. Dasgupta, and A. C. T. van Duin, INDEEDopt: A Deep Learning-Based ReaxFF Parameterization Framework, Npj Computational Materials **7**, 68 (2021).

[121] M. C. Kaymak, A. Rahnamoun, K. A. O'Hearn, A. C. T. van Duin, K. M. Merz Jr., and H. M. Aktulga, JAX-ReaxFF: A Gradient-Based Framework for Fast Optimization of Reactive Force Fields, Journal of Chemical Theory and Computation **18**, 5181 (2022).

[122] L. Komissarov, L. Krep, F. Schmalz, W. A. Kopp, K. Leonhard, and T. Verstraelen, A Reactive Molecular Dynamics Study of Chlorinated Organic Compounds. Part I: Force Field Development, Chemphyschem :

a European Journal of Chemical Physics and Physical Chemistry **24**, e202200786 (2023).

[123] M. F. Gustavo, M. Hellström, and T. Verstraelen, Sensitivity Analysis for ReaxFF Reparametrization Using the Hilbert–Schmidt Independence Criterion, Journal of Chemical Theory and Computation **19**, 2557 (2023).

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

[125] O. T. Unke and M. Meuwly, PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges, Journal of Chemical Theory and Computation **15**, 3678 (2019).

[126] M. Cools-Ceuppens, J. Dambre, and T. Verstraelen, Modeling Electronic Response Properties with an Explicit-Electron Machine Learning Potential, Journal of Chemical Theory and Computation **18**, 1672 (2022).

[127] M. V. Fedkin et al., Development of the ReaxFF Methodology for Electrolyte–Water Systems, Journal of Physical Chemistry a **123**, 2125 (2019).

[128] I. Brown, R. Smith, and S. D. Kenny, A ReaxFF Potential for Al–ZnO Systems, Modelling and Simulation in Materials Science and Engineering **30**, 35001 (2022).

[129] T. Verstraelen, S. V. Sukhomlinov, V. Van Speybroeck, M. Waroquier, and K. S. Smirnov, Computation of Charge Distribution and Electrostatic Potential in Silicates with the Use of Chemical Potential Equalization Models, Journal of Physical Chemistry C **116**, 490 (2012).

[130] S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy, and R. Schmid, MOF-FF – A Flexible First-Principles Derived Force Field for Metal-Organic Frameworks, Physica Status Solidi B **250**, 1128 (2013).

[131] S. Grimme, C. Bannwarth, and P. Shushkov, A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All Spd-Block Elements (Z = 1–86), Journal of Chemical Theory and Computation **13**, 1989 (2017).

[132] L. Komissarov and T. Verstraelen, Improving the Silicon Interactions of GFN-xTB, Journal of Chemical Information and Modeling **61**, 5931 (2021).

[133] J. Li, X. Song, P. Li, and J. Herzfeld, A Carbon Is a Carbon Is a Carbon: Orbital-free Simulations of Hydrocarbon Chemistry without Resort to Atom Types, Journal of Physical Chemistry a **126**, 8468 (2022).

[134] A. Włodarczyk, M. Uchroński, A. Podsiadły-Paszkowska, J. Irek, and B. M. Szyja, Mixing ReaxFF Parameters for Transition Metal Oxides Using Force-Matching Method, Journal of Molecular Modeling **28**, 8 (2022).

[135] L. Komissarov, R. Rüger, M. Hellström, and T. Verstraelen, ParAMS: Parameter Optimization for Atomistic and Molecular Simulations, Journal of Chemical Information and Modeling **61**, 3737 (2021).

[136] M. F. Gustavo and T. Verstraelen, GloMPO (Globally Managed Parallel Optimization): A Tool for Expensive, Black-Box Optimizations, Application to ReaxFF Reparameterizations, J. Cheminformatics **14**, 7 (2022).

[137] J. S. Smith, O. Isayev, and A. E. Roitberg, ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost, Chemical Science **8**, 3192 (2017).

[138] C. Chen and S. P. Ong, A Universal Graph Deep Learning Interatomic Potential for the Periodic Table, **2**, 718 (2022).

[139] S. Takamoto et al., Towards Universal Neural Network Potential for Material Discovery Applicable to Arbitrary Combination of 45 Elements, Nature Communications **13**, 2991 (2022).

[140] G. Lefèvre, M. Duc, P. Lepeut, R. Caplain, and M. Fédoroff, Hydration of \gamma-Alumina in Water and Its Effects on Surface Reactivity, Langmuir : the ACS Journal of Surfaces and Colloids **18**, 7530 (2002).

