



Politecnico
di Torino

ScuDo

Scuola di Dottorato ~ Doctoral School

WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Physics (38th cycle)

Addressing the complexity of electrochemical interfaces via machine-learning-accelerated molecular dynamics

By

Michele Giovanni Bianchi

Supervisors:

Prof. Giancarlo Cicero, Supervisor

Dr. Michele Re Fiorentin, Co-Supervisor

Dr. Francesca Risplendi, Co-Supervisor

Doctoral Examination Committee:

Prof. Georg Kastlunger, *Referee*, Catalysis Theory Center, Department of Physics,
Technical University of Denmark, Denmark

Dr. Giacomo Melani, *Referee*, New Energies and Material Science Research Center,
Eni S.p.A, Italy

Politecnico di Torino

2026

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Michele Giovanni Bianchi

2026

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Acknowledgements

I would like to thank Professor Giancarlo Cicero for his support, patience, availability, and constant encouragement to overcome the many challenges encountered throughout my PhD journey. I also acknowledge Professor Michele Parrinello for sharing his expertise, for making the collaboration with his group possible and for hosting me at the Italian Institute of Technology – Genoa.

I am also very grateful to Dr. Luigi Bonati for his careful mentorship and continuous optimism, to Dr. Francesca Risplendi and Dr. Michele Re Fiorentin for their supervision since the beginning of my academic path, and to Dr. Federico Raffone for his continuous advice.

I also acknowledge the important role of all the researchers and friends in my groups at Politecnico di Torino and in the Atomistic Simulation group at the Italian Institute of Technology. My PhD experience would have been less enjoyable and complete without Anna C., Clara S., Enrico T., Lan H., Mattia S., Pedro J. B., Rhoy L. M., Shweta M., Simone P., Wei W., and many others.

My sincere thanks go to Prof. Georg Kastlunger and Dr. Giacomo Melani for their time and suggestions during the review of my thesis.

Finally, I wish to extend my special thanks to my family for its understanding and support during all my studies.

Contents

Introduction	1
1 Electrochemical interfaces: motivations and investigation approaches	5
1.1 The role of electrochemical interfaces in sustainable technologies . .	6
1.2 Metal/water interfaces in electrochemical cells	8
1.2.1 Working principles and applications of electrolyzers	10
1.2.2 The hydrogen evolution reaction	12
1.2.3 The CO ₂ reduction reaction	14
1.3 Study of the electrochemical interfaces	16
1.3.1 Canonical models	16
1.3.2 Experimental characterisation	19
1.3.3 Atomistic modelling	21
2 Methods: atomistic simulations	28
2.1 Density functional theory	29
2.1.1 The many-body problem and the Born-Oppenheimer approx- imation	29
2.1.2 Basics of DFT	30
2.1.3 The Hellmann-Feynman theorem	35
2.2 Molecular Dynamics	36

2.2.1	Classification of the molecular dynamics approaches	39
2.3	Enhanced sampling techniques for reactive events	42
2.3.1	CV-based formulation of the sampling problem	43
2.3.2	Altering the sampling probability distribution	44
2.3.3	Recovery of the properties of the unbiased distribution . . .	45
2.3.4	Well-tempered metadynamics	46
2.3.5	On-the-fly Probability Enhanced Sampling	47
2.3.6	Considerations on the collective variables	49
3	Methods: Machine learning interatomic potentials	51
3.1	Characteristics of a ML potential	51
3.1.1	Central role of the dataset	52
3.1.2	Data representation via descriptors	54
3.1.3	Architectures of the ML models	57
3.2	Approaches for the development of ML potentials	66
3.2.1	Quantification of the uncertainty	66
3.2.2	Active learning	67
3.2.3	Transfer learning	73
4	TRECI: transfer learning for electrochemical interfaces	75
4.1	Data-efficient workflow for electrochemical interfaces	76
4.1.1	High-quality description at multiple applied bias values . . .	77
4.1.2	ML force-fields via transfer learning	79
4.1.3	Data-efficient active learning	80
4.1.4	Complete strategy	81
4.2	Validation of the workflow	83
4.2.1	Preliminary step: model for the PZC	84

4.2.2	Starting point for the constant potential models	85
4.2.3	From stability to accuracy	86
4.2.4	From accuracy to convergence	87
4.2.5	Comparison of different training approaches for the constant potential models	89
4.3	Conclusion	92
5	Properties of electrified copper/water interfaces via ML-accelerated MD	94
5.1	Computational details	95
5.1.1	MD simulations	95
5.1.2	Generation of the ML models	97
5.2	Copper/water properties at PZC	98
5.2.1	Water density profile	98
5.2.2	Water molecule orientation	100
5.2.3	Hydrogen bond analysis	102
5.2.4	Water vibrational density of states	104
5.2.5	Ordering of the interfacial water structure	106
5.2.6	Evaluation of the PZC value	109
5.3	Copper/water properties at constant potential	111
5.3.1	Water density profile	111
5.3.2	Water molecule orientation	114
5.3.3	Hydrogen bond analysis	116
5.3.4	Water vibrational density of states	118
5.3.5	Analysis of the bond strength via critical point analysis . . .	121
5.3.6	Ordering of the interfacial water structure	123
5.4	Conclusion	125
6	Electrochemical reaction steps elucidated via ML-accelerated MD	128

6.1	Computational details	129
6.1.1	Generation of the ML models	129
6.1.2	MD simulations with enhanced sampling techniques	130
6.2	CO ₂ RR: the CO dimerisation step	131
6.2.1	CO diffusion	131
6.2.2	CO dimerisation: activation barriers and insights	133
6.3	HER: the Volmer step	139
6.3.1	H adatom diffusion	139
6.3.2	Volmer step: activation barriers and insights	141
6.4	Conclusion	145
Conclusion		148
References		153
Appendix A Double Reference Method		172
A.1	Considerations on the homogeneous background method	174
A.2	Considerations on the chain of potential references in the DRM	175
A.3	Code for the fully converged constant potential calculator via DRM	176
Appendix B Franken descriptors: characteristics, accuracy and usage		183
B.1	Comparison of the different Franken models	183
B.2	Models for the interface without reactive events	185
B.2.1	Comparative test on the descriptors and the number of random features	185
B.2.2	Comparative test of PZC and constant potential models	186
B.3	Models for the reactivity of the interfaces	188
Appendix C Assessments of the modelling framework		190

C.1	Locality of the ML models and long-range interactions	190
C.2	Effects of the DFT functional on the density profile	193

Introduction

Electrochemical technologies are the basis of a growing range of applications, including electrochemical manufacturing, heterogeneous catalysis, and, especially in the last few years, sustainable energy conversion and storage. For example, electrochemical processes can be employed to convert abundant resources such as water, carbon dioxide, and nitrogen into valuable products and fuels using renewable electricity, providing an attractive alternative to thermocatalytic and fossil-fuel-based routes. For these reasons, electrochemical technologies are central to the transition to a low-carbon, sustainable energy economy. Despite their broad appeal, many electrochemical technologies remain expensive and difficult to scale, mainly because the technology is still immature. Key issues are the low energy efficiency and poor control of the core electrochemical reactions. Achieving precise control of nanoscale processes at the electrode/electrolyte interface (e.g., metal/aqueous solution systems) raises a deeper and largely unresolved issue: a comprehensive understanding of the physics governing electrochemical interfaces. How does an external electric potential modify the electronic structure of the electrode surface and the configurations of adsorbed species? What is the response of the interfacial electrolyte to an applied bias? How are its structural, dynamical, and dielectric properties affected by surface charge accumulation? What role do the hydrogen bond network and, more generally, solvation shells play in determining the stability of the different species in the system? Addressing these fundamental questions is essential for understanding the interfacial phenomena in electrochemical systems and constitutes a major advance in surface physics.

In this respect, atomistic modelling approaches are the most common way to directly access a nanoscale picture of the electrochemical interfaces. *Ab initio* simulations at the atomic level allow for accurately capturing the charge transfer processes occurring at the electrode/electrolyte interfaces. Specifically, molecular dynamics

(MD) at the *ab initio* level is commonly regarded as one of the most common and comprehensive approaches to access a picture of the complex processes occurring in a dynamical environment such as solid electrode/liquid electrolyte interfaces, accounting for finite-temperature effects and characterising these phenomena from the statistical point of view. Nonetheless, the predictive capabilities of these simulations are limited by their high computational cost and by the difficulty of accurately capturing at once all the key elements that are at the basis of the physics of the electrochemical interfaces. Main challenges for MD in electrochemical systems regard, for example, a faithful description of the behaviour of the electrolyte via an explicit picture with a large number of solvent molecules and the simulation of the charge transfer processes in an electronically grand canonical ensemble where the electronic charge fluctuates to maintain the applied potential constant. These challenges are typically addressed by adopting a compromise between physical accuracy and computational tractability, which generally implies discarding more advanced, accurate and computationally demanding simulation protocols, such as the use of fully explicit representations of the solvent instead of implicit solvent solutions and robust constant potential schemes that do not require the presence of artificial interfaces with implicit solvent or vacuum regions.

In this thesis we address these challenges by developing machine learning interatomic potentials (MLIPs) to accelerate MD simulations of electrochemical interfaces, focusing on metal/water systems. By retaining the accuracy of the data computed via Density Functional Theory (DFT) over which the ML model is trained, but without the need to solve the original quantum problem, MLIPs drastically speed up MD simulations to a cost comparable with classical interatomic potentials.

Even if MLIPs are extremely promising, the requirement of a large amount of training data computed at the *ab initio* level can result in a high computational cost, particularly in electrochemical simulations. At present, studies employing MLIPs in electrochemical contexts remain constrained in their dataset generation by the DFT-level trade-offs between accuracy and computational feasibility, frequently omitting the explicit treatment of the applied potential and of any other element that require additional simulation efforts at the DFT level.

Here we follow a different paradigm: MLIPs are developed not only to overcome the spatial and time limits of *ab initio* MD but also to unlock more advanced but computationally expensive simulation setups. These include rigorous DFT calculations in an electronically grand canonical ensemble and a full explicit solvent

representation of the electrolyte at the meta-GGA exchange-correlation level. To address the problem of generating a dataset with this high level of accuracy, we propose TRECI (TRansfer learning for ElectroChemical Interfaces), a general and data-efficient workflow for developing and testing MLIPs for electrochemical interfaces. By leveraging different ML techniques, we are able to train stable and accurate MLIPs for metal/water interfaces with a few hundred configurations computed at the DFT level. This relatively low amount of required *ab initio* data allows for generating a small but high-level dataset.

We show the potentiality of the TRECI scheme by developing MLIPs for a representative electrochemical interface, a copper/water system, at several reducing potentials. From the comparison of low-index Cu surfaces in contact with water, we are able to revisit these systems with a level of accuracy and resolution not achievable with standard *ab initio* approaches and evidencing new structuring effects in the water network. The TRECI workflow, expanded with MD simulations and enhanced sampling techniques, also enables the development of MLIPs describing some selected key steps for reactions commonly occurring on Cu-based electrocatalysts, including hydrogen evolution and CO₂ electroreduction. This allows us to correlate the activation barriers of the reactions with the effects of the applied potential, surface charge and the structuring phenomena of the interfacial H₂O solvent. In this way, we do not only lay the foundations for future, more detailed works on the reactivity of Cu-based electrocatalysts, but we also present a complete example of how MLIPs, coupled with effective strategies such as TRECI, can markedly expand the comprehensions of electrochemical processes.

The results presented in this thesis were obtained within the framework of a collaboration between the *Material and Processes for Micro and Nano Technologies* at Politecnico di Torino and the *Atomistic Simulations* group, led by Prof. Michele Parrinello at the Italian Institute of Technology (IIT). The progress of the PhD project benefited from an active and continuous exchange of ideas between the two research teams, facilitated by the guest studentship of the PhD candidate in the *Atomistic Simulations* group under the supervision of Prof. Michele Parrinello and Dr. Luigi Bonati. They supervised the research activities primarily concerning the machine learning methodologies and enhanced sampling modelling aspects of the project.

The discussion of the thesis is organised as follows. Chapter 1 introduces the relevance of electrochemical interfaces, with a focus on metal/water systems for green energy transition technologies. This chapter also presents the main simulation challenges at the atomistic level for a full comprehension of electrochemical processes occurring in these systems. In the following, Chapter 2 introduces the fundamental concepts underlying the atomistic simulation techniques used in this work, highlighting the strengths and limitations of *ab initio* and force field approaches. Beyond this dualism, Chapter 3 presents the rapidly emerging field of MLIP models. Chapter 4 presents TRECI, our strategy that enables fast and data-efficient development of ML models tailored for electrochemical interfaces. The following chapters describe the application of TRECI to gain nanoscale insights into the behaviour of interfacial water near Cu (111) and (100) electrodes under applied potential (Chapter 5) and to elucidate some key reaction steps on this electrocatalyst (Chapter 6). In the last chapter, we summarise the key findings presented throughout the thesis and outline their possible implications for future works.

Chapter 1

Electrochemical interfaces: motivations and investigation approaches

This first chapter introduces the central topic of this thesis: electrochemical interfaces. An electrochemical interface is the boundary between an electron conductor, the electrode (typically a solid metallic material), and an ionic conductor, the electrolyte (such as an aqueous solution). These interfaces host complex mass and charge exchange/transport processes occurring across multiple length and time scales which are governed by redox reactions that mediate electron transfer between the electrode and the electrolyte. These reactions form the basis of electrochemical cells, devices that convert chemical energy into electrical energy (or vice versa) through spontaneous or externally driven reactions. In this chapter, we discuss the relevance of electrochemical interfaces in contemporary technological and environmental scenarios, with a focus on sustainable technologies (Section 1.1). We then examine electrolyzers as representative systems in which electrochemical reactions, such as hydrogen evolution and CO₂ reduction, occur, enabling the renewable production of a broad range of value-added chemicals using electricity as an energy source (Section 1.2). Considering the challenges associated with the poor energy efficiency and product selectivity of the current devices, the discussion will then move toward strategies for studying electrochemical interfaces to deepen the understanding and improve control over interfacial phenomena (Section 1.3). The chapter will focus on atomistic simulation approaches for studying electrochemical interfaces, since these

methods are the modelling framework employed in this thesis. The goal is to introduce the key concepts and present the state-of-the-art atomistic methodologies used in this field, highlighting common practices and current limitations. This background motivates the innovative machine learning-based strategy developed in this work, as well as the need for a data-efficient approach to constructing these models in order to describe electrochemical interfaces under conditions closer to experimental *in situ* characterisations.

1.1 The role of electrochemical interfaces in sustainable technologies

Electrochemical interfaces govern reactivity, charge transport, and stability across a broad range of natural and technological systems [1–3]. They play a central role in processes such as corrosion, lubrication, environmental chemistry, electrochemistry, and heterogeneous catalysis. Among electrochemical interfaces, metal/water interfaces are by far the most ubiquitous and technologically relevant [1]. They are central to industrial corrosion and corrosion-protection strategies, electrochemical manufacturing processes, and modern energy-conversion and energy-storage technologies such as fuel cells, electrolyzers, and electrochemical reactors.