[141] C. Arrouvel, M. Digne, M. Breysse, H. Toulhoat, and P. Raybaud, Effects of Morphology on Surface Hydroxyl Concentration: A DFT Comparison of Anatase–TiO2 and \gamma-Alumina Catalytic Supports, Journal of Catalysis **222**, 152 (2004).

[142] M. Lagauche, K. Larmier, E. Jolimaitre, K. Barthelet, C. Chizallet, L. Favergeon, and M. Pijolat, Thermodynamic Characterization of the Hydroxyl Group on the \gamma-Alumina Surface by the Energy Distribution Function, Journal of Physical Chemistry C **121**, 16770 (2017).

[143] T. Pigeon, C. Chizallet, and P. Raybaud, Revisiting \gamma-Alumina Surface Models through the Topotactic Transformation of Boehmite Surfaces, Journal of Catalysis **405**, 140 (2022).

[144] P. Euzen, P. Raybaud, X. Krokidis, H. Toulhoat, J.-L. Le Loarer, J.-P. Jolivet, and C. Froidefond, Alumina, Handbook of Porous Solids 1591 (2002).

[145] H. Knözinger, Dehydration of Alcohols on Aluminum Oxide, Angewandte Chemie. International Edition in English **7**, 791 (1968).

[146] H. Knözinger, H. Bühl, and K. Kochloefl, The Dehydration of Alcohols on Alumina: XIV. Reactivity and Mechanism, Journal of Catalysis **24**, 57 (1972).

[147] T. K. Phung, A. Lagazzo, M. Á. Rivero Crespo, V. Sánchez Escribano, and G. Busca, A Study of Commercial Transition Aluminas and of Their Catalytic Activity in the Dehydration of Ethanol, Journal of Catalysis **311**, 102 (2014).

[148] A. L. Kohl and R. B. Nielsen, Chapter 8 - Sulfur Recovery Processes, Gas Purification (Fifth Edition) 670 (1997).

[149] K. Larmier, C. Chizallet, N. Cadran, S. Maury, J. Abboud, A.-F. Lamic-Humblot, E. Marceau, and H. Lauron-Pernot, Mechanistic Investigation of Isopropanol Conversion on Alumina Catalysts: Location of Active Sites for Alkene/Ether Production, Acs Catalysis **5**, 4423 (2015).

[150] P. Raybaud and H. Toulhoat, *Catalysis by Transition Metal Sulphides: From Molecular Theory to Industrial Application* (Editions TECHNIP, 2013).

[151] C. Copéret, A. Comas-Vives, M. P. Conley, D. P. Estes, A. Fedorov, V. Mougel, H. Nagae, F. Núñez-Zarur, and P. A. Zhizhko, Surface Organometallic and Coordination Chemistry toward Single-Site Heterogeneous Catalysts: Strategies, Methods, Structures, and Activities, Chemical Reviews **116**, 323 (2016).

[152] M. Boudart, Catalysis by Supported Metals, Advances in Catalysis 153 (1969).

[153] R. Prins, On the Structure of \gamma-Al_{2}O_{3}, Journal of Catalysis **392**, 336 (2020).

[154] M. C. Valero and P. Raybaud, Computational Chemistry Approaches for the Preparation of Supported Catalysts: Progress and Challenges, Journal of Catalysis **391**, 539 (2020).

[155] X. Krokidis, P. Raybaud, A.-E. Gobichon, B. Rebours, P. Euzen, and H. Toulhoat, Theoretical Study of the Dehydration Process of Boehmite to \gamma-Alumina, Journal of Physical Chemistry B **105**, 5121 (2001).

[156] R. Wischert, P. Laurent, C. Copéret, F. Delbecq, and P. Sautet, \gamma-Alumina: The Essential and Unexpected Role of Water for the Structure, Stability, and Reactivity of "Defect" Sites, Journal of the American Chemical Society **134**, 14430 (2012).

[157] M. Digne, P. Sautet, P. Raybaud, P. Euzen, and H. Toulhoat, Hydroxyl Groups on \gamma-Alumina Surfaces: A DFT Study, Journal of Catalysis **211**, 1 (2002).

[158] R. Réocreux, É. Girel, P. Clabaut, A. Tuel, M. Besson, A. Chaumonnot, A. Cabiac, P. Sautet, and C. Michel, Reactivity of Shape-Controlled Crystals and Metadynamics Simulations Locate the Weak Spots of Alumina in Water, Nature Communications **10**, 3139 (2019).

[159] P. P. Mardilovich, A. N. Govyadinov, N. I. Mukhurov, A. M. Rzhevskii, and R. Paterson, New and Modified Anodic Alumina Membranes Part I. Thermotreatment of Anodic Alumina Membranes, Journal of Membrane Science **98**, 131 (1995).

[160] G. E. Brown et al., Metal Oxide Surfaces and Their Interactions with Aqueous Solutions and Microbial Organisms, Chemical Reviews **99**, 77 (1999).

[161] A. T. F. Batista et al., Beyond \gamma-Al_{2}O_3 Crystallite Surfaces: The Hidden Features of Edges Revealed by Solid-State 1H NMR and DFT Calculations, Journal of Catalysis **378**, 140 (2019).

[162] A. Batista et al., Structure, Location and Spatial Proximities of Hydroxyls on \gamma-Alumina Crystallites by High-Resolution Solid-State NMR and DFT Modelling: Why Edges Hold the Key, Acs Catalysis **13**, 6536 (2023).

[163] M. C. Pitman and A. C. T. van Duin, Dynamics of Confined Reactive Water in Smectite Clay–Zeolite Composites, Journal of the American Chemical Society **134**, 3042 (2012).

[164] J. M. Rimsza, J. Yeon, A. C. T. van Duin, and J. Du, Water Interactions with Nanoporous Silica: Comparison of ReaxFF and Ab Initio Based Molecular Dynamics Simulations, Journal of Physical Chemistry C **120**, 24803 (2016).

[165] A. J. Porter and A. J. O'Malley, A Classical Molecular Dynamics Study on the Effect of Si/al Ratio and Silanol Nest Defects on Water Diffusion in Zeolite HY, Journal of Physical Chemistry C **125**, 11567 (2021).

[166] W. F. van Gunsteren and H. J. C. Berendsen, Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry, Angewandte Chemie. International Edition in English **29**, 992 (1990).

[167] A. Nakata et al., Large Scale and Linear Scaling DFT with the CONQUEST Code, Journal of Chemical Physics **152**, 164112 (2020).

[168] T. D. Kühne et al., CP2K: An Electronic Structure and Molecular Dynamics Software Package - Quickstep: Efficient and Accurate Electronic Structure Calculations, Journal of Chemical Physics **152**, 194103 (2020).

[169] P. Raybaud, M. Digne, R. Iftimie, W. Wellens, P. Euzen, and H. Toulhoat, Morphology and Surface Properties of Boehmite (\gamma-AlOOH): A Density Functional Theory Study, Journal of Catalysis **201**, 236 (2001).

[170] A. Motta, M.-P. Gaigeot, and D. Costa, AIMD Evidence of Inner Sphere Adsorption of Glycine on a Stepped (101) Boehmite AlOOH Surface, Journal of Physical Chemistry C **116**, 23418 (2012).

[171] B. F. Ngouana-Wakou, P. Cornette, M. Corral Valero, D. Costa, and P. Raybaud, An Atomistic Description of the \gamma-Alumina/Water Interface Revealed by Ab Initio Molecular Dynamics, Journal of Physical Chemistry C **121**, 10351 (2017).

[172] Q. Zhang, T.¸C¸Căg˘ g\imath ın, A. van Duin, W. A. Goddard, Y. Qi, and L. G. Hector, Adhesion and Nonwetting-Wetting Transition in the Al/ \alpha-Al_{2}O_3 Interface, Physical Review B **69**, 45423 (2004).

[173] M. F. Russo, R. Li, M. Mench, and A. C. van Duin, Molecular Dynamic Simulation of Aluminum–Water Reactions Using the ReaxFF Reactive Force Field, International Journal of Hydrogen Energy **36**, 5828 (2011).

[174] K. L. Joshi and A. C. T. van Duin, Molecular Dynamics Study on the Influence of Additives on the High-Temperature Structural and Acidic Properties of ZSM-5 Zeolite, Energy & Fuels **27**, 4481 (2013).

[175] F. G. Sen, A. T. Alpas, A. C. T. van Duin, and Y. Qi, Oxidation-Assisted Ductility of Aluminium Nanowires, Nature Communications **5**, 3959 (2014).

[176] K. L. Joshi, G. Psofogiannakis, A. C. T. van Duin, and S. Raman, Reactive Molecular Simulations of Protonation of Water Clusters and Depletion of Acidity in H-ZSM-5 Zeolite, Physical Chemistry Chemical Physics **16**, 18433 (2014).