The societal and economic impact of processes occurring at metal/water interfaces is particularly evident in the context of corrosion. Corrosion is fundamentally an electrochemical phenomenon driven by the dissolution of metals at their interface with aqueous environments [4]. It has been estimated that the global cost of corrosion amounts to approximately 2.5 trillion US dollars per year, and up to 15–35% of these costs could be avoided through the implementation of optimal corrosion prevention [5]. In this sense, a detailed understanding of the interfacial mechanisms that initiate and sustain corrosion can directly enable the development of more effective corrosion-resistant materials, resulting in remarkable economic benefits. Beyond traditional surface science and corrosion engineering, the importance of electrochemical interfaces has become even more pronounced in the 2020s due to their central role in addressing global challenges related to climate change and sustainable energy production (Figure 1.1) [7, 8]. Electrified metal/electrolyte interfaces host the electrochemical reactions that drive a wide range of low-carbon energy

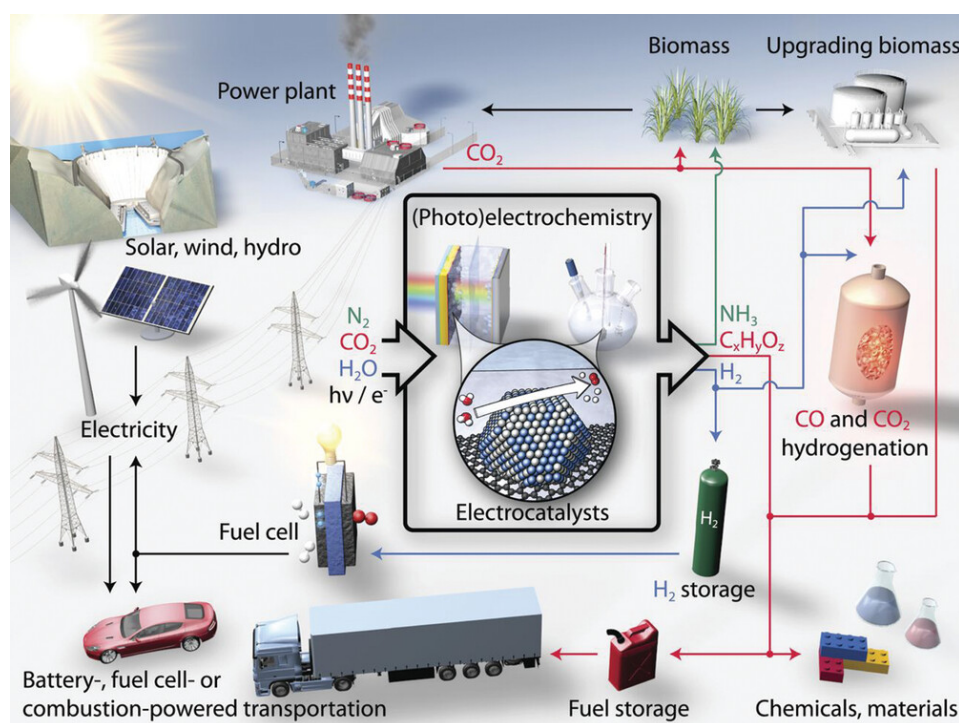


Fig. 1.1 Schematic of a renewable energy landscape related to electrocatalysis. Reproduced from [6]

technologies, including fuel cells, water electrolyzers, and electrochemical reactors for green fuel and chemical synthesis. These technologies enable the reversible conversion between electrical energy and chemical energy associated with bond formation and breaking during electrochemical reactions [9, 10]. When powered by renewable energy sources, electrochemical processes provide a sustainable route for the synthesis of fuels and chemicals, reducing dependence on thermocatalysis. The latter strategy typically involves a high energy demand, as it operates at elevated temperatures and pressures that are generally attained through the combustion of conventional fossil fuels, resulting in a high carbon footprint. Moreover, the resulting green fuels produced via electrocatalysis represent efficient energy vectors, enabling the storage and transport of renewable energy for use when and where direct generation is unavailable. Through the reactions occurring at electrochemical interfaces, electrolyzers and fuel cells emerge as key components of a sustainable and circular energy system and central pillars of current energy-transition strategies. In all these systems, the structure and dynamics of the electrified metal/liquid interface control key processes, including electron transfer, ion adsorption, surface reconstruction, and solvent reorganisation. Consequently, the identification, characterisation, and

control of electrochemical interfaces in these devices constitute key enabling elements for the development and large-scale deployment of electrochemical energy technologies. In this context a molecular-level understanding of these interfaces becomes a prerequisite for the rational design of materials and operating conditions to improve performance and long-term stability.

1.2 Metal/water interfaces in electrochemical cells

Electrochemical interfaces, and metal/water interfaces in particular, are at the heart of electrochemical cells, including fuel cells, electrolyzers, and, more generally, electrocatalytic reactors [3]. Despite their different configurations, these devices share a common architecture composed of the electrodes, the electrolyte and the separator (Figure 1.2) [8, 9, 11, 12]. Clearly identifying these components makes it possible to determine where metal/water interfaces appear within an electrochemical cell and which minimal elements must be considered to understand the interfacial processes. Because the present work focuses on a nanoscale understanding of electrochemical interfaces, we do not discuss the optimised designs and the details of the full electrochemical device. These aspects are extensively reported elsewhere [11–13]. Instead, we provide a brief description of the cell architecture to emphasise that the metal/water interface represents only a subregion of a much more complex system.

The electrodes, namely the cathode and the anode, are the conductive components across which an electric potential difference is generated or applied, depending on the operating mode of the device. They can be made of various materials, but they typically have excellent electrical conductivity and are often metallic or metal-supported. In many implementations, the electrodes consist of high-surface-area support materials on which metallic nanoparticles are dispersed. These nanoparticles contain the actual active sites in direct contact with the electrolyte, where electrochemical reactions take place. Such reactions involve electron exchange between species in the electrolyte and the metallic electrodes, resulting in redox processes. By convention, the electrode where reduction occurs is the cathode, and the electrode where oxidation occurs is the anode. The electrodes are electrically connected via an external circuit that allows for the electron flow of the redox, coupling the reduction and the oxidation reactions occurring at the opposite electrodes. The support materi-

als themselves do not participate in the redox reactions, but they provide chemical and mechanical stability while ensuring electron transport between the metallic active sites and the external circuit. Their high surface area maximises the exposure of metal particles to the electrolyte. Typical support materials include carbon-based structures (e.g., carbon black, graphene, and mesoporous carbons) for their high surface area or metal oxides (e.g., Ti_2O , Ce_2O , Zr_2O), which offer enhanced stability.

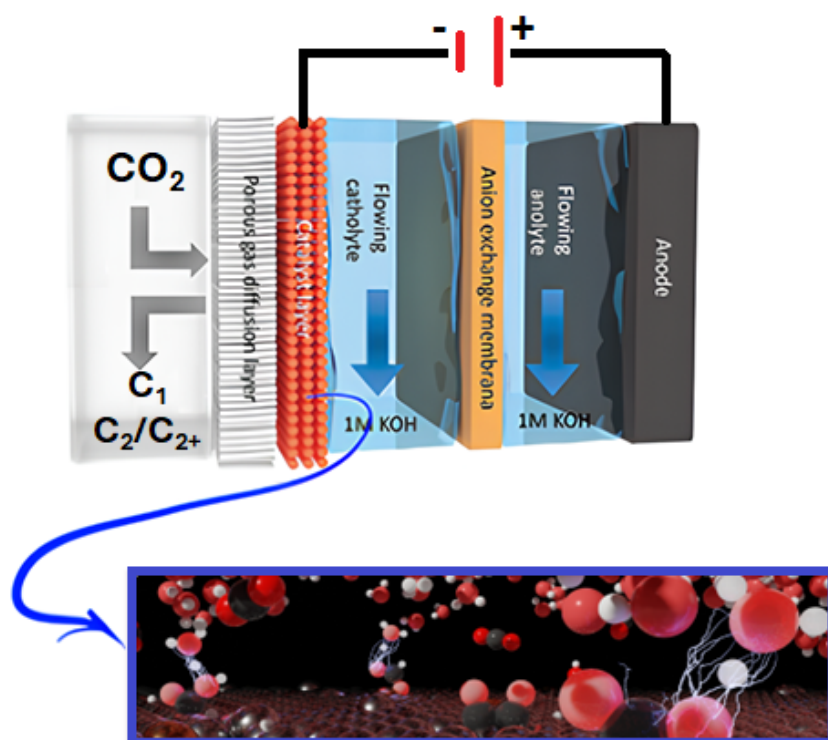


Fig. 1.2 Simplified scheme of a CO_2 electrolyser with the typical components of any electrochemical cell. The inset highlights the metal/water interface, the central region that will be studied in this thesis.

The electrolyte, located between the electrodes, provides the ionic conduction pathway that closes the circuit inside the device. While electron transport dominates within the electrodes, charge transport in the electrolyte occurs via ions. Electrolytes often consist of aqueous solutions where acids, bases, or salts dissociate into mobile anions and cations. Under an applied electric field, these ions migrate toward the oppositely charged electrodes, enabling ionic conduction. Common charge carriers in water-based electrolytes include H^+ , OH^- , K^+ , Na^+ , Cl^- , and SO_3^- ions. Depending

on the device, additional molecular species may be present near the electrode surfaces and participate directly in redox reactions: water (in fuel cells or electrolyzers), CO_2 , N_2 , or their corresponding derived products (in CO_2 or N_2 reduction reactors).

Electrochemical devices also include a separator or membrane that physically divides the electrolyte into two sub-cells associated with the anode and cathode. This barrier prevents the direct mixing of reaction products from the two electrodes while still allowing ionic transport. In addition to the discussed components, there are also other elements that are necessary for the device to work. A current collector is required to conduct electrons between the electrode active sites and the external circuit. Flow fields distribute reactants to the electrodes and remove reaction products to optimise mass transport within the cell.

Although a full electrochemical device requires the optimisation of all its components, its core part is the interfacial region between the electrode and the electrolyte, where the fundamental reaction steps occur. For this reason, the present study concentrates exclusively on this specific region. The full device spans length scales far beyond those accessible with the simulation methods employed in this thesis and is outside the scope of this work. Focusing on the electrode/electrolyte interface, we model the electrode solely as a metal surface, since the metallic component is the part directly exposed to the electrolyte and responsible for the electrochemical activity. The support material is therefore omitted from our nanoscale picture. Concerning the electrolyte, although many types exist (organic liquids, ionic liquids, solid or polymer materials), aqueous electrolytes are widely used due to their high ionic conductivity, low cost, and non-hazardous, environmentally friendly properties [12]. Accordingly, in this thesis we treat the electrode/electrolyte interface specifically as a metal/water interface.

1.2.1 Working principles and applications of electrolyzers

A typical class of electrochemical cells in which the electrode/electrolyte interfaces are metal/water interfaces is represented by electrolyzers. Even in systems where the electrolyte is not strictly aqueous, humidity plays a crucial role in the device operation, meaning that metal/water interfaces remain central to their working [9, 12]. Electrolyzers are electrochemical devices in which non-spontaneous reactions are driven by an external electrical energy source. They thus provide a technological

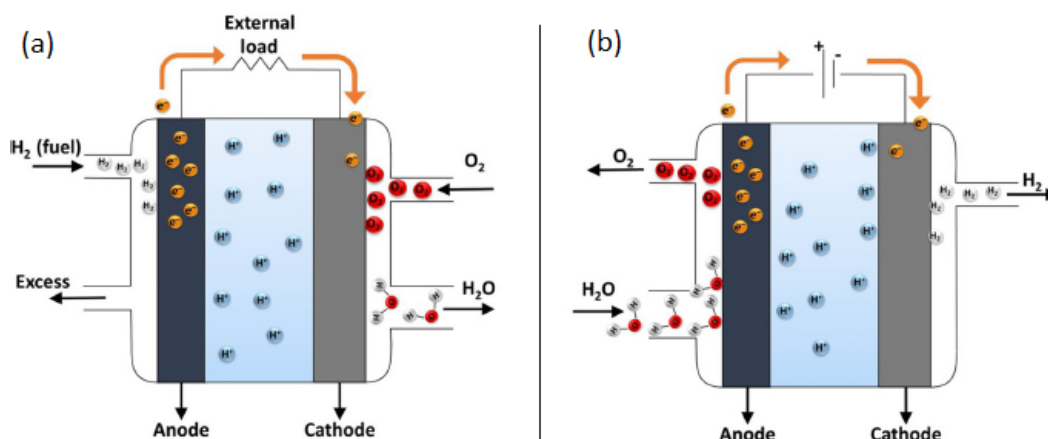


Fig. 1.3 Comparison of (a) a hydrogen fuel cell performing the exergonic conversion of hydrogen and oxygen to water and (b) a water electrolysis device for endergonic hydrogen production. Reproduced from [17].

route for activating chemical transformations using electrical, rather than thermal, energy [14, 15]. Electrolysers enable the synthesis of a variety of chemical products and are considered a promising and more environmentally sustainable alternative to conventional thermocatalytic processes. Electrolysers differ from other electrochemical devices, such as batteries and fuel cells, primarily in the role of the electrical energy source and in the direction of the redox reactions. Figure 1.3 illustrates the working principles of a hydrogen fuel cell and a water electrolyser [16, 17]. In a hydrogen fuel cell, molecular hydrogen is supplied at the anode, where it undergoes oxidation. The resulting H^+ ions migrate through the electrolyte to the cathode, where molecular oxygen is reduced to form water. The overall redox reaction produces a flow of electrons from the anode to the cathode through the external circuit, generating useful electrical power. In contrast, a water electrolyser operates in a reverse working condition. An external power source forces electrons to flow from the anode to the cathode. At the anode, water is oxidised to form O_2 and protons, while at the cathode, these protons are reduced to produce H_2 . In this configuration, the anode is held at a positive potential, and the cathode at a negative potential.

A water electrolyser represents just one example of an electrochemical cell designed to drive non-spontaneous redox reactions and synthesise new chemical species. While water electrolysers are used to produce molecular hydrogen from water via the hydrogen evolution reaction (HER) [18], other electrochemical systems target different transformations. For instance, some cells can convert CO_2 to a

variety of more reduced carbon-based products with higher value through the CO₂ reduction reaction (CO₂RR) [10, 19], or reduce N₂ to NH₃ via the N₂ reduction reaction (N₂RR) [20]. Among these processes, HER and CO₂RR are currently the two most studied and will be discussed in the following sections. Although N₂RR has attracted increasing attention as a more sustainable route to ammonia production [20], it remains the least developed of the three technologies and will not be examined further in this thesis.

1.2.2 The hydrogen evolution reaction

The hydrogen evolution reaction (HER) is one of the most representative and extensively studied electrochemical reactions [18, 21]. In this process, protons or water molecules are reduced at the cathode of an electrochemical cell to produce molecular hydrogen. The reaction is activated by applying an external potential and proceeds at the surface of the negatively biased cathode, which also acts as the catalyst. Typical cathode materials used to promote HER include precious metals such as platinum, as well as cheaper alternatives such as nickel and nickel-based catalysts [22]. Producing pure and green hydrogen using electrolyzers powered by renewable energy, without direct CO₂ emissions, plays a central role in contemporary energy transition scenarios [9]. Hydrogen may be used directly as a fuel or, more commonly, as a clean energy carrier to store surplus energy from intermittent renewable sources such as wind and solar. When renewable energy is abundant, it can be used to feed electrolyzers that produce molecular hydrogen, which is subsequently stored. During periods of energy deficit, the stored hydrogen is oxidised in a fuel cell, recovering the energy previously stored.

Given its importance, the HER mechanism has been the focus of extensive research. HER typically proceeds through two sequential proton–electron transfer steps¹ (Figure 1.4) [21, 23]:

- $\text{H}_2\text{O} + \text{e}^- + * \longrightarrow \text{H}^* + \text{OH}^-$ (Volmer step)
- $\text{H}_2\text{O} + \text{e}^- + \text{H}^* \longrightarrow \text{H}_2 + \text{OH}^-$ (Heyrovsky step)

where $*$ identifies the binding site on the metal surface. Alternatively, the second step might occur via the following:

¹formulation in neutral or alkaline medium

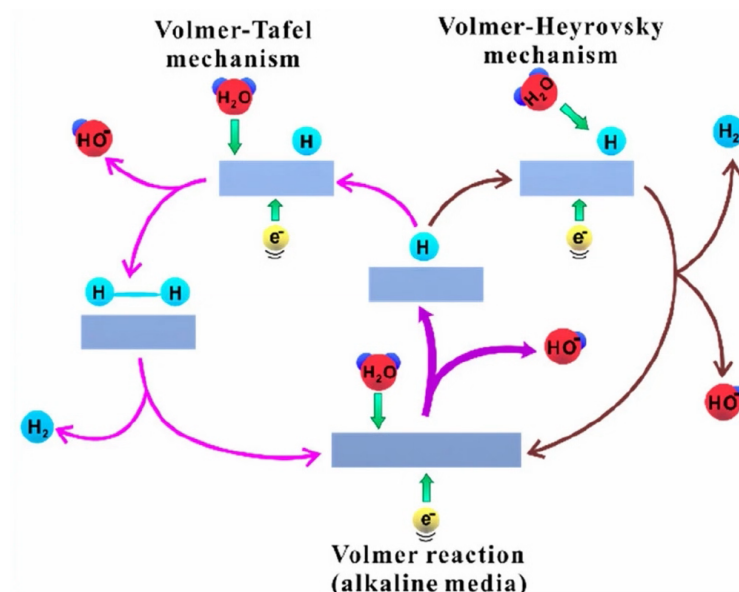
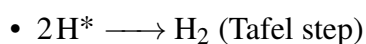


Fig. 1.4 Schematic diagram of HER mechanisms in alkaline media. Reproduced from [24].