[177] S. Hong and A. C. van Duin, Molecular Dynamics Simulations of the Oxidation of Aluminum Nanoparticles Using the ReaxFF Reactive Force Field, Journal of Physical Chemistry C **119**, 17876 (2015).

[178] N. Gunkelmann, E. M. Bringa, and Y. Rosandi, Molecular Dynamics Simulations of Aluminum Foams under Tension: Influence of Oxidation, Journal of Physical Chemistry C **122**, 26243 (2018).

[179] M. Ramírez, R. I. González, S. E. Baltazar, J. Rojas-Nunez, S. Allende, J. A. Valdivia, J. Rogan, M. Kiwi, and F. J. Valencia, Thermal Stability of Aluminum Oxide Nanoparticles: Role of Oxygen Concentration, Inorganic Chemistry Frontiers **6**, 1701 (2019).

[180] Y. Rosandi, H.-T. Luu, H. M. Urbassek, and N. Gunkelmann, Molecular Dynamics Simulations of the Mechanical Behavior of Alumina Coated Aluminum Nanowires under Tension and Compression, RSC Advances **10**, 14353 (2020).

[181] D. Chiche, C. Chizallet, O. Durupthy, C. Chanéac, R. Revel, P. Raybaud, and J.-P. Jolivet, Growth of Boehmite Particles in the Presence of Xylitol: Morphology Oriented by the Nest Effect of Hydrogen Bonding, Physical Chemistry Chemical Physics **11**, 11310 (2009).

[182] W. J. Mortier, S. K. Ghosh, and S. Shankar, Electronegativity-equalization method for the calculation of atomic charges in molecules, J. Am. Chem. Soc. **108**, 4315 (1986).

[183] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier, ACKS2: Atom-Condensed Kohn-Sham DFT Approximated to Second Order., Journal of Chemical Physics **138**, 74108 (2013).

[184] ReaxFF Force Field Format Specification, (n.d.).

[185] The Amsterdam Modeling Suite, (n.d.).

[186] S. Plimpton, Fast Parallel Algorithms for Short-Range Molecular Dynamics, Journal of Computational Physics **117**, 1 (1995).

[187] A. P. Thompson et al., LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales, Computer Physics Communications **271**, 108171 (2022).

[188] ParAMS 2023.101 Documentation, (n.d.).

[189] M. Gaus, Q. Cui, and M. Elstner, DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB), Journal of Chemical Theory and Computation **7**, 931 (2011).

[190] A. H. Larsen et al., The Atomic Simulation Environment—a Python Library for Working with Atoms, Journal of Physics: Condensed Matter **29**, 273002 (2017).

[191] W. Humphrey, A. Dalke, and K. Schulten, VMD: Visual Molecular Dynamics, Journal of Molecular Graphics **14**, 33 (1996).

[192] K. Kennes et al., Multiscale Visualization and Quantification of the Effect of Binders on the Acidity of Shaped Zeolites, Acs Catalysis **12**, 6794 (2022).

[193] J. P. Perdew, M. Ernzerhof, and K. Burke, Rationale for Mixing Exact Exchange with Density Functional Approximations, Journal of Chemical Physics **105**, 9982 (1996).

[194] J. Hafner, Ab-Initio Simulations of Materials Using VASP: Density-functional Theory and Beyond, Journal of Computational Chemistry **29**, 2044 (2008).

[195] P. E. Blöchl, Projector Augmented-Wave Method, Physical Review B **50**, 17953 (1994).

[196] S. N. Steinmann and C. Corminboeuf, Comprehensive Benchmarking of a Density-Dependent Dispersion Correction, Journal of Chemical Theory and Computation **7**, 3567 (2011).

[197] M. R. LaBrosse, J. K. Johnson, and A. C. T. van Duin, Development of a Transferable Reactive Force Field for Cobalt, Journal of Physical Chemistry a **114**, 5855 (2010).

[198] E. Iype, M. Hutter, A. P. J. Jansen, S. V. Nedea, and C. C. M. Rindt, Parameterization of a Reactive Force Field Using a Monte Carlo Algorithm, Journal of Computational Chemistry **34**, 1143 (2013).

[199] J. R. Boes, M. C. Groenenboom, J. A. Keith, and J. R. Kitchin, Neural Network and ReaxFF Comparison for Au Properties, International Journal of Quantum Chemistry **116**, 979 (2016).

[200] G. T. Kokotailo, S. L. Lawton, D. H. Olson, and W. M. Meier, Structure of Synthetic Zeolite ZSM-5, Nature **272**, 437 (1978).

[201] A. C. T. van Duin, J. M. A. Baas, and B. van de Graaf, Delft Molecular Mechanics: A New Approach to Hydrocarbon Force Fields. Inclusion

of a Geometry-Dependent Charge Calculation, J. Chem. Soc. Faraday Trans. **90**, 2881 (1994).

[202] N. Hansen, S. D. Müller, and P. Koumoutsakos, Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES), Evolutionary Computation **11**, 1 (2003).

[203] N. Hansen, The CMA Evolution Strategy: A Tutorial, (2023).

[204] P. Raybaud, C. Chizallet, and T. Pigeon, NOMAD Dataset: Gamma-Alumina Surfaces, (2021).

[205] I. Giagkiozis and P. Fleming, Methods for Multi-Objective Optimization: An Analysis, **293**, 338 (2015).

[206] Y. Tian, L. Si, X. Zhang, R. Cheng, C. He, K. C. Tan, and Y. Jin, Evolutionary Large-Scale Multi-Objective Optimization: A Survey, Acm Computing Surveys **54**, 1 (2021).

[207] P. Blanchard, D. J. Higham, and N. J. Higham, Accurately Computing the Log-Sum-Exp and Softmax Functions, Ima Journal of Numerical Analysis **41**, 2311 (2021).

[208] D. Palaz, G. Synnaeve, and R. Collobert, *Jointly Learning to Locate and Classify Words Using Convolutional Networks*, in *Proc. Interspeech 2016* (2016), pp. 2741–2745.

[209] R. Nian, J. Liu, and B. Huang, A Review On Reinforcement Learning: Introduction and Applications in Industrial Process Control, Computers & Chemical Engineering **139**, 106886 (2020).

[210] Gradient Bandit Algorithms, (2018).

[211] K. S. Brown and J. P. Sethna, Statistical Mechanical Approaches to Models with Many Poorly Known Parameters, Physical Review E: Statistical Physics, Plasmas, Fluids, And Related Interdisciplinary Topics **68**, 21904 (2003).

[212] J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna, Sloppy-Model Universality Class and the Vandermonde Matrix, Physical Review Letters **97**, 150601 (2006).

[213] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, Universally Sloppy Parameter Sensitivities in Systems Biology Models, Plos Computational Biology **3**, e189 (2007).

[214] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, Perspective: Sloppiness and Emergent Theories in

Physics, Biology, and Beyond, Journal of Chemical Physics **143**, 10901 (2015).

[215] T. Verstraelen, P. Bultinck, V. Van Speybroeck, P. W. Ayers, D. Van Neck, and M. Waroquier, The Significance of Parameters in Charge Equilibration Models, Journal of Chemical Theory and Computation **7**, 1750 (2011).

[216] O. Rahaman, A. C. T. van Duin, V. S. Bryantsev, J. E. Mueller, S. D. Solares, W. A. I. Goddard, and D. J. Doren, Development of a ReaxFF Reactive Force Field for Aqueous Chloride and Copper Chloride, Journal of Physical Chemistry a **114**, 3556 (2010).

[217] E. G. Goken, K. L. Joshi, M. F. Russo, A. C. T. van Duin, and A. W. Castleman, Effect of Formic Acid Addition on Water Cluster Stability and Structure, Journal of Physical Chemistry a **115**, 4657 (2011).

[218] X. Hu, J. Schuster, and S. E. Schulz, Multiparameter and Parallel Optimization of ReaxFF Reactive Force Field for Modeling the Atomic Layer Deposition of Copper, Journal of Physical Chemistry C **121**, 28077 (2017).

[219] L. W. Bertels, L. B. Newcomb, M. Alaghemandi, J. R. Green, and M. Head-Gordon, Benchmarking the Performance of the ReaxFF Reactive Force Field on Hydrogen Combustion Systems, Journal of Physical Chemistry a **124**, 5631 (2020).

[220] C. M. Daksha, J. Yeon, S. C. Chowdhury, and J. W. Gillespie Jr., Automated ReaxFF Parametrization Using Machine Learning, Computational Materials Science **187**, 110107 (2021).