In the latter case, H_2 originates from two neighbouring adsorbed hydrogens rather than from a new water dissociation event. The preferred reaction pathway (Heyrovsky vs Tafel) and the rate-limiting step depend strongly on factors such as the catalyst material, the crystalline orientation of the metal surface and the pH [23]. For instance, on platinum, the rate-limiting step is the Heyrovsky one on Pt(100) and the Tafel one on Pt(110) [18].

Understanding the HER mechanism is essential not only for optimising hydrogen electrolyzers but also for a broad range of electrocatalytic reactions carried out in aqueous media [19, 20]. Whenever an applied potential is used to drive a redox reaction in water, the presence of water molecules near the electrode inevitably leads to water splitting and H_2 formation, making HER an unavoidable competing reaction. Consequently, the design of electrocatalysts for selective production of a target chemical species must always incorporate strategies to suppress HER in order to maximise catalytic efficiency and selectivity.

1.2.3 The CO₂ reduction reaction

The electrochemical CO₂ reduction reaction (CO₂RR) is an electrochemical process that has attracted considerable attention in the context of energy transition and circular economy [10]. It enables the conversion of CO₂ into a wide range of value-added chemicals and fuels, including CO, methane, formic acid, ethylene, and ethanol, using electrical energy that can be supplied by renewable sources. As a result, CO₂RR simultaneously addresses two major challenges: it promotes the capture and conversion of CO₂, mitigating its release into the atmosphere, and it provides a pathway for the production of carbon-neutral fuels and industrial chemicals without relying on conventional fossil-based processes. In this context, methane, methanol, and ethanol can be used as renewable fuels, while CO, formic acid, and ethylene represent key products for green chemical manufacturing [25].

Unlike water electrolysis for the HER, CO₂RR technology is still under development and faces two major challenges: energy efficiency and product selectivity [10, 19, 25]. The high thermodynamic stability of the CO₂ molecule requires significant energy input to activate and break its double bonds, requiring high electrical bias values. In addition, CO₂RR can yield a broad spectrum of products. Rather than forming a single compound, the reaction typically produces a mixture of gaseous and liquid carbon-based products, as well as molecular hydrogen arising from the competing HER. This lack of selectivity increases the complexity and cost of subsequent product separation and purification.

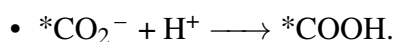
The complexity of CO₂RR is intrinsic in the multitude of possible reaction pathways that can occur following CO₂ activation. The reaction can begin with the reduction of CO₂ to adsorbed CO (Figure 1.5) [25]²:

- $\text{CO}_2 + \text{H}_2\text{O} + \text{e}^- + * \longrightarrow * \text{COOH} + \text{OH}^-$,
- $* \text{COOH} + \text{e}^- \longrightarrow * \text{CO} + \text{OH}^-$.

Beyond this more widely recognised mechanism, other research indicates that a hydrogen species already adsorbed on the catalyst surface may carry out the first hydrogenation step instead of a proton coming from solution [26]:

- $\text{CO}_2 + \text{e}^- + * \longrightarrow * \text{CO}_2^-$,

²formulation in neutral or alkaline medium



Nevertheless, it is still unclear if an adsorbed CO_2^- intermediate is truly present. Once CO adsorbed on the catalyst surface is formed, the reaction can proceed through multiple pathways leading to single-carbon compounds (C_1 products), two-carbon compounds (C_2), or even higher-carbon (C_{2+}) species.

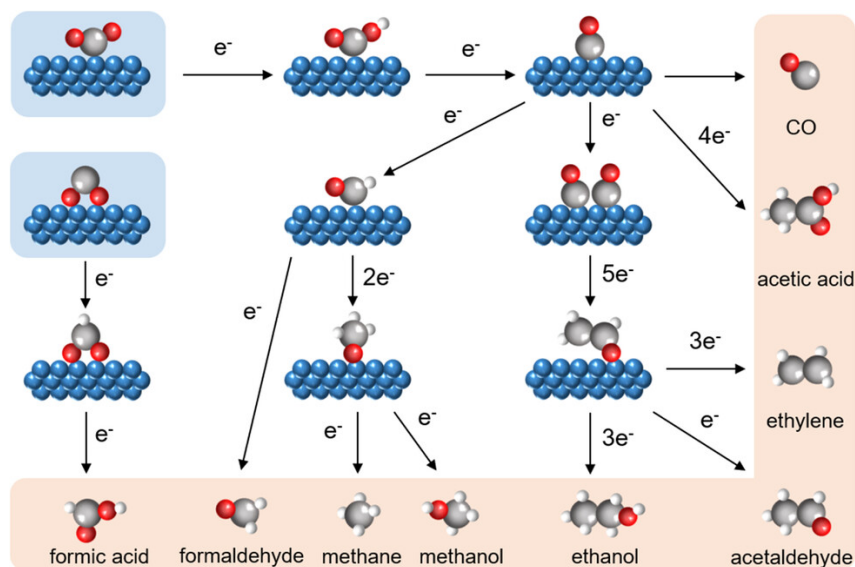
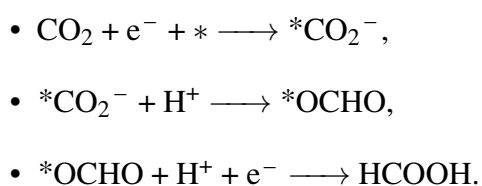
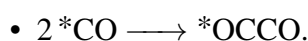


Fig. 1.5 Main reaction pathways for CO_2RR towards different products. Reproduced from [25].

In addition to the reaction mechanisms throughout the *CO intermediate, CO_2 can also be reduced to formic acid via the following reaction steps:



A schematic overview of these reaction pathways is shown in Figure 1.5. Despite the complexity of the reaction network, it is commonly recognised that the formation of C_2 products usually proceeds through the dimerisation of two adsorbed CO molecules, which establishes the crucial C–C bond:



For this reason, CO dimerisation is commonly studied in the optimisation of C₂ selective electrocatalysts and is often identified as the rate-limiting step of the overall reaction [19, 25].

Electrocatalyst materials play a key role in directing the CO₂RR toward specific products [27]. For instance, Ni, Fe, and Pt are generally ineffective for CO₂RR, as they predominantly catalyse HER. In contrast, Au, Ag, Pd, and Zn primarily yield CO, while post-transition metals such as Pb, Hg, In, Sn, and Bi favour the production of formic acid [28]. Among pure metals, copper is unique in its ability to produce significant amounts of C₂ and C₂₊ compounds, making it the most promising material for CO₂RR in complex carbon-based compounds [19, 28]. However, the multitude of reaction pathways following CO dimerisation presents many challenges in controlling product selectivity on Cu electrodes. Consequently, copper-based CO₂RR technologies remain limited by high energy requirements, catalyst stability issues, and poor selectivity. For these reasons, Cu-based electrocatalysts are the subject of extensive research aimed at overcoming these limitations. In this thesis, copper is, therefore, considered a representative metal electrode due to its central role in CO₂RR and the fundamental challenges associated with understanding and controlling its complex reaction mechanisms.

1.3 Study of the electrochemical interfaces

The discussed importance and ubiquity of electrochemical interfaces, and, in particular, of metal/water interfaces, has resulted in decades of scientific interest in these systems. These include efforts both in experimental characterisation and theoretical modelling, aiming at understanding the microscopic processes occurring at these surfaces [19, 29–34]. Nonetheless, the studies of electrified metal/water interfaces remain a major challenge in surface science, mainly due to the difficulty of directly probing the interfacial processes within a water-based electrolyte characterised by a various range of dynamical phenomena.

1.3.1 Canonical models

The conventional textbook description of the interface between a solid electrode and a liquid electrolyte is based on the works of Gouy, Chapman, Stern, and subsequent

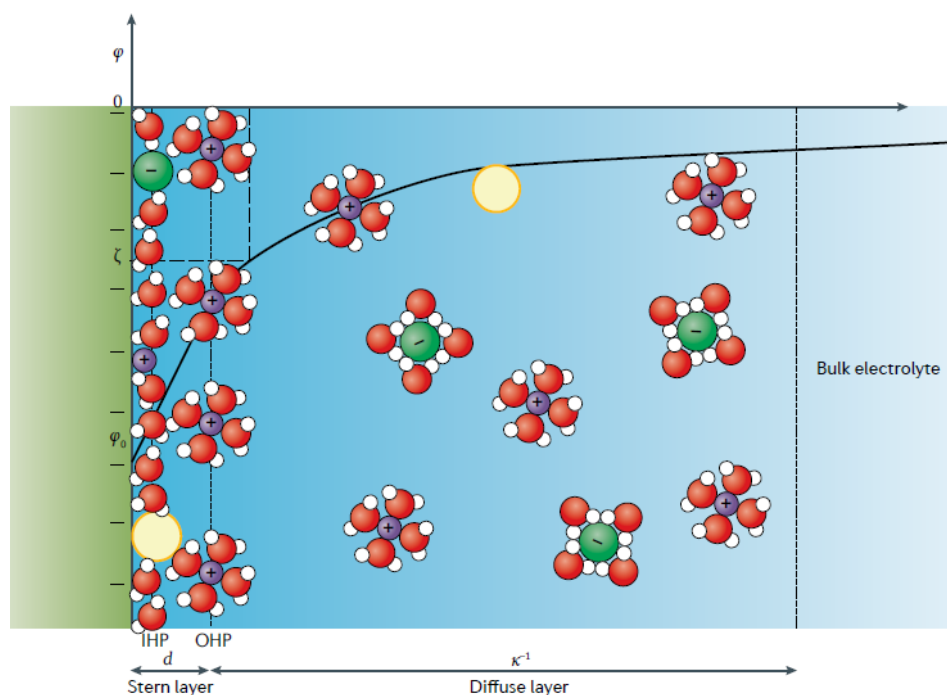


Fig. 1.6 Schematic description of the EDL structure. The picture represents the inner Helmholtz plane (IHP), the outer Helmholtz plane (OHP), the thickness of the Stern layer, the thickness of the diffuse layer and the potential profile ϕ as a function of the distance from the surface. The purple, green and yellow spheres represent the cations, the anions and neutral species, respectively. The hydration water around the ions as well as at the surface is explicitly represented. Reproduced from [2].

developments [35, 36]. These models established the key concept of the electric double layer (EDL), a complex interfacial region governed by the interactions among electrons, solvent molecules, and solute species (e.g., ions) over multiple spatial and temporal scales. The EDL forms in response to the charging of an electrode surface in contact with an electrolyte. Upon surface charging, ions and solvent molecules in the vicinity of the electrode rearrange their spatial distribution to restore electroneutrality. This reorganisation gives rise to a spatially varying electrostatic potential drop across the interface, which governs the energetics of species at the electrode/electrolyte boundary. The resulting potential profile reflects a complex balance between several competing effects, including the ordering imposed by the electric field, the entropic tendency associated with thermal motion, the screening action of polar solvent molecules, and long-range electrostatic interactions among charged species. Within this model, the EDL is commonly described as composed of two main regions (Figure 1.6) [36]: a compact layer, known as the Stern layer, located adjacent to

the electrode surface, and a diffuse layer that extends into the bulk electrolyte. The Stern layer is subdivided into the inner Helmholtz plane (IHP) and the outer Helmholtz plane (OHP). The IHP contains specifically adsorbed ions and partially desolvated molecules in direct contact with the electrode, which interact chemically with the surface. In contrast, the OHP is located slightly farther from the electrode and consists of fully solvated ions that are electrostatically attracted to the surface charge but remain separated from it by their solvation shells, preventing specific adsorption. Beyond the Stern layer lies the diffuse layer, where ions redistribute in response to the surface potential, and the electrostatic potential gradually relaxes toward its bulk value. Overall, the total charge stored within the EDL (Stern and diffuse layers) exactly compensates for the excess charge on the electrode surface, maintaining global charge neutrality. As a consequence, the structure and spatial extent of the EDL depend on numerous factors, including the surface charge density of the electrode, the size, the valence and the concentrations of the ions, and the dielectric and screening properties of the solvent. Depending on these quantities, the diffuse layer extends from approximately 1 nm far from the surface up to several hundred nanometres into the electrolyte, constituting the largest contribution to the EDL, since the Stern layer typically has a thickness below 1 nm [37].

This picture of the EDL is at the heart of the Gouy–Chapman–Stern model and subsequent developments [35, 36]. Despite the increasing complexity and completeness of these models, they share the same mean-field approach, where both the electrode and electrolyte are described as continuous media in which particles respond only to an average field generated by all other components. The strength of these models lies in their capacity to connect a macroscopic observable like the electrode potential to the electrode charge and the intricate interfacial structure of the EDL, which spans tens to hundreds of nanometres. Hence, it provides a conceptual framework for understanding the capacitive behaviour of electrode/electrolyte interfaces. Nevertheless, because these models rely on mean-field approximations, they inherently lack an explicit atomistic-level description of the interfaces, which is particularly important in the thin Stern layer. Key questions, such as how ions interact with each other and with the electrode surface while retaining their solvation shells, how water molecules reorganise and reorient to screen the surface charge, how reactants adsorb and react at the interface, and what local chemical environment governs interfacial reactions, cannot be fully answered within a mean-field model. These atomistic details are essential for an understanding of electrochemical processes in

electrolysers. Local electric fields, interfacial hydration, and hydrogen-bond networks influence proton availability, surface coverage, and the stabilisation of reaction intermediates, modulating reaction rates and product selectivity. For these reasons, extensive experimental investigations and atomistic modelling approaches have been developed over the years to achieve a more detailed description of electrochemical interfaces.

1.3.2 Experimental characterisation

From a characterisation perspective, accessing electrified interfaces buried within a liquid solvent requires advanced *in situ* and *operando* experimental techniques. Achieving nanoscale insight into these systems remains highly challenging, as the interfacial region comprises only a minute fraction of the total number of atoms relative to the bulk. Moreover, resolving its morphological and structural features demands extreme sensitivity and experimental control [38]. For these reasons, a wide range of complementary experimental techniques have been developed to characterise electrified metal/liquid interfaces.

Broadly, experimental studies of electrochemical interfaces can be classified into two main categories: investigations of interfacial water under ultra-high vacuum (UHV) conditions and *in situ* studies of electrode/electrolyte interfaces under electrochemical control [39]. In the first category, highly idealised systems consisting of atomically clean and well-defined metal substrates are examined under UHV conditions. These studies, employing techniques such as scanning tunnelling microscopy (STM), atomic force microscopy (AFM), and low-energy electron diffraction (LEED), aim to identify water adsorption sites, bonding motifs, and ordered overlayers on metal surfaces. Typically, only thin water films or small water clusters can be stabilised and characterised using these approaches [2, 38]. Despite their high spatial resolution, UHV-based studies suffer from intrinsic limitations when applied to water. Stabilising adsorbed water typically requires cryogenic temperatures (≤ 200 K), which suppress thermal fluctuations, freeze dynamical disorder, and eliminate entropic effects that are central to liquid-phase behaviour. Moreover, most UHV measurements are restricted to the first adsorbed water layer, rendering multilayer adsorption and the evolution of interfacial structures with increasing water coverage difficult to access [2]. As a result, the relevance of UHV studies to realistic electrochemical environments is inherently limited.