[221] J. Yeon, S. C. Chowdhury, C. M. Daksha, and J. W. Gillespie, Development of Mg/al/Si/O ReaxFF Parameters for Magnesium Aluminosilicate Glass Using an Artificial Neural Network-Assisted Genetic Algorithm, Journal of Physical Chemistry C **125**, 18380 (2021).

[222] D. J. Evans and B. L. Holian, The Nose–Hoover Thermostat, Journal of Chemical Physics **83**, 4069 (1985).

[223] I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. C. Simm, R. Drautz, C. Ortner, B. Kozinsky, and G. Csányi, The design space of E(3)-equivariant atom-centred interatomic potentials, Nat. Mach. Intell. **7**, 56 (2025).

[224] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, E(3)-Equivariant Graph

Neural Networks for Data-Efficient and Accurate Interatomic Potentials, Nat. Commun. **13**, 2453 (2022).

[225] S. Vandenhaute, M. Cools-Ceuppens, S. DeKeyser, T. Verstraelen, and V. Van Speybroeck, Machine Learning Potentials for Metal-Organic Frameworks Using an Incremental Learning Approach, Npj Comput. Mater. **9**, 1 (2023).

[226] E. Tarasova, I. Korotkin, V. Farafonov, S. Karabasov, and D. Nerukh, Complete Virus Capsid at All-Atom Resolution: Simulations Using Molecular Dynamics and Hybrid Molecular Dynamics/Hydrodynamics Methods Reveal Semipermeable Membrane Function, J. Mol. Liq. **245**, 109 (2017).

[227] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models, J. Chem. Theory Comput. **6**, 459 (2010).

[228] J. Zhang and P. S. Branicio, Molecular Dynamics Simulations of Plane Shock Loading in SiC, Procedia Eng. **75**, 150 (2014).

[229] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, Achieving DFT Accuracy with a Machine-Learning Interatomic Potential: Thermomechanics and Defects in Bcc Ferromagnetic Iron, Phys. Rev. Mater. **2**, 13808 (2018).

[230] Y. Zuo et al., Performance and Cost Assessment of Machine Learning Interatomic Potentials, J. Phys. Chem. A **124**, 731 (2020).

[231] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, Machine Learning Prediction Errors Better than DFT Accuracy, J. Chem. Theory Comput. **13**, 5255 (2017).

[232] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, Quantifying the Carbon Emissions of Machine Learning, (2019).

[233] C. Schran, F. L. Thiemann, P. Rowe, E. A. Müller, O. Marsalek, and A. Michaelides, Machine Learning Potentials for Complex Aqueous Systems Made Simple, Proc. Natl. Acad. Sci. **118**, e2110077118 (2021).

[234] C. Schran, K. Brezina, and O. Marsalek, Committee Neural Network Potentials Control Generalization Errors and Enable Active Learning, J. Chem. Phys. **153**, 104105 (2020).

[235] N. Wilson, D. Willhelm, X. Qian, R. Arróyave, and X. Qian, Batch Active Learning for Accelerating the Development of Interatomic Potentials, Comput. Mater. Sci. **208**, 111330 (2022).

[236] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, The MLIP Package: Moment Tensor Potentials with MPI and Active Learning, Mach. Learn.: Sci. Technol. **2**, 25002 (2020).

[237] A. R. Tan, J. C. B. Dietschreit, and R. Gomez-Bombarelli, Enhanced Sampling of Robust Molecular Datasets with Uncertainty-Based Collective Variables, J. Chem. Phys. **162**, 34114 (2025).

[238] A. Zhu, S. Batzner, A. Musaelian, and B. Kozinsky, Fast Uncertainty Estimates in Deep Learning Interatomic Potentials, J. Chem. Phys. **158**, 164111 (2023).

[239] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, The Mahalanobis Distance, Chemom. Intell. Lab. Syst. **50**, 1 (2000).

[240] G. Seob~Jung, J. Youl~Choi, and S. Matthew~Lee, Active Learning of Neural Network Potentials for Rare Events, Digit. Discov. **3**, 514 (2024).

[241] Y.-C. Chen, A Tutorial on Kernel Density Estimation and Recent Advances, Biostat. \ Epidemiol. **1**, 161 (2017).

[242] K. D. Pandiangan, W. Simanjuntak, E. Pratiwi, and M. Rilyanti, Characteristics and Catalytic Activity of Zeolite-a Synthesized from Rice Husk Silica and Aluminium Metal by Sol-Gel Method, J. Phys.: Conf. Ser. **1338**, 12015 (2019).