Experiments conducted at ambient temperature and pressure, under conditions more representative of operating electrochemical systems, are particularly informative. Interface-sensitive spectroscopic techniques, such as surface-enhanced infrared absorption spectroscopy (SEIRAS), surface-enhanced Raman spectroscopy (SERS), and vibrational sum-frequency generation (VSFG), enable direct probing of the structure, orientation, and hydrogen-bonding network of interfacial water as a function of electrode potential [1, 2, 39]. These methods provide experimentally accessible observables, including vibrational frequency shifts, intensity variations, and polarisation-dependent signals, from which information about the local interfacial environment can be inferred. In addition, electrochemical atomic force microscopy, surface-sensitive X-ray reflectivity and electron holography allow for atomic-scale mapping of surface morphology, reconstruction processes, and solvent layering under electrochemical control [40–44].

Despite this wide range of experimental tools, achieving a quantitative, nanoscale description of electrified metal/water interfaces remains an open challenge. Key difficulties associated with experimentally probing these systems include [39]:

- the limited selectivity of measurements in isolating signals originating from interfacial water from the overwhelming contribution of bulk water,
- the extreme low-scale sensitivity required to detect weak interfacial interactions, as hydrogen bonding and electrostatic interactions are often orders of magnitude weaker than chemical bonds,
- the potentially invasive or destructive nature of the experimental probes, which can locally perturb or disrupt the delicate interfacial structure.

As a result, obtaining quantitative and reproducible measurements under well-controlled conditions, particularly under applied potentials, remains difficult, complicating the unambiguous interpretation of interfacial processes. In this context, theoretical modelling is an essential complementary tool. By explicitly controlling system composition, geometry, and external conditions, theoretical approaches allow individual effects to be isolated and systematically analysed. Consequently, modelling can provide critical guidance for the interpretation of experimental observations. In particular, atomistic simulations offer sub-nanometre spatial resolution and access to dynamical and structural details of electrochemical interfaces that are

currently beyond experimental detection, enabling a deeper understanding of the microscopic mechanisms governing interfacial phenomena.

1.3.3 Atomistic modelling

As anticipated in Section 1.3.1, atomistic modelling, explicitly accounting for individual atom–atom interactions, can enrich the description of the EDL with respect to the Gouy–Chapman–Stern picture. By providing a discrete representation of the system, these approaches start from the position and chemical nature of the atoms and, given a physical model for the interatomic interactions, they are able to capture nanoscale phenomena with a much higher level of detail [17, 45, 46]. Owing to their atomic-scale resolution, atomistic methods are inherently limited in the maximum length scales they can access, typically ranging from a few Å up to several tens or, in some cases, hundreds of nanometres. As a consequence, only the near-surface region of the electrode is usually included in the model and represented as a slab consisting of a limited number of atomic layers [45, 47]. Furthermore, the full electrochemical cell with two electrodes is often not explicitly simulated; instead, studies commonly focus only on a half-cell. When both electrodes are included, the electrolyte thickness separating them is generally much smaller than that found in experimental systems [48]. In all cases, the modelling effort is focused on the region of the EDL closest to the electrode surface, i.e., the Stern layer and the innermost portion of the diffuse layer, while the full diffuse layer extending over hundreds of nanometres is typically not represented [49].

Depending on the accessible length scales, as well as the desired level of accuracy, atomistic approaches can be broadly divided into two main categories: *ab initio* methods and force field–based methods [50–53]. *Ab initio* approaches describe interatomic interactions by explicitly solving the electronic structure problem, accounting directly for the behaviour of electrons. As a result, these methods are well suited to describe interfacial processes involving electron transfer, bond formation and breaking, and chemical reactions, which are at the core of electrolyser working [54]. In contrast, force field methods rely on analytical, parametrised expressions to represent interatomic interactions. In these models, electronic effects are incorporated implicitly through the functional form of the force field and its fitted parameters, making the accuracy of the results strongly dependent on the quality and transferability of the chosen parametrisation. The fundamental difference in

how interatomic interactions are evaluated (solving a quantum mechanics problem versus computing a predefined force field) leads to different computational costs. Consequently, force field methods can access significantly larger system sizes than *ab initio* approaches, albeit at the expense of a reduced ability to describe electronic effects explicitly. A more detailed discussion of the differences between these two classes of atomistic modelling techniques, as well as their respective advantages, limitations, and relevance in the context of this thesis, is provided in Chapter 2. This dualism of the atomistic approach represents the starting point for a new class of simulations accelerated by machine learning techniques that represent the core of this thesis.

Regardless of the atomistic approach employed, modelling electrochemical interfaces remains intrinsically complex due to the heterogeneous nature of these systems, which comprise metal surfaces, solvents, and solutes, including both neutral and charged species that interact with one another [3]. From a statistical-mechanical perspective, describing a liquid in contact with a solid requires an explicit treatment of thermal fluctuations and an appropriate ensemble description [46, 53]. The highly dynamic nature of solvents such as water, where molecules continuously reorganise, solutes diffuse, and interfacial species respond to an externally applied potential, cannot be adequately captured by approaches restricted to absolute minimum-energy configurations. To capture these essential features of electrochemical interfaces under *operando* conditions, dynamical atomistic approaches such as molecular dynamics (MD) simulations are required [55]. MD provides a natural framework for sampling the configurational space of the solvent and interfacial species, explicitly accounting for thermal fluctuations, hydrogen-bond rearrangements, and transient interactions between solvent molecules, ions, and surface adsorbates. This time-dependent treatment is particularly important for understanding solvation structures and dynamic interfacial phenomena that directly influence electrochemical reactivity. Further details on MD simulations and their role in modelling electrochemical interfaces are provided in Chapter 2.

Approaches for the representation of the solvent

In addition to capturing the dynamic behaviour of the system, an accurate description of the solvent is essential for understanding interfacial processes [46]. Solvent/solute interactions play a central role in electrochemical systems: for instance, charged

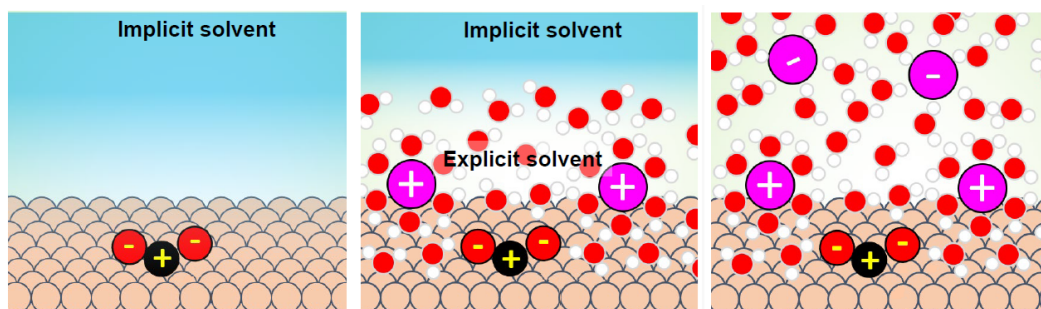


Fig. 1.7 Scheme for the different solvation approaches: from left to right, implicit, hybrid explicit-implicit and explicit solvent. Courtesy of Duy Le.

or polar reaction intermediates can be stabilised through hydrogen bonding with water. Studies of electrochemical interfaces that neglect the solvent, i.e., those performed in a vacuum, miss this fundamental aspect and often fail to reproduce realistic behaviours [56].

A common strategy to account for solvation effects is the use of implicit solvent models. Instead of explicitly representing individual solvent molecules, the electrolyte is coarse-grained into a continuous polarisable medium (Figure 1.7) [17, 45, 54, 57]. The widespread use of such approaches arises from their computational efficiency: they enable the simulation of electrochemical environments without the cost associated with treating large numbers of solvent molecules explicitly [46]. By replacing the discrete solvent with a dielectric continuum, they provide a simplified yet physically based description. Their efficiency and flexibility make them especially attractive for modelling extended systems such as charged metal interfaces. Moreover, many implicit schemes can also account for dissolved ions by modelling their charge and spatial distribution through the Poisson–Boltzmann equation [46, 54]. Despite their advantages, implicit models are intrinsically limited in their ability to capture short-range, directional interactions between solvent molecules and adsorbates. In particular, they cannot explicitly describe hydrogen bonding among water, intermediates, and the electrode surface, which systematically leads to an underestimation of the stabilisation of many adsorbed species [58, 59].

Explicit solvation strategies overcome these limitations by directly representing solvent molecules, capturing both the global dielectric response and the specific local interactions occurring at the interface [3, 46]. A practical compromise between accuracy and computational cost is offered by hybrid explicit–implicit schemes [57, 60]. In these approaches, a few layers of explicit water molecules, typically corresponding

to the first solvation shells, are placed near the electrode surface to reproduce local structural and hydrogen-bonding effects, while an implicit solvent description is used to model the bulk electrolyte above this region (Figure 1.7). However, hybrid methods are associated with well-known limitations: many continuum solvation models allow the implicit medium to penetrate regions in the explicit solvent that would be inaccessible to real water molecules, resulting in unphysical additional solvation of the explicit solvent [61, 62]. The presence of the artificial boundary between explicit and implicit domains may further perturb the properties of the explicit solvent, preventing it from recovering bulk-like behaviour far from the electrode surface [63, 64].

For these reasons, fully explicit solvent models, without vacuum gaps or implicit regions, remain the only rigorous way to capture simultaneously the dielectric response and the specific local interactions of the solvent without introducing artificial effects [3, 46]. This approach is also naturally suited for molecular dynamics simulations to optimise the solvent density, as the absence of vacuum or continuum regions prevents solvent molecules from drifting away from the explicit domain. Nevertheless, such simulations are often avoided due to their high computational cost, arising from the large number of atoms required to represent the solvating environment. The search for strategies enabling the use of explicit solvents while mitigating their computational cost, without introducing additional approximations, will be one of the central motivations driving the developments presented in this thesis.

Approaches for the representation of the applied potential

If a suitable description of the solvent is essential, an even more crucial component of atomistic models of electrochemical interfaces is the treatment of the external applied potential. Models that neglect the applied bias often yield unreliable predictions, since they omit the primary driving force that activates interfacial processes [17, 45, 46]. A key effect of the applied potential is the modification of the electrode surface charge. To reproduce the resulting variations in the electrostatic potential and local electric field near the interface, atomistic simulations typically alter the surface charge by adding or removing electrons [54, 57]. Two general strategies are commonly used to control the surface charge and thus model the applied potential [46, 47, 57]: constant charge schemes, which fix the total number of electrons, and constant

potential methods, in which the electron number is allowed to fluctuate to maintain a target electrode potential. The electrode potential U is typically defined with respect to a reference electrode, such as the Standard Hydrogen Electrode (SHE), according to [45, 46, 54]:

$$U = (-\tilde{\mu}_e + \Phi_{SHE})/e \quad (1.1)$$

where $\tilde{\mu}_e$ is the electrochemical potential of the electrons, often associated with the system work function [62, 65, 66], Φ_{SHE} is the potential of the SHE with respect to a vacuum, and e is the elementary electron charge.

In constant charge schemes, the electrification of the electrode is achieved by introducing ionisable species, such as alkali metals (K, Na, and Cs), halogens (F and Cl), or $\text{H}_3\text{O}^\bullet$ radicals, into the simulation cell (Figure 1.8a) [46, 60, 67]. These species readily donate or accept an electron with the surface, becoming charged and forming part of the EDL, while the electrode acquires a net surface charge. Adjusting the number of added species modulates the electrode charge. Despite their simplicity, such canonical-ensemble approaches have important limitations: because the number of electrons is fixed, the electrode potential cannot be held constant. Any electron transfer between the electrode and the electrolyte changes the work function, and therefore the electrode potential, during the simulation. This artificial potential drift leads to unphysical variations in reaction energetics and activation barriers, effects particularly severe in small simulation cells [46, 65]. To mitigate the size-dependent effects of constant charge methods, extrapolation schemes exist, but they are generally cumbersome and might become computationally demanding.

Constant potential methods overcome this limitation by treating the electron number as a fluctuating variable and fixing instead the electrochemical potential of electrons. This corresponds to modelling the system in the electronic grand canonical ensemble, where the electrode potential is externally controlled [46, 57]. These approaches repeatedly adjust the electron number and evaluate the resulting electrode potential (often via the work function), forming a feedback loop that enforces the target potential [60, 68, 69]. Because electrons are added or removed during this process, a charge-neutralisation mechanism is required. Under periodic boundary conditions, the presence of a net charge would lead to an infinite lattice of charged cells and a divergent Coulomb energy [46]

Many implementations achieve neutrality by coupling the constant potential scheme with an implicit solvent model (Figure 1.8b). Here, the electrolyte is rep-

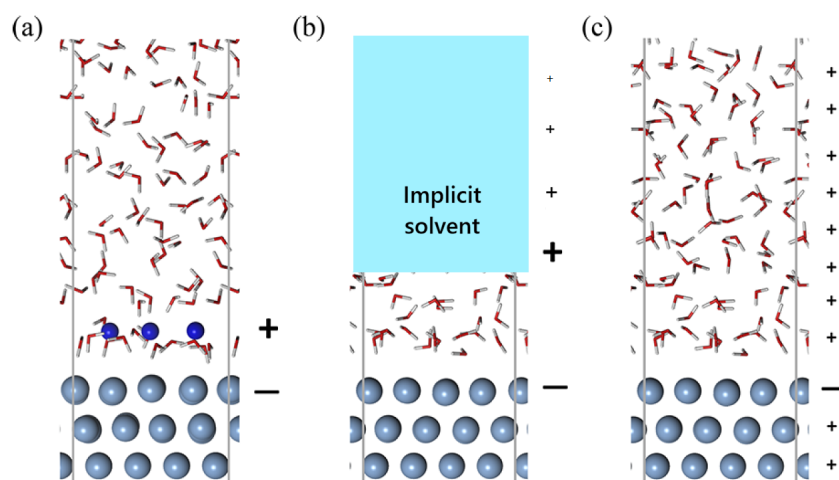


Fig. 1.8 Scheme for the different electrification approaches: (a) constant charge scheme with ions, (b) constant potential scheme with distributed countercharge in the implicit solvent region and (c) constant potential scheme with uniformly distributed countercharge. Adapted from [57].

represented as a polarisable dielectric continuum hosting a counterion distribution that responds to the local electrostatic environment, thereby screening the excess electrode charge [62, 69, 70]. As a result, the total electrostatic potential remains well-defined, and spurious interactions between periodic images are avoided. This feature enables simulations of polarised electrodes, where the metal slab carries a net charge, as well as systems containing explicitly charged species. In this case the electrolyte represented by the implicit model plays a dual role: it captures the physical screening behaviour of the electrochemical environment and ensures consistency in simulations of charged interfaces. However, these implementations inherit all limitations of implicit solvent models discussed earlier [64].

A more minimalistic strategy for maintaining charge neutrality is the use of a homogeneous background charge when electrons are added or removed from the electrode (Figure 1.8c) [66, 71]. Although such a background charge is unphysical, the presence of an explicit solvent has been shown to yield realistic potential profiles within the inner double layer, comparable to those obtained with explicit counterions [66]. This compensation scheme is employed in the double reference method, which provides a robust determination of the electrode potential [66]. In this method, a well-defined connection between surface charge and work function enables one to adjust the number of electrons to match a target applied potential within a grand

canonical framework. This approach allows for a fine control over the surface charge, similar to implicit solvent-based schemes, but without requiring a vacuum or implicit region to define the work function. Additional details are discussed in Appendix A. The main drawback of the double reference method is the need to evaluate auxiliary systems to obtain reference energy levels to rigorously define the work function. This comes at an increased computational cost. Consequently, despite its robustness and the absence of complex counterion placement, this method is seldom used, particularly in molecular dynamics simulations. The possibility of employing the double reference method in constant potential simulations for studying the dynamical effects of electrified interfaces will be one of the core motivations driving the developments presented in this thesis.

In summary, electrochemical cells, such as electrolyzers designed to generate added-value chemical products (e.g., H_2 and carbon-based molecules) from abundant compounds like water and CO_2 via the HER and CO_2RR , are the technological applications that motivate the focus of this thesis on electrochemical interfaces. The challenges of controlling energy efficiency and product selectivity in these systems demand a nanoscale understanding of the processes occurring at the electrode/electrolyte interface. In this thesis, we aim to provide an atomistic description of these complex environments, concentrating on the interplay between applied potential and interfacial reactivity. As a representative electrocatalyst for CO_2RR , we focus on copper electrodes in contact with a water-based electrolyte. The solvent and electric potential effects will be captured through a fully explicit solvation strategy and the constant potential double reference method, respectively. These features will be embedded in a dynamic picture provided by molecular dynamics simulations. The computational cost and complexity of simulating electrochemical interfaces will be faced by accelerating the molecular dynamics simulations via machine learning interatomic potentials. Machine learning potentials are a rapidly emerging field, combining the accuracy of *ab initio* simulations with the computational efficiency of force fields required for large-scale molecular dynamics simulations. The following chapters will present in detail the methodology associated with the machine learning potentials alongside our methodological improvement for the study of electrochemical interfaces and the efficacy of this strategy to study real-case scenarios.

Chapter 2

Methods: atomistic simulations

As discussed in the previous chapter, the thesis investigates the nanoscale processes occurring at electrochemical interfaces, deepening the knowledge on the effects of an externally applied potential and surface reactivity. To unravel these complex processes, atomistic simulations based on *ab initio* methods have emerged as indispensable tools, offering insights that complement and guide experimental observations. *Ab initio* methods derive their accuracy and predictive power from the fundamental principles of quantum mechanics, making them particularly suited for systems where electronic effects are critical. Among the various strategies for solving the many-body Schrödinger equation, Density Functional Theory (DFT) stands out as one of the most widely adopted frameworks. This thesis is entirely developed within the DFT paradigm, and all machine learning (ML) models presented herein are trained on DFT-generated data. Consequently, DFT serves as the reference (“ground truth”) for the ML predictions, and any inherent limitations or approximations in DFT will inevitably propagate into the ML results. Consequently, this chapter first revisits the essential principles of DFT (Section 2.1). However, static electronic structure calculations alone cannot fully capture the dynamic and complex nature of real surfaces. For this reason, Molecular Dynamics (MD) is introduced as an atomistic technique for exploring time-dependent phenomena at surfaces (Section 2.2). Bare MD is unable to cover the time-scale at which surface reaction processes occur, but enhanced sampling techniques are required to analyse the surface reactivity. This chapter concludes with an overview of these methods, focusing on the specific strategies adopted in this work (Section 2.3).

2.1 Density functional theory

2.1.1 The many-body problem and the Born-Oppenheimer approximation

Matter is characterised by the co-presence and reciprocal interaction among electrons and ions. The many-body problem of this system is described by the following Hamiltonian (in atomic units) 2.1:

$$\begin{aligned} \hat{H} = & -\frac{1}{2} \sum_i \nabla_i^2 - \sum_I \frac{1}{2M_I} \nabla_I^2 + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \\ & - \sum_{i,I} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}, \end{aligned} \quad (2.1)$$

where:

- M_I and Z_I are the mass and the atomic number of the I -th ion,
- $\mathbf{r}_i$ and $\mathbf{R}_I$ are the positions of the i -th electron and of the I -th ion, respectively.

The first and second terms in 2.1 are the sums of the kinetic energy contributions of each electron and each ion, respectively. The third contribution describes the interaction between electron pairs, the fourth considers the attraction between electrons and ions, while the last is associated with the repulsion of the ion couples. The complexity of this formulation is associated with the coupling of the electronic and ionic systems. For that reason, a series of approximations is required to deal with this many-body problem. The main approximation is the adiabatic or Born-Oppenheimer approximation [72]. The rationale behind this assumption is associated with the large difference between the velocity of electrons and ions, making the ions essentially stationary for the electrons, whereas ions only see a time-averaged potential (V_i) associated with the electronic system. According to this assumption, the Hamiltonian in 2.1 can be rewritten as:

$$\hat{H} = \hat{H}_{ions}(\{\mathbf{R}_I\}) + \hat{H}_e(\{\mathbf{r}_i\}, \{\mathbf{R}_{I0}\}) + \hat{H}_{e-ion}(\{\mathbf{r}_i\}, \{\delta\mathbf{R}_I\}). \quad (2.2)$$

The first term in 2.2 is the Hamiltonian associated only with the ionic system, and it explicitly depends only on the ion coordinates. The second term in 2.2 is the

electronic Hamiltonian:

$$\hat{H}_e(\{\mathbf{r}_i\}, \{\mathbf{R}_{I0}\}) = -\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} - \sum_{i,I} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_{I0}|}. \quad (2.3)$$

Notice that the electron-ion interaction in 2.3 considers only the condition in which ions are at their equilibrium position $\mathbf{R}_{I0}$, hence the ion coordinates become a parameter in the electronic Hamiltonian. The variations of the ion positions $\delta \mathbf{R}_I$ and the associated variations in the electron-ion interaction are encoded in the third term in 2.2, which is commonly described as the interaction between electrons and phonons.

The main advantage of the adiabatic approximation is the partition of the original problem into sub-parts that can be studied separately. In view of the electronic structure study, the electronic Hamiltonian in equation 2.3 becomes the central point of the following discussion. Despite the huge problem simplification due to the adiabatic approximation, the eigenproblem associated with equation 2.3 is still a many-body problem. Different approaches have been proposed to solve it; the DFT solution is described in the next section.

2.1.2 Basics of DFT

Density Functional Theory (DFT) is a method developed to describe correlated many-body systems. The original formulation is associated with the works of Hohenberg-Kohn [73] and Kohn-Sham [74]. The peculiarity of the DFT approach is the central role of the electron density $n(\mathbf{r})$. Indeed, if other methods identify the electronic wavefunction as the basic unknown in which the physical information is encoded, DFT replaces it with the electron density. In other words, all system properties are formulated as functionals of the ground state density.

Hohenberg-Kohn theorems

The Hohenberg-Kohn (H-K) theorems hold for a generic system of N interacting electrons in the presence of a time-independent local external potential $V_{ext}(\mathbf{r})$. According to the first H-K theorem, for **any** system of interacting particles in an external potential, $V_{ext}(\mathbf{r})$ is determined uniquely, except for a constant, by the ground state particle density $n_0(\mathbf{r})$. This theorem has relevant consequences; indeed, once

the ground state electron density $n_0(\mathbf{r})$ is known, a unique $V_{ext}(\mathbf{r})$ is determined up to an additive constant. Hence, the Hamiltonian is uniquely assembled given $n_0(\mathbf{r})$, and if the associated eigenproblem is solved, all wavefunctions and energy eigenvalues (both ground and excited states) can be evaluated. Therefore, all system properties are completely determined given only the ground state density. According to the H-K scheme, the ground state expectation value of any observable is a unique functional of n_0 . For example, the total energy $E[n]$ is expressed as:

$$E[n] = \int n(\mathbf{r}) V_{ext}(\mathbf{r}) d\mathbf{r} + F_{HK}[n]. \quad (2.4)$$

F_{HK} is universal since it is independent of the specific $V_{ext}(\mathbf{r})$, but includes only energy terms (electron kinetic energy and electron repulsion energy) that strictly depend on the electronic system alone.

The second H-K theorem states that the minimisation of the total energy functional $E[n]$ with respect to n allows for the identification of the ground state. The ground state energy E_0 is the minimum of $E[n]$ and the ground state density n_0 is the density associated with the minimum of $E[n]$. Hence, the minimisation of the total energy in equation 2.4 with respect to $n(\mathbf{r})$, imposing the constraint of the constant number N of electrons, is the path to follow to determine the exact ground state energy and density.

The Kohn-Sham *ansatz*

The H-K theorems can be considered an effective reformulation of the original many-body problem through functionals of the electron density. However, they do not solve the complexity of the many-body system since they do not specify any way to identify the functional required in equation 2.4. For that reason, the effectiveness of the H-K approach is associated with the Kohn-Sham (K-S) *ansatz* [74]. The basic idea of the K-S scheme is the association between the original, complex many-body system and a non-interacting electron system. The Hamiltonian of this auxiliary problem must be formulated so that the exact ground state density of the original many-body system can be represented by the ground state density of the associated system of non-interacting particles. Once the association between the original and the auxiliary problem is established, the application of the H-K approach is by far more convenient in the non-interacting electron system.

In the K-S formulation, the total energy functional $E[n(\mathbf{r})]$ in equation 2.4 can be rewritten as:

$$E[n(\mathbf{r})] = \int n(\mathbf{r}) V_{ext}(\mathbf{r}) d\mathbf{r} + T_s[n(\mathbf{r})] + \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[n(\mathbf{r})], \quad (2.5)$$

where $T_s[n(\mathbf{r})]$ is the kinetic energy functional of the non-interacting system, the third term is the electron density-density Coulomb interaction, and $E_{xc}[n(\mathbf{r})]$ is the exchange-correlation contribution.

The minimisation of equation 2.5 with respect to the density will provide the ground-state density of the non-interacting system. Equivalently, it is possible to prove (cf. ref. [50]) that the minimisation can be performed with respect to the electronic wavefunction ψ_i of the auxiliary K-S system, resulting in the solution of a Schrödinger-like single particle equation for the electronic wavefunction of the non-interacting system:

$$\hat{h}_{KS} \psi_i = \left[-\frac{1}{2} \nabla^2 + V_{KS}(\mathbf{r}) \right] \psi_i = \epsilon_i \psi_i, \quad (2.6)$$

where the Kohn-Sham potential $V_{KS}(\mathbf{r})$ includes the Coulomb potential (Hartree term), the external potential, and the exchange-correlation potential. Notice that equation 2.6 appears as the eigenvalue equation of a non-interacting electron system. This means that interacting electrons can be described as independent electrons in the presence of an effective potential V_{KS} . This is the key point of the K-S approach. Once the wavefunctions ψ_i of the non-interacting system are evaluated by equation 2.6, the associated electron density is finally computed as:

$$n(\mathbf{r}) = \sum_i^N |\psi_i(\mathbf{r})|^2. \quad (2.7)$$

Formally, once equation 2.6 is solved, the problem of the interacting electrons is sorted out since the ground state density of the auxiliary system can be directly associated with the original one, and then all other observables can be derived from $n_0(\mathbf{r})$. Practically, the many-body complexity is simply hidden inside an effective expression of V_{KS} and, more specifically, of $E_{xc}[n]$. Indeed, the exact correspondence between the original and auxiliary problems holds only if the correct expression for the exchange-correlation term is employed. Unfortunately, excluding the simple case of the uniform electron gas, exact forms for $E_{xc}[n]$ are unknown. So, even

if DFT is an exact theory, the result accuracy strongly depends on the employed exchange-correlation functional. For that reason, a complex hierarchy of functionals has been developed in the last decades, as discussed in the section below.

Exchange-correlation functionals

The association between the many-body system and the auxiliary independent-particle one is guaranteed only if the exact form of the exchange-correlation functional is known. Since the H-K and K-S equations do not provide any expression for $E_{xc}[n]$, only approximated functionals can be employed.

The simplest functional form is provided by the Local Density Approximation (LDA) [75, 76]. It was originally developed for a homogeneous electron gas where $n(\mathbf{r})$ is constant everywhere. The total exchange-correlation functional results in the sum of contributions of N equivalent electrons:

$$E_{xc}^{hom}[n] = N\epsilon_{xc}(n). \quad (2.8)$$

This idea can be extended to systems in which the density is not constant but depends on position. So, the exchange-correlation energy per particle $\epsilon_{xc}(n)$ expression, previously used for the homogeneous case, now depends on the local density. E_{xc} becomes a spatial average of $\epsilon_{xc}^{hom}(n(\mathbf{r}))$ weighted by the local density:

$$E_{xc}^{LDA} = \int n(\mathbf{r})\epsilon_{xc}^{hom}(n(\mathbf{r})) d\mathbf{r}. \quad (2.9)$$

The basic idea justifying the LDA approach is the quite short range of the effects of exchange and correlation. Nonetheless, this assumption does not hold in all systems: LDA is effective in systems resembling homogeneous electron gases (e.g., nearly free electron metals), and it works poorly in cases of inhomogeneous electron distribution. For that reason, the LDA idea is generalised to more complex expressions for E_{xc} as the Generalised Gradient Approximation (GGA) [77]. The non-locality is introduced in equation 2.9, considering that ϵ_{xc} depends not only on the local value of $n(\mathbf{r})$, but also on its gradient:

$$E_{xc}^{GGA} = \int n(\mathbf{r})\epsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r})) d\mathbf{r}. \quad (2.10)$$

Different strategies have been proposed to formulate $\varepsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r}))$, often modifying the original expression ε_{xc}^{hom} of the homogeneous electron gas. The simplest and most used expression for E_{xc}^{GGA} is the PBE (Perdew, Burke and Ernzerhof) form [78]. Despite the success and the popularity of GGA functionals, there are still several systems, including water, that are not accurately described by these functionals [79]. For that reason, a new class of functionals beyond GGA has been developed: the Meta-GGAs [50]. In addition to the density and its gradients, the Meta-GGAs are also functionals of the kinetic energy density $t(\mathbf{r}) = n(\mathbf{r})\tau(\mathbf{r})$, where τ is the kinetic energy density per particle, such that:

$$E_{xc}^{Meta-GGA} = \int n(\mathbf{r}) \varepsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r})) d\mathbf{r}. \quad (2.11)$$

There are different ways to formulate the kinetic energy density; among these, a possible definition is:

$$t(\mathbf{r}) = \frac{1}{2} \sum_{i=1}^N |\nabla \psi_i(\mathbf{r})|^2, \quad (2.12)$$

where the summation runs over the occupied states. Notice that the kinetic energy density is an implicit non-local functional of the density.

The most common Meta-GGA functional, which will also be employed in this work, is the "Strongly Constrained and Appropriately Normed" (SCAN) functional [80]. This functional is constructed in terms of the dimensionless variable:

$$\alpha(\mathbf{r}) = \tau_x(\mathbf{r}) / \tau^{unif}(n(\mathbf{r})) > 0, \quad (2.13)$$

where τ_x is the exchange kinetic energy density, and τ^{unif} is the kinetic energy density of the electron gas system with density $n(\mathbf{r})$. The specific form of the exchange energy as a function of n , ∇n and α was determined to satisfy all 17 known constraints of the true exchange-correlational functional accessible to a Meta-GGA (the "Strongly Constrained" in the acronym stands for this achievement). At the same time, the authors added in the SCAN formulation some appropriate norms from systems in which semilocal functionals are expected to be accurate (the "Appropriately Normed" in the acronym stands for this formulation) [80]. Even if bonded systems are not used to impose the norms, the SCAN formulation distinguishes among different types of interactions, including covalent single (for which $\alpha = 0$), metallic ($\alpha \approx 1$) and weak bonds ($\alpha \gg 1$). Indeed, it has been shown that SCAN is superior to LDA and common GGAs to describe different types of bonded materials,

including covalent, metallic, ionic, hydrogen, and van der Waals bonds [81]. Thanks to this capability, SCAN is able to carefully describe the structure of liquid water, with minimal over-structuring, since it is able to properly reproduce the relative strength of O-H covalent and hydrogen bonds [79]. At the same time, SCAN is also able to reproduce both the energies and the geometries of metallic systems [81]. Due to these capabilities, SCAN will be used in the following chapters to describe metal/water interfaces.

Considering the numerical instability and criticality in the convergence of the original SCAN version, a regularized formulation r^2 SCAN has been developed to improve the numerical performance, at the expense of breaking some constraints of the exact exchange-correlational functional [82]. Nonetheless, a subsequent reformulation of the regularised version, r^4 SCAN, recovering all the constraints of the original SCAN form [83], has been proved not to outperform r^2 SCAN, making the latter the handiest version within the SCAN family.