[243] M. Sathupunya, E. Gulari, and S. Wongkasemjit, Na-A (LTA) Zeolite Synthesis Directly from Alumatrane and Silatrane by Sol-Gel Microwave Techniques, J. Eur. Ceram. Soc. **23**, 1293 (2003).

[244] T. P. Senftle et al., The ReaxFF reactive force-field: development, applications and future directions, Npj Comput. Mater. **2**, 15011 (2016).

[245] L. Komissarov, R. Rüger, M. Hellström, and T. Verstraelen, ParAMS: Parameter Optimization for Atomistic and Molecular Simulations, J. Chem. Inf. Model. **61**, 3737 (2021).

[246] M. Freitas Gustavo and T. Verstraelen, GloMPO (Globally Managed Parallel Optimization): A Tool for Expensive, Black-Box Optimizations, Application to ReaxFF Reparameterizations, J. Cheminformatics **14**, 7 (2022).

[247] M. Freitas Gustavo, M. Hellström, and T. Verstraelen, Sensitivity Analysis for ReaxFF Reparametrization Using the Hilbert–Schmidt Independence Criterion, J. Chem. Theory Comput. **19**, 2557 (2023).

[248] L. Komissarov, L. Krep, F. Schmalz, W. A. Kopp, K. Leonhard, and T. Verstraelen, A Reactive Molecular Dynamics Study of Chlorinated Organic Compounds. Part I: Force Field Development, Chemphyschem **24**, e202200786 (2023).

[249] L. Dumortier, C. Chizallet, B. Creton, T. de Bruin, and T. Verstraelen, Managing Expectations and Imbalanced Training Data in Reactive Force Field Development: An Application to Water Adsorption on Alumina, J. Chem. Theory Comput. **20**, 3779 (2024).

[250] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, PACKMOL: A package for building initial configurations for molecular dynamics simulations, J. Comput. Chem. **30**, 2157 (2009).

[251] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Structural Relaxation Made Simple, Phys. Rev. Lett. **97**, 170201 (2006).

[252] K. Xu, M. Zhang, J. Li, S. S. Du, K.-i. Kawarabayashi, and S. Jegelka, How Neural Networks Extrapolate: From Feedforward to Graph Neural Networks, (2021).

[253] G. R. Terrell and D. W. Scott, Variable Kernel Density Estimation, Ann. Stat. **20**, 1236 (1992).

[254] D. W. Scott, Scott's Rule, Wires Comput. Stat. **2**, 497 (2010).

[255] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Quantum-chemical insights from deep tensor neural networks, Nat. Commun. **8**, 13890 (2017).

[256] K. T. Schütt, O. T. Unke, and M. Gastegger, Equivariant message passing for the prediction of tensorial properties and molecular spectra, (2021).

[257] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A. G. Wilson, Averaging Weights Leads to Wider Optima and Better Generalization, (2019).

[258] I. Batatia et al., A Foundation Model for Atomistic Materials Chemistry, (2024).

[259] G. Bussi and M. Parrinello, Accurate Sampling Using Langevin Dynamics, Phys. Rev. E **75**, 56707 (2007).

[260] E. Paquet and H. L. Viktor, Molecular Dynamics, Monte Carlo Simulations, and Langevin Dynamics: A Computational Review, Biomed Res. Int. **2015**, 183918 (2015).

[261] G. Kresse and J. Furthmüller, Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set, Physical Review B **54**, 11169 (1996).

[262] G. Kresse and J. Furthmüller, Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set, Computational Materials Science **6**, 15 (1996).

[263] P. E. Blöchl, Projector Augmented-Wave Method, Phys. Rev. B **50**, 17953 (1994).

[264] G. Kresse and D. Joubert, From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method, Phys. Rev. B **59**, 1758 (1999).

[265] S. Grimme, S. Ehrlich, and L. Goerigk, Effect of the Damping Function in Dispersion Corrected Density Functional Theory, J. Comput. Chem. **32**, 1456 (2011).

[266] K. Chenoweth, A. C. T. van Duin, and W. A. Goddard, ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation, J. Phys. Chem. A **112**, 1040 (2008).

[267] A. C. T. van Duin, J. M. A. Baas, and B. van de Graaf, Delft Molecular Mechanics: A New Approach to Hydrocarbon Force Fields. Inclusion of a Geometry-Dependent Charge Calculation, J. Chem. Soc. Faraday Trans. **90**, 2881 (1994).

[268] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, ReaxFF: \, A Reactive Force Field for Hydrocarbons, J. Phys. Chem. A **105**, 9396 (2001).