2.1.3 The Hellmann-Feynman theorem

The Hellmann-Feynman, or force theorem identifies an effective way to evaluate the force acting on nuclei in a system composed of interacting nuclei and electrons. The force acting on a nucleus I is associated with the gradient of the total energy with respect to the nucleus coordinates $\mathbf{R}_I$ as:

$$\mathbf{F}_I = -\nabla_{\mathbf{R}_I} E. \quad (2.14)$$

The total energy can be calculated through the evaluation of the expectation value of the system Hamiltonian $\hat{H}$:

$$E = \frac{\langle \psi(\mathbf{R}) | \hat{H}(\mathbf{R}) | \psi(\mathbf{R}) \rangle}{\langle \psi(\mathbf{R}) | \psi(\mathbf{R}) \rangle}. \quad (2.15)$$

Combining equations 2.14 and 2.15, it is possible to evaluate the force acting on the nucleus I . Nevertheless, the calculation of the gradient of the expectation value is not trivial since both the Hamiltonian $\hat{H}(\mathbf{R})$ and the system wavefunction $\psi(\mathbf{R})$ depend on the nucleus coordinates. The Hellmann-Feynman theorem [84, 85] proves that the gradient of the energy expectation value is equivalent to the gradient acting only on the Hamiltonian, if $\psi(\mathbf{R})$ is an eigenstate of $\hat{H}$. Hence, the force acting on the

nucleus I is simply:

$$\mathbf{F}_I = - \frac{\langle \psi | \nabla_{\mathbf{R}_I} \hat{H}(\mathbf{R}) | \psi \rangle}{\langle \psi | \psi \rangle}. \quad (2.16)$$

In other words, forces can be evaluated by knowing only the explicit dependence of each Hamiltonian term with respect to the nucleus coordinates.

This theorem has a central role since it offers an effective way to evaluate the forces acting on the atoms, relying on quantum mechanical principles. The knowledge of the forces is crucial for MD simulations to propagate the dynamics of the system over time. All the ML interatomic potentials will be trained on the forces evaluated through the Hellmann-Feynman theorem.

2.2 Molecular Dynamics

Molecular Dynamics (MD) is a computational technique that simulates the time-dependent behaviour of a system of atoms by numerically integrating their equations of motion. Unlike static approaches, which focus on finding minimum-energy configurations and neglect thermal fluctuations, MD inherently accounts for temperature effects on both atomic positions and dynamics. This makes MD particularly powerful for exploring properties that depend on atomic motion, such as diffusion, vibrational spectra, and phase transitions. By providing access to trajectories over time, MD enables the calculation of a wide range of observables that can be directly compared with experimental measurements [86].

The basics of MD are connected to the fundamental classical laws of dynamics [50]. Given a set of atoms (usually treated as classical particles) with positions $\mathbf{R}_I$, masses M_I and interaction potential energy $E[\mathbf{R}_I]$ (or, equivalently, the reciprocal interacting forces $\mathbf{F}_I$), the equations of motion according to Newton's law are:

$$M_I \ddot{\mathbf{R}}_I = - \frac{\partial E}{\partial \mathbf{R}_I} = \mathbf{F}_I[\{\mathbf{R}_I\}]. \quad (2.17)$$

This set of equations in the atomic positions is not solved analytically but numerically via a time discretisation strategy, commonly using the velocity-Verlet algorithm [87]:

$$\begin{aligned}
 \mathbf{R}_I(t + \Delta t) &= \mathbf{R}_I(t) + \mathbf{v}_I(t)\Delta t + \frac{\mathbf{F}_I(t)}{2M_I}\Delta t^2, \\
 \mathbf{v}_I(t + \Delta t/2) &= \mathbf{v}_I(t) + \frac{\Delta t}{2} \frac{\mathbf{F}_I(t)}{M_I}, \\
 \mathbf{F}_I(t + \Delta t) &= \mathbf{F}_I(\mathbf{R}_I(t + \Delta t)), \\
 \mathbf{v}_I(t + \Delta t) &= \mathbf{v}_I(t + \Delta t/2) + \frac{\Delta t}{2} \frac{\mathbf{F}_I(t + \Delta t)}{M_I},
 \end{aligned} \tag{2.18}$$

where $\mathbf{v}_I$ is the velocity of I -atom, and Δt is the time step defined in the discretisation approach. In this way, given the positions $\mathbf{R}_I(t)$ and the velocities $\mathbf{v}_I(t)$ of atoms at time t , it is possible to obtain their positions $\mathbf{R}_I(t + \Delta t)$ and velocities $\mathbf{v}_I(t + \Delta t)$ at the new time step $t + \Delta t$, knowing the forces acting on the atoms (cf. Section 2.2.1). Recursively applying equations 2.18, the atom motion is propagated for a defined time window (usually ranging in the scale of ps - ns).

Along the MD trajectory, a generic physical quantity A_k can be evaluated at every time step k (or a multiple of it) such that the time-averaged quantity $\bar{A}$ can be approximated as a discrete summation:

$$\bar{A} \approx \frac{1}{n_k} \sum_{k=1}^{n_k} A_k, \tag{2.19}$$

where n_k is the number of samples collected along the trajectory. Under the assumption of ergodicity, the average behaviour of the system over time (i.e., what can be obtained from an MD trajectory via equation 2.19) is equivalent to the average behaviour of many equivalent systems at a single point in time. Consequently, the ensemble average $\langle A \rangle$ of the physical quantity of a system with probability distribution $\rho(\mathbf{p}, \mathbf{q})$ (where $\mathbf{p}$ and $\mathbf{q}$ represent positions and momenta) can be computed as:

$$\langle A \rangle = \int d\mathbf{p} d\mathbf{q} \rho(\mathbf{p}, \mathbf{q}) A(\mathbf{p}, \mathbf{q}) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau dt A(\{\mathbf{R}_I(t)\}, \{\mathbf{v}_I(t)\}) = \bar{A}. \tag{2.20}$$

In this way an experimental observable quantity A_{exp} can be reasonably evaluated as a discrete time average $\bar{A}$ employing the statistics collected along an MD trajectory.

Thermostats

Equations 2.18 enforce energy conservation, so statistics are gathered in a micro-canonical (NVE) ensemble, rather than in a canonical (NVT) ensemble where the temperature is constant, as is typically the case under experimental conditions. In MD, the temperature is monitored by calculating the average kinetic energy (i.e., the velocities of the particles) [51]. According to the Maxwell-Boltzmann distribution, the average kinetic energy for each degree of freedom in a system with N particles is associated with the system temperature T via:

$$\frac{1}{2}M\overline{(\mathbf{v}^2)} = \frac{1}{6N} \sum_{i=1}^{3N} M_I v_I^2 \equiv \frac{k_B T}{2} \quad \text{with } k_B \text{ the Boltzmann constant.} \quad (2.21)$$

Temperature control is achieved by using a thermostat which adjusts the particle velocities, thereby ensuring the sampling of the canonical ensemble. Numerous thermostat schemes have been proposed, each relying on a distinct mechanism for regulating the velocities [86]. Here we briefly present the two methods most widely accepted for canonical ensemble sampling: the Nosé-Hoover thermostat and the "stochastic velocity rescaling" approach (also known as Bussi-Donadio-Parrinello). The Nosé-Hoover thermostat modifies the equations of motion, adding an extra "friction" force to increase or decrease the velocities to target a specific temperature [88]. An additional equation of motion is introduced to control the relaxation of the dynamics of the "friction" term and control the fluctuations of the instantaneous temperature with respect to the target temperature. In this way the temperature is not continuously fixed at a constant value (isokinetic ensemble), but its average value is kept constant as required in the canonical ensemble. However, it has been proved that the ergodicity of the system can be broken in the original formulation of the Nosé-Hoover thermostat [88]. This can occur in complex and heterogeneous systems, and this thermostat can generate unphysical temperature gradients within the simulation domain [89]. For this reason, the more elaborate Nosé-Hoover Chain formulation [90] is frequently necessary, although it entails greater computational expense and increased effort in identifying suitable parameters entering in the analytical formula defining the thermostat.

In addition to Nosé-Hoover, another simple but at the same time accurate thermostat is the "stochastic velocity rescaling" one [91]. Also in this case an extra force is added to the equations of motion to rescale the velocities, but with a stochastic

approach (in contrast to the deterministic one of the Nosé-Hoover). The velocities are rescaled to match a kinetic energy value extracted from a canonical equilibrium distribution, allowing the temperature to oscillate around a target value. The rescaling procedure is distributed among a number of time steps to allow a smooth evolution according to a set of equations for a stochastic dynamics [91]. It has been shown that, across a broad range of parameter choices, both the structural and most dynamical properties remain essentially unchanged by the specific thermostat settings, thereby enabling efficient sampling of the NVT ensemble with minimal artificial influence from the thermostat [91]. For that reason, this thermostat will be preferred in this thesis.

2.2.1 Classification of the molecular dynamics approaches

As visible in equations 2.18, a central element to propagate the motion is the knowledge of the force acting on the particles. Because the force must be calculated at every time step, it is clear that force evaluation accounts for most of the computational expense in MD. At the same time, the reliability of the observables that can be derived from an MD trajectory is strongly linked to how accurately these forces are computed. Consequently, force calculation in MD usually reflects a compromise between accuracy and computational cost and the different MD approaches are classified based on the strategies followed to evaluate the forces.

Ab initio molecular dynamics

The most accurate, but computationally expensive, way to compute the forces relies on an *ab initio* approach. This strategy is the so-called *ab initio* molecular dynamics (AIMD). In this technique the forces are obtained at each time step by solving a quantum mechanical problem for electrons: DFT is usually employed as the *ab initio* approach, and forces are evaluated through the Hellmann-Feynman theorem (cf. Section 2.1.3).

Even if forces are derived in a fully quantum mechanical approach, then the atoms are treated as classical particles, and they are moved following the laws of classical dynamics (Eqs. 2.18). This approach is valid thanks to the Born-Oppenheimer approximation, and this MD method is also called Born-Oppenheimer molecular dynamics (BOMD). Nonetheless, there are some systems with light species (e.g., protons) in which the Born-Oppenheimer approximation fails, and a fully

quantum approach is also required for these atomic species. In these cases, more advanced MD techniques are required, such as the path-integral molecular dynamics (PIMD) [92, 93].

AIMD is the most accurate and appropriate approach to describe systems characterised by chemical reactions, catalytic processes, and chemisorption events, i.e., all the processes in which the transfer or sharing of electrons are the essential elements to understand the physics of the problem [94]. This is because AIMD naturally includes the electronic degree of freedom. However, the solution of a quantum problem at each time step to compute the forces is computationally expensive. Hence, the potentiality of AIMD is hindered, limiting its applicability to small systems (up to a few hundred atoms) and short timescales (up to a few tens of picoseconds). This implies that results of AIMD can be affected by finite-size effects and poor statistical sampling.

Classical molecular dynamics

Classical MD is essentially the mirror image of AIMD: while it might not be as accurate, its computational efficiency is significantly higher. This is because the forces in classical MD are calculated by evaluating empirical force-fields (FFs), i.e., they map the potential energy surface (PES) as a function of the atomic coordinates. Since no quantum problem is solved on-the-fly at each time step, but forces are "simply" derived by evaluating a function, the force calculation is extremely efficient, allowing classical MD to explore quite large systems (several thousands of atoms) and long time scales (tens or hundreds of nanoseconds). The efficiency of empirical FFs is also associated with their "relatively" simple functional form: basic FFs reproduce pair interactions (e.g., the Lennard-Jones potential). More complete FFs include three-body or multi-body interactions and can reproduce long-range electrostatic attraction/repulsion [95]. These expressions contain a limited number of parameters that are fitted to reproduce some specific observables derived from empirical measurements or high-level *ab initio* calculations. The accuracy of the classical FFs is strongly connected to the functional form of the interatomic potential and its parametrisation. The FF design naturally restricts their flexibility and transferability because they typically have hand-crafted functional forms fitted to replicate specific properties. This frequently leads to inadequate accuracy for systems where the interaction is difficult to determine with an analytical expression

and/or to parametrise. Moreover, since classical FFs do not explicitly include the electronic degrees of freedom, classical MD is usually not accurate enough for systems in which the electronic contributions are predominant (e.g., for chemical reactions). It is possible to develop classical FFs specifically tailored for chemical reactions (the so-called Reactive Force Fields - ReaxFFs) [96], but their development is challenging, and, in general, they are characterised by poor transferability, making them not sufficiently accurate for systems far from the conditions in which they were parametrised. In this sense, AIMD remains the most reliable approach for reactive systems.

Machine-learning-based molecular dynamics

The best of classical and *ab initio* MD worlds can be bridged together in a new class of interatomic potentials: the machine-learning interatomic potentials (MLIPs) [52, 97]. MLIPs, also known as machine-learning force-fields (ML-FFs), are nominally classical-like FFs in the sense that forces are derived by evaluating a function that connects the atomic coordinates to the PES, without the need to directly solve a quantum problem. However, differently from the hand-crafted expressions of classical FFs, the MLIPs fit the PES through complex ML models, offering much greater flexibility through the large numbers of trainable parameters and without the need to define an *a priori* functional form for the FFs. Moreover, an MLIP is directly fitted on the forces entering into the equations of motion and not on some indirect physical quantities that derive from the forces. More precisely, an ML-FF is trained on a dataset of atomistic configurations, described at the *ab initio* level (i.e., DFT or at a higher level of theory), from which it learns to predict total energies, atomic forces, and, when required, other quantities such as virials, atomic charges [98], Wannier centroids [99], or atomic electronegativities [100]. In this way, ML potentials are able to fit the PES to predict the energy/forces of a new configuration of MD with an accuracy comparable to an *ab initio* approach but without the need to solve a quantum problem. For this reason, ML-FFs promise to combine the accuracy of *ab initio* methods with an efficiency and scalability comparable to the classical ones¹. Chapter 3 is fully devoted to a detailed analysis of ML models used to fit *ab initio* data and strategies to train and validate the MLIPs.

¹MLIPs implement more intricate functional forms compared to empirical ones, making them somewhat more computationally intensive, although they remain considerably more efficient than AIMD.

2.3 Enhanced sampling techniques for reactive events

Among interfacial phenomena occurring at the electrode/solvent interfaces, chemical reactions are one of the most investigated processes. A reaction and its kinetics are typically characterised via its activation barrier. According to the Transition State Theory, the rate constant k of a reaction can be computed as [101]:

$$k = \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger}{k_B T}}. \quad (2.22)$$

$\Delta G^\ddagger$ is the Gibbs free energy of activation, T is the temperature, and h the Planck constant. Therefore, the assessment of $\Delta G^\ddagger$ provides insights into an experimentally measurable quantity such as the rate constant and offers a means to compare the velocity of various reaction steps. In this thesis, the analysis will be restricted to the evaluation of $\Delta G^\ddagger$ values for the reactions under consideration, which will be employed as the principal metric for relative comparison. To achieve this objective, MD trajectories need to sample the basins of both reactants and products, as well as all possible pathways that connect them. This approach offers insights into the ensemble of high-energy configurations, the transition states (TSs), that separate the various basins. Unlike static methods such as the nudged elastic band or dimer method [46], which investigate the PES, MD directly gives access to the Free Energy Surface (FES). While the PES focuses solely on potential energy from interatomic interactions, the FES accounts for both potential and entropic contributions, offering a more thorough thermodynamic representation of the system [102].

Evaluating activation barriers using MD is a challenging task. Indeed, reactions occur on a timescale which is well beyond the capabilities of standard MD simulations. For example, electrocatalytic reactions can occur over timescales ranging from microseconds to milliseconds [103]. Consequently, initiating an MD simulation from a metastable state A will keep the system confined to that state for the entire duration of the simulation, because it does not have the possibility to cross the free energy barriers required to transition to another metastable state B . For that reason, MD is combined with enhanced sampling techniques designed to boost the occurrence of these rare transitions.

Here we will focus on a particular class of enhanced sampling methods that make use of collective variables (CVs) and introduce an external bias along these CVs,

thereby amplifying their fluctuations and reducing the free-energy barriers that hinder adequate sampling [104, 105].

2.3.1 CV-based formulation of the sampling problem

The problems of rare event sampling can be cast in terms of probability. In a canonical ensemble the probability associated with a potential energy $U(\mathbf{R})$, where $\mathbf{R}$ is the vector of the Cartesian coordinates of N atoms in the system, is evaluated through Boltzmann distribution:

$$p(\mathbf{R}) = \frac{e^{-\beta U(\mathbf{R})}}{Z} \quad \text{with the normalization factor} \quad Z = \int d\mathbf{R} e^{-\beta U(\mathbf{R})}, \quad (2.23)$$

where $\beta = 1/k_B T$.

Since $\mathbf{R} \in \mathbb{R}^{3N}$, it is convenient to formulate the problem in a smaller subspace of degrees of freedom, referred to as 'collective variables'. A CV is a coarse-grained coordinate, usually selected to be capable of distinguishing between relevant states. A set of CVs, denoted as $\mathbf{s}(\mathbf{R})$, is formulated as a collection of functions based on the atomic coordinates $\{\mathbf{R}\}$. In this way, the probability of observing the system in a state characterised by a specific $\mathbf{s}(\mathbf{R})$ value is:

$$p(\mathbf{s}) = \int d\mathbf{R} \delta(\mathbf{s} - \mathbf{s}(\mathbf{R})) p(\mathbf{R}). \quad (2.24)$$

The FES $F(\mathbf{s})$ as a function of the CVs is defined statistically, under the assumption of ergodicity, as:

$$F(\mathbf{s}) = -\frac{1}{\beta} \log[p(\mathbf{s})] \approx -\frac{1}{\beta} \lim_{t \rightarrow \infty} \log[N(\mathbf{s}, t)], \quad (2.25)$$

where $p(\mathbf{s})$ is the equilibrium distribution of the CVs, and $N(\mathbf{s}, t)$ is the normalised histogram obtained from sampling the MD trajectory. Once the FES is available, the energy difference between two metastable states and, more generally, the minimum free energy path connecting these states can be identified, making it possible to extract all the barriers that separate the states.

Accurate estimations of the FES via equation 2.25, require uniform samplings of $N(\mathbf{s}, t)$ over the relevant CV space within a finite simulation time, but this is hindered

by the presence of free energy barriers. For these reasons, we need to resort to enhanced sampling techniques to alter the sampling probability.

2.3.2 Altering the sampling probability distribution

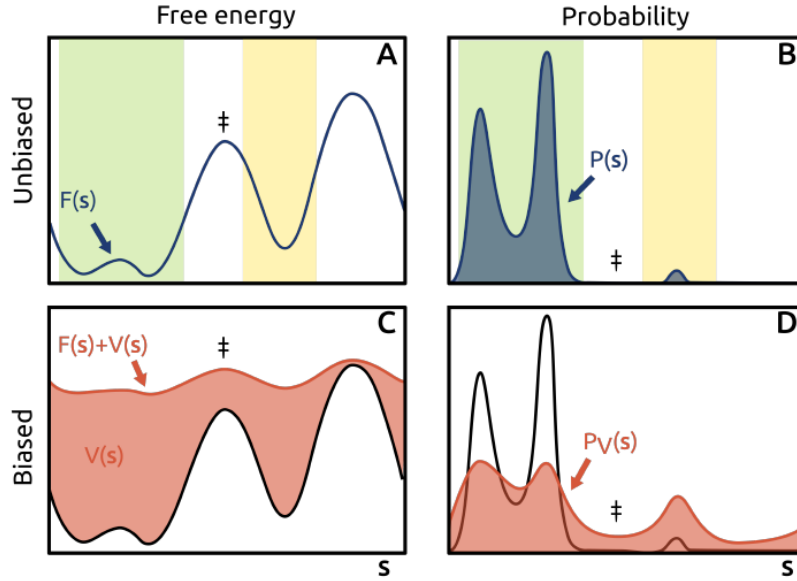


Fig. 2.1 Scheme of the relation between free energy (A, C) and probability distribution (B, D) along a collective variable. For an unbiased system with free energy $F(s)$ (top row), the probability distribution is strongly peaked on the metastable states, highlighted by the green and yellow shaded boxes, and almost negligible in correspondence to the free energy barriers associated with the TS ($\ddagger$). If a proper bias potential $V(s)$ is applied to the system (bottom row), the relative free energy barriers are lowered, and, as a consequence, the probability distribution is broadened, also covering the TS region. Reproduced from [106].

We focus on the family of enhanced sampling methods that use a bias potential $V(s)$ based on CVs. This potential, which is intended to improve sampling in under-visited phase space regions, depends on the CVs to discriminate among the different regions of the FES. In this manner, the system is sampled during a biased MD trajectory in accordance with the distribution in equation 2.26 rather than equation 2.23:

$$p_V(\mathbf{R}) = \frac{e^{-\beta(U(\mathbf{R})+V(s(\mathbf{R})))}}{Z_V} \quad \text{with} \quad Z_V = \int d\mathbf{R} e^{-\beta(U(\mathbf{R})+V(s(\mathbf{R})))}. \quad (2.26)$$

The bias is designed such that p_V is a smoother distribution in comparison to the original distribution, facilitating the evaluation of the normalised histogram $N_V(s, t)$

during the biased trajectory. Nonetheless, the resulting biased MD does not give access to the original unbiased FES $F(\mathbf{s})$, but to a biased one defined as:

$$F_V(\mathbf{s}) = F(\mathbf{s}) + V(\mathbf{s}). \quad (2.27)$$

2.3.3 Recovery of the properties of the unbiased distribution

A crucial element is the possibility of computing the properties of the original unbiased distribution from the sampled statistics of the biased MD trajectory. Given a generic observable $O(\mathbf{R})$, the average ensemble observable in the original distribution is:

$$\langle O(\mathbf{R}) \rangle = \int d\mathbf{R} O(\mathbf{R}) p(\mathbf{R}). \quad (2.28)$$

The same quantity can be formulated by introducing the biased probability distribution:

$$\langle O(\mathbf{R}) \rangle = \int d\mathbf{R} \left(O(\mathbf{R}) \frac{p(\mathbf{R})}{p_V(\mathbf{R})} \right) p_V(\mathbf{R}) = \frac{\langle O(\mathbf{R}) e^{\beta V(\mathbf{s}(\mathbf{R}))} \rangle_V}{\langle e^{\beta V(\mathbf{s}(\mathbf{R}))} \rangle_V}, \quad (2.29)$$

where $\langle \cdot \rangle_V = \int d\mathbf{R} \cdot p_V(\mathbf{R})$ is the ensemble average in the biased distribution.

Supposing ergodicity holds, the ensemble average can be connected with the discrete time average, such that equation 2.29 becomes:

$$\langle O(\mathbf{R}) \rangle \approx \sum_k O(\mathbf{R}_k) \frac{e^{\beta V(\mathbf{s}(\mathbf{R}_k))}}{\sum_k e^{\beta V(\mathbf{s}(\mathbf{R}_k))}} = \sum_k O(\mathbf{R}_k) w(\mathbf{R}_k), \quad (2.30)$$

in which the sampled observables $O(\mathbf{R}_k)$ extracted from a biased MD trajectory are multiplied by normalised weights $w(\mathbf{R}_k)$ to be associated with the corresponding samples of the unbiased distribution. This procedure is named "reweighting", and it is valid only if the bias $V(\mathbf{s}(\mathbf{R}_k))$ is static (or, practically, it can be approximated as static), i.e., it is known and time-independent. Employing equation 2.30, it is also possible to evaluate the unbiased probability distribution of equation 2.24 via reweighting:

$$p(\mathbf{s}) \approx \sum_k^{n_k} \delta(\mathbf{s} - \mathbf{s}(\mathbf{R}_k)) w(\mathbf{R}_k). \quad (2.31)$$

Similarly, it is also possible to have access to the unbiased FES from the biased trajectory. Combining equations 2.25 for the biased case and 2.27, one obtains:

$$F(\mathbf{s}) = F_V(\mathbf{s}) - V(\mathbf{s}) = -\frac{1}{\beta} \log[N_V(\mathbf{s})] - V(\mathbf{s}), \quad (2.32)$$

where $N_V(\mathbf{s})$ is the histogram collected during the biased MD trajectory in which the barriers have been lowered [105].

2.3.4 Well-tempered metadynamics

Various enhanced sampling methods are based on these general principles, differing in how the bias potential is constructed. The target distribution p^{tg} , i.e., the distribution that one would like to sample during the biased trajectory, is usually selected as a smoother version of the original distribution to make the rare events easier to observe. At the same time, it should preserve a good overlap with the original distribution to avoid the sampling of highly unfeasible CV spaces [104].

For example, in the well-tempered metadynamics (WT-MetaD) [107], the target distribution is:

$$p^{tg}(\mathbf{s}) = \frac{[p(\mathbf{s})]^{1/\gamma}}{\int d\mathbf{s} [p(\mathbf{s})]^{1/\gamma}}, \quad (2.33)$$

while the associated bias potential $V(\mathbf{s})$ is:

$$V(\mathbf{s}) = -\left(1 - \frac{1}{\gamma}\right) F(\mathbf{s}). \quad (2.34)$$

γ is a positive parameter greater than 1 that is used to smooth the original distribution p , increasing the fluctuations of the CVs and lowering the barriers [105]. Indeed, the limit $\gamma \rightarrow 1$ represents the unbiased ensemble, while for $\gamma \rightarrow \infty$, $V(\mathbf{s}) \rightarrow -F(\mathbf{s})$ and p^{tg} tends to a uniform distribution where all the barriers have been smoothed out. However, the latter case is not optimal since it leads to an uncontrolled exploration of the full FES, also in highly unphysical regions. For that reason it is important to tune γ to preserve a good overlap with the original distribution.

As visible in equation 2.34, the definition of the bias requires the knowledge of the FES but $F(\mathbf{s})$ is not known in advance. For that reason, the bias is constructed in an adaptive manner through an iterative scheme. In WT-MetaD, one iteratively builds a bias potential that asymptotically satisfies equation 2.34 [107]. This is possible

through a recursive addition at every τ_G time interval of repulsive biasing kernels $G(\mathbf{s} - \mathbf{s}')$, centred in the current $\mathbf{s}'$ value of the CVs:

$$G(\mathbf{s} - \mathbf{s}') = W e^{-\frac{1}{2} \Sigma_{i,j} (\mathbf{s}_i - \mathbf{s}'_i) \Sigma_{ij}^{-1} (\mathbf{s}_j - \mathbf{s}'_j)}, \quad (2.35)$$

where Σ_{ij} is a diagonal matrix where the diagonal element σ_i is the width of the Gaussian for the i -th CV. In this way, at time t , after adding n Gaussian kernels, the total bias is:

$$V(\mathbf{s}, t) = \sum_{k=1}^n G(\mathbf{s} - \mathbf{s}_k) e^{\left[-\frac{1}{\gamma-1} \beta V_{k-1}(\mathbf{s}_k)\right]}, \quad (2.36)$$

where each Gaussian kernel is modulated by the exponential term $e^{\left[-\frac{1}{\gamma-1} \beta V_{k-1}(\mathbf{s}_k)\right]}$, which scales as $\frac{1}{n}$. In this way the variations of the bias along the trajectory become progressively smaller, reaching a quasi-static condition.

Even though WT-Metad has significantly improved upon earlier metadynamics formulations, it still exhibits hysteresis behaviour between the various basins and slow FES convergence, particularly for sub-optimal CVs and poorly chosen algorithmic parameters. To overcome these limitations, more advanced approaches for metadynamics can be employed.

2.3.5 On-the-fly Probability Enhanced Sampling

On-the-fly Probability Enhanced Sampling (OPES) represents a new approach to the metadynamics framework, emphasising the reconstruction of the probability distribution rather than the bias [108]. In OPES, the bias is defined as:

$$V(\mathbf{s}) = -\frac{1}{\beta} \log \frac{p^{tg}(\mathbf{s})}{p(\mathbf{s})}, \quad (2.37)$$

where in the original formulation (OPES-Metad) $p^{tg}(\mathbf{s})$ is the well-tempered distribution of equation 2.33 ($p^{tg}(\mathbf{s}) \propto p(\mathbf{s})^{\frac{1}{\gamma}}$).

Since $p(\mathbf{s})$ is unknown, $V(\mathbf{s})$ is optimised in a self-consistent way during the trajectory (i.e., on-the-fly), following these iterative steps [106]:

- update the estimation of $p_n(\mathbf{s})$ at step n via reweighting, i.e.

$$p_n(\mathbf{s}) = \frac{\sum_k^n w_k G(\mathbf{s} - \mathbf{s}_k)}{\sum_k^n w_k} \quad \text{with} \quad w_k = e^{\beta V_{k-1}(\mathbf{s}_k)}; \quad (2.38)$$

- update the estimation of the bias $V_n(\mathbf{s})$ at step n with the new estimation of $p_n(\mathbf{s})$, combining equation 2.33 and 2.37 to obtain:

$$V_n(\mathbf{s}) = \left(1 - \frac{1}{\gamma}\right) \frac{1}{\beta} \log \left(\frac{p_n(\mathbf{s})}{Z_n} + \varepsilon \right), \quad (2.39)$$

with $\varepsilon \ll 1$ a regularization term and Z_n a normalization factor only associated with the CV space explored up to step n (i.e., not the full CV space);

- sampling in the biased trajectory with the updated bias $V_n(\mathbf{s})$ until a new estimation of p_n .

The regularisation term ε in equation 2.39 has the important role of avoiding the exploration of unphysical regions, setting a soft threshold for the maximum bias. Indeed, ε is linked to the height ΔE of the target free energy barrier one wants to overcome via:

$$\varepsilon = e^{-\beta \Delta E / (1 - 1/\gamma)}. \quad (2.40)$$

Among the several advantages of OPES with respect to WT-MetaD, we stress the iterative scheme of OPES allows for quickly reaching a regime in which the bias is quasi-static, i.e., more easily reaching the convergence of the properties. At the same time, OPES is more efficient in the reweighting step since the Gaussian kernels in equation 2.38 are not stored on a grid, but it uses a compression algorithm where closed kernels are merged together, preventing the number of kernels from growing linearly along the simulation time [106]. For these reasons, the OPES-Metad algorithm will be preferred to the WT-MetaD as enhanced sampling technique in this thesis.

In addition to the discussed OPES-Metad scheme previously discussed, the other variants of OPES will also be employed in specific conditions.

OPES-Explore

If OPES-Metad is meant for precise evaluation of free energy difference, OPES-Explore is designed for faster exploration of new basins, easily escaping from the original bias also in case of sub-optimal CVs [106]. For OPES-Explore, the bias V_n is not iteratively updated using the estimation of the unbiased distribution p_n at step n , but it directly uses the estimation of the actual distribution p_n^{tg} sampled in the

biased trajectory. Hence, V_n is:

$$V_n(\mathbf{s}) = (\gamma - 1) \frac{1}{\beta} \ln \left(\frac{p_n^{tg}(\mathbf{s})}{Z_n} + \varepsilon \right), \quad (2.41)$$

where the target distribution is always the well-tempered one, and its estimation at step n is:

$$p_n^{tg}(\mathbf{s}) = \frac{\sum_k^n G(\mathbf{s} - \mathbf{s}_k)}{n}. \quad (2.42)$$

The possibility of a fast exploration of the FES via OPES-Explore makes this technique suitable for a preliminary exploration of the FES during the rounds of active learning in the construction of ML potentials (cf. Chapter 3 for more details about the active learning procedure).

OPES-Flooding

Different from other OPES schemes, OPES-Flooding is meant to partially fill only one of the metastable basins and allow the system to spontaneously complete the overcoming of the barrier, escaping from the basin without adding bias in the region of the TSs. In this way the TSs are not influenced by the bias. This scheme is effective only if the biased CVs are able to distinguish the starting basin from the TSs. Even if this scheme was originally developed for kinetic calculations, it is also applied to generate reactive trajectories to specifically target the sampling of the region close to the TSs. For example, this scheme will be employed during the active learning step for the generation of ML potentials, stopping the MD simulations as soon as the system overcomes the TS regions to finalise the trajectory to sample only the TS regions, without exploring and sampling other basins.