[269] S. H. Hahn, J. Rimsza, L. Criscenti, W. Sun, L. Deng, J. Du, T. Liang, S. B. Sinnott, and A. C. T. van Duin, Development of a ReaxFF Reactive Force Field for NaSiOx/Water Systems and Its Application to Sodium and Proton Self-Diffusion, J. Phys. Chem. C **122**, 19613 (2018).

[270] A. P. Thompson et al., LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales, Comput. Phys. Commun. **271**, 108171 (2022).

[271] D. J. Evans and B. L. Holian, The Nose-Hoover Thermostat, J. Chem. Phys. **83**, 4069 (1985).

[272] G. J. Martyna, D. J. Tobias, and M. L. Klein, Constant Pressure Molecular Dynamics Algorithms, J. Chem. Phys. **101**, 10 (1994).

[273] W. Shinoda, M. Shiga, and M. Mikami, Rapid Estimation of Elastic Constants by Molecular Dynamics Simulation under Constant Stress, Phys. Rev. B **69**, 134103 (2004).

[274] Y. Ikeda, yuzie007/mpltern: 1.0.4, (2024).

[275] A. A. Hagberg, D. A. Schult, and P. J. Swart, *Exploring Network Structure, Dynamics, And Function Using NetworkX*, in *Proc. 7th Python Sci. Conf. (Scipy 2008)* (Pasadena, California, August 19–24, 2008), pp. 11–16.

[276] D. Furman and D. J. Wales, A Well-Behaved Theoretical Framework for ReaxFF Reactive Force Fields, J. Chem. Phys. **153**, 21102 (2020).

[277] B. Courty et al., mlco2/codecarbon: v2.4.1, (2024).

[278] A. Barducci, M. Bonomi, and M. Parrinello, Metadynamics, Wires Comput. Mol. Sci. **1**, 826 (2011).

[279] J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, and L. Delemotte, Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0], Living J. Comput. Mol. Sci. **4**, 1583 (2022).

[280] R. Barrett, J. C. B. Dietschreit, and J. Westermayr, Incorporating Long-Range Interactions via the Multipole Expansion into Ground and Excited-State Molecular Simulations, (2025).

[281] D. M. Anstine and O. Isayev, Machine Learning Interatomic Potentials and Long-Range Physics, J. Phys. Chem. A **127**, 2417 (2023).

[282] M. Cools-Ceuppens, J. Dambre, and T. Verstraelen, Modeling Electronic Response Properties with an Explicit-Electron Machine Learning Potential, J. Chem. Theory Comput. **18**, 1672 (2022).

[283] M. C. Kaymak, A. Rahnamoun, K. A. O'Hearn, A. C. T. van Duin, K. M. J. Merz, and H. M. Aktulga, JAX-ReaxFF: A Gradient-Based Framework for Fast Optimization of Reactive Force Fields, Journal of Chemical Theory and Computation **18**, 5181 (2022).

[284] S. B. Kylasa, H. M. Aktulga, and A. Y. Grama, PuReMD-GPU: A Reactive Molecular Dynamics Simulation Package for GPUs, Journal of Computational Physics **272**, 343 (2014).

[285] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler, Chemistry with ADF, Journal of Computational Chemistry **22**, 931 (2001).

[286] SCM - Accelerate Your Chemistry & Materials Research, (n.d.).

[287] P. Eastman et al., OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics, PLOS Computational Biology **13**, e1005659 (2017).

[288] P. Eastman et al., OpenMM 8: Molecular Dynamics Simulation with Machine Learning Potentials, The Journal of Physical Chemistry B **128**, 109 (2024).

[289] P. Eastman and V. Pande, OpenMM: A Hardware-Independent Framework for Molecular Simulations, Computing in Science & Engineering **12**, 34 (2010).

[290] Reproducible-Reporting/Parman: ParMan Extends Python Concurrent.Futures to Facilitate Parallel Workflows, (n.d.).

[291] Reproducible-Reporting/Stepup-Core, (2025).

[292] Reproducible-Reporting/Stepup-Reprep, (2025).

[293] L. Mädje, M. Haug, and The Typst Project Developers, Typst, (2025).

We acknowledge IFP Energies nouvelles for funding the first 3 years of the PhD program.

We acknowledge IFP School for providing expert training sessions.

We acknowledge the UGent for a BOF project fellowship (grant no. BOF/24J/2023/121) that funded the 4th year of the Phd program.

The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.