2.3.6 Considerations on the collective variables

No matter which CV-based advanced sampling algorithm is utilised, selecting appropriate collective variables is fundamental to the success of the method. Exploration efficiency and FES convergence are influenced by the quality of the biased CVs. An optimal CV should distinctly separate metastable basins and include all the slow modes, i.e., the degrees of freedom that evolve on long timescales and are therefore the most difficult to sample [105, 109]. If an important slow mode is omitted or a

poorly discriminating CV is chosen, convergence on the FES becomes inefficient. Being a coarse-grained coordinate, an optimal CV should also perform an effective dimensionality reduction. Indeed, even if biasing multiple CVs can help capture all the relevant modes, most CV-based enhanced sampling methods are computationally effective only with few CVs (usually no more than three or four) [106].

Finally, when a CV is used for biasing, it has to be continuous and differentiable across the configurational space. This requirement is associated with the evaluation of the forces $\mathbf{f}^{\text{bias}}(\mathbf{R})$ during the biased simulation since:

$$\mathbf{f}^{\text{bias}}(\mathbf{R}) = -\nabla_{\mathbf{R}}(U^{\text{unbias}}(\mathbf{R}) + V(\mathbf{s}(\mathbf{R}))) = \mathbf{f}^{\text{unbias}}(\mathbf{R}) - \nabla_{\mathbf{s}}V(\mathbf{s})\nabla_{\mathbf{R}}\mathbf{s}(\mathbf{R}). \quad (2.43)$$

In addition, the CVs should respect the relevant symmetries of the analysed system to correctly represent the physical properties such that the bias potential does not introduce artefacts. Considering the several requirements associated with CVs, choosing optimal CVs is thus a delicate task that requires both physical intuition and system-specific expertise [105]. We will discuss the selection of the CVs in our specific cases in Chapter 6.

In summary, this chapter has introduced the basic principles of DFT, the *ab initio* atomistic approach that will be used to describe the phenomena occurring at electrochemical interfaces. The ability of DFT to reliably capture the electronic effects at play in these systems is crucial, as this level of theory will serve as the reference ground truth for the ML models developed in the following chapters. Particular emphasis was placed on the choice of exchange–correlation functional: rather than adopting the widely used PBE approximation, we employ the more advanced meta-GGA SCAN functional, selected for its superior accuracy in reproducing the properties of aqueous systems without evident over-structuring. Beyond static electronic-structure calculations with DFT, we also introduced the core ideas of molecular dynamics (including, the Verlet integration scheme and NVT thermostats such as the stochastic velocity-rescaling thermostat) and its principal variants, providing the conceptual basis for modern ML potential frameworks that aim to bridge the gap between *ab initio* and classical molecular dynamics. This topic will be covered in detail in Chapter 3. Finally, we outlined how molecular dynamics, combined with enhanced sampling methods such as OPES and its variants, can be applied to investigate reactive processes and to determine the corresponding free-energy activation barriers. These approaches will be utilised in Chapter 6.

Chapter 3

Methods: Machine learning interatomic potentials

The development of machine learning interatomic potentials (MLIPs), also known as machine learning force fields (ML-FFs) for electrochemical interfaces, is the central element of this thesis. This chapter is fully devoted to the methodology associated with MLIPs and the approach to generate them. ML methods learn functional relationships between inputs (chemical descriptors) and outputs (properties) by identifying patterns in *ab initio* data. This makes them particularly effective for constructing versatile and accurate force fields that map atomic configurations to the corresponding region of the PES, enabling predictions of energies and forces.

Modern MLIP formulations trace back to the seminal work of Behler and Parrinello (2007) [110], which introduced the key ingredients of MLIP models. Since then, the field has evolved rapidly, integrating advanced descriptors and more complex ML architectures to enhance accuracy and transferability. In this chapter, Section 3.1 discusses the main components of MLIPs, while Section 3.2 outlines the standard methodology adopted to generate them.

3.1 Characteristics of a ML potential

A common characteristic of the ML potentials, derived from the Behler-Parrinello work [110], is the decomposition of the complex chemical system into local atomic environments, allowing training of the models in systems with a small number of

atoms (i.e., *ab initio*-affordable) and then transferring them to larger structures in the production use. Given an atomic configuration, the ML model is commonly applied to each atom whose properties (e.g., forces and local energy contributions) are computed considering the interaction of the atom with its neighbours within a local environment defined by a finite cut-off radius (typically around 6 Å). Global properties are then computed as the sum of contributions from local environments. For example, the total energy of the system is computed as follows:

$$E = \sum_{i=1}^N E_i, \quad (3.1)$$

where E_i is the local energy contribution associated with atom i .

Another key characteristic of MLIPs derived from the Behler-Parrinello work is the presence of three elements, each of which has important effects on the final performance of the MLIP: the dataset used to train the model, the descriptors used to map the raw data in a more effective representation in view of the ML fitting and the ML architectures for the regression of some physical properties, such as energy and forces. (Figure 3.1).

3.1.1 Central role of the dataset

The dataset, i.e., the data on which the ML model is trained and validated, has a central role, and it is composed of several atomic configurations labelled (i.e., with assigned energies, forces and other relevant properties) at the *ab initio* level (DFT in this thesis). Indeed, in this work we mainly refer to supervised ML algorithms that require the data to be characterised by input features (e.g., a representation of the local atomic environment that will be used as input for the model predictions) and labels, i.e., target properties that the ML model will learn to predict (e.g., energy and forces).

The generation of the dataset is the most computationally expensive step but also the most important one [52]. The accuracy of the ML model is limited by the underlying accuracy of the data. For this reason, extreme care should be devoted to the labelling of the data to grant accuracy and, above all, consistency of the data within the whole dataset. In addition to accuracy and consistency, the dataset has to be complete in the sense that all the relevant regions of the target PES should be adequately sampled and represented in the data since ML models are able to interpolate new points within

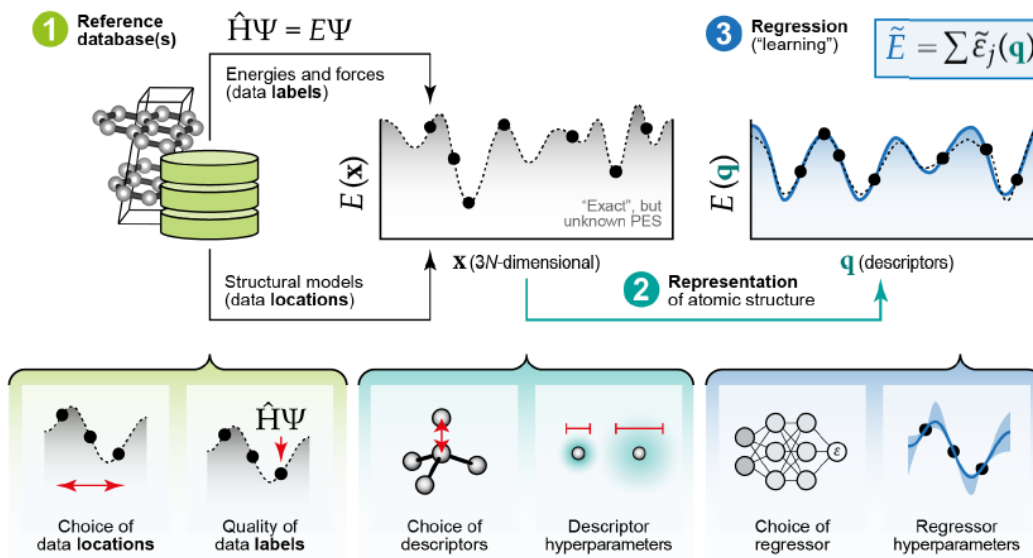


Fig. 3.1 General overview of the MLIP methodology, with the three main "ingredients" highlighted: 1) the dataset containing representative structures labelled with ground truth energies and forces, 2) the representation (descriptors) encoding the atomic environment, and 3) the ML architecture for the regression task. The second row sketches criteria and choices associated with each ingredient that affect the quality of the ML potential. Reproduced from [111].

the dataset domain, but they can present risks in extrapolating outside it. Ensuring a complete sampling of the PES is a challenging task, above all when the extension of the target PES is unknown and/or some regions are accessible only via high-energy configurations (transition states).

For this reason, nowadays, active learning strategies (cf. Section 3.2.2) are commonly employed to assemble an exhaustive dataset. Within these approaches, the dataset creation is performed iteratively alongside the exploration of the PES and the refinement of the ML models. Moreover, these strategies can be coupled with enhanced sampling techniques (cf. Section 2.3) to adequately include also rare events in the dataset.

Practically, the dataset is not only used for the training of the model (training set), but a portion is employed for its validation and test [52, 111]. If the data of the training set are used to optimise the learnable parameters of a model, the validation set is used to compare the accuracy of models with different architectures and/or hyperparameters (i.e., the parameters that define the learning behaviour of the ML models and are not learnable but set before the training) to identify the best one.

Finally, the test set is used to check the accuracy of the best model, evaluating its generalisation capability on unforeseen configurations.

3.1.2 Data representation via descriptors

ML regression models that predict energies and forces cannot be used directly with raw input in the form of atomic cartesian coordinates. Atomic configurations are mapped to the input features of the regression model via descriptors $G(\mathbf{R})$, i.e., many-body functions of the atomic coordinates $\mathbf{R}$ that are used to characterise the local environment around each atom. Effective descriptors enforce physical constraints in the representation, such as spatial translation, rotational and atom permutation invariance [112]. These constraints do not limit the generality of the model, but they are properties that characterise any target PES. These invariances are used to reduce the set of all the possible functions that the ML model could learn, decreasing the degrees of freedom for the learning. Hence, the descriptors allow for presenting the data to the ML model in a better functional form that makes the regression step easier, increasing the accuracy of the predictions. Imposing physical constraints on the descriptors also increases the data efficiency of the ML model since different atomic configurations, but equivalent under the discussed invariances, are automatically presented to the regression model as identical, without the need for the ML model to learn their equivalence from the data.

Due to the important role of the descriptors, several formulations were proposed: the accuracy and the data efficiency of the different MLIP codes are strongly correlated with the implemented descriptors [112]. Here we focus on a specific class of descriptors that will be used in this thesis and are based on the atomic cluster expansion (ACE) formalism [113]. Differently from other descriptor families that rely on a predefined functional form, ACE provides a general approach to represent the atomic environments via many-body interactions through an expansion in terms of local basis functions. In this way the PES is expressed as a sum of contributions from atomic clusters of increasing order (e.g., single atoms, pairs, triplets, etc.).

Given a system of N particles with positions $\mathbf{R} = [r_{j1}, r_{j2}, \dots, r_{jN}]$, the standard approach to compute the energy of a system via cluster expansion consists of sums of

the contributions of single atoms $f^{(1)}$, pairs $f^{(2)}$, triplets $f^{(3)}$ and high-order clusters:

$$E(\mathbf{R}) = \sum_{j_1} f^{(1)}(r_{j_1}) + \sum_{j_1 < j_2} f^{(2)}(r_{j_1}, r_{j_2}) + \cdots + \sum_{j_1 < \cdots < j_{n+1}} f^{(n+1)}(r_{j_1}, \dots, r_{j_{n+1}}). \quad (3.2)$$

However, the cost of this summation quickly increases, adding high-order cluster contributions with correlation order n ($\propto N_c^n$ for the leading term $n+1$, with N_c the number of neighbour atoms), and it is a formulation that cannot be easily adapted to build descriptors that satisfy the required symmetry operations.

For this reason, the problem can be formulated following the so-called "density trick" [114]. It is possible to introduce a set of *one-particle basis functions* in $\mathbb{R}^3$, where the k -element $\phi_k(r_{ij})$ of the basis set is used to describe the interaction of atom i with its neighbour atom j . For example, $\phi_k(r_{ij})$ can be defined as a product of a radial basis R_{nl} and a real spherical harmonic Y_l^m [115]:

$$\phi_k(\mathbf{r}) = R_{nl}(r) Y_l^m(\hat{\mathbf{r}}), \quad (3.3)$$

where n, l, m are the indexes commonly employed in the formalism of radial and spherical harmonics, while the k index is used to identify the elements of the *one-particle basis function*-set and depends on the triplet (n, l, m) . From this, it is possible to introduce an *atomic basis* A_k as a sum of the contributions of *one-particle basis functions* of all the neighbour atoms j around i [113]:

$$A_k = \sum_j^{N_c} \phi_k(r_{ij}), \quad (3.4)$$

where the summation over j runs only on the N_c neighbour atoms around atom i . This basis set is invariant to permutation of atoms ($i \leftrightarrow j$), but it represents only a 2-body contribution (pairs of atoms $i-j$). For this reason, it can be generalised to a many-body permutation-invariant basis $\mathbf{A}_{k_1 \dots k_n}$ (the $\mathbf{A}$ – *basis*):

$$\mathbf{A}_{k_1 \dots k_n} = \prod_{t=1}^n A_{k_t} \quad \text{where } n \text{ denotes the correlation order.} \quad (3.5)$$

For example, it is possible to obtain a 3-body basis function for $n = 2$ (Figure 3.2):

$$\mathbf{A}_{k_1 k_2} = A_{k_1} A_{k_2}. \quad (3.6)$$

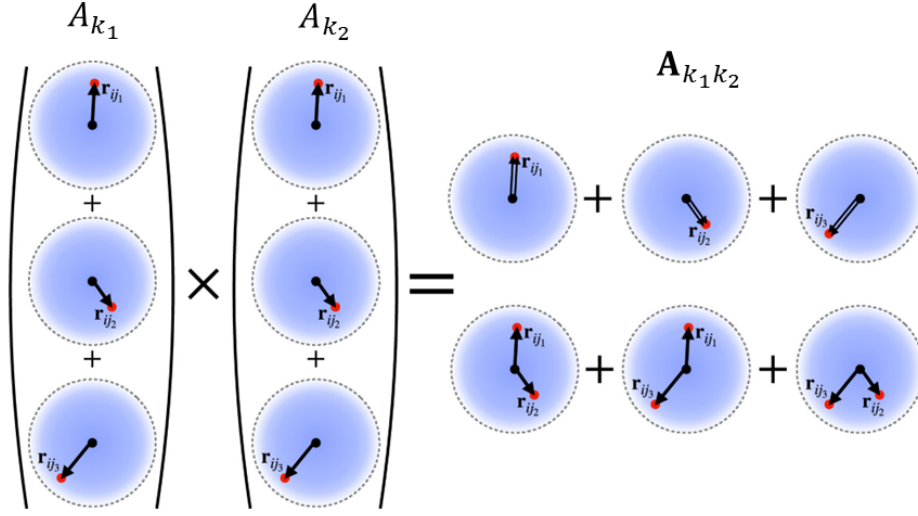


Fig. 3.2 Scheme for the construction of the $\mathbf{A}$ -basis function for the case of the 3-body term (right side) from the product of $n = 2$ atomic basis A_k (left side), considering 3 neighbouring atoms. Adapted from [114].

Constructing a many-body term in this way (firstly, the sum of 2-body terms ϕ_k to obtain A_k and then the product to obtain a many-body term $\mathbf{A}_{k_1 \dots k_n}$) is more convenient than summing over all possible many-body terms of the same order, as performed in equation 3.2 [114]. Indeed, the evaluation cost of the basis functions scales linearly with the number of neighbours ($\propto N_c$) rather than exponentially ($\propto N_c^n$ with $n + 1$ the maximum body term) by doing the neighbour summation first (eq. 3.4) and then taking the products to obtain higher-order basis functions (eq. 3.5).

Following this strategy, the energy associated with the local environment of atom i can be expressed as:

$$E_i = \sum_k c_k \prod_{t=1}^n \sum_j^{N_c} \phi_{k_t}(r_{ij}) = \sum_k c_k \prod_{t=1}^n A_{k_t} = \sum_k c_k \mathbf{A}_{k_1 \dots k_n}, \quad (3.7)$$

where k is the same index defined in eq. 3.3 and it also identifies the elements of the $\mathbf{A}$ – basis set. Explicitly expanding equation 3.7, it is possible to prove that the terms correspond to a sum of $(n + 1)$ -body terms as in the traditional body order expansion of equation 3.2 [114]. The analogy between the two formulations for the atomic cluster expansion is reported in Figure 3.3. This approach to building many-body term descriptors can be further generalised, averaging the $\mathbf{A}$ -basis over the three-dimensional rotation group $O(3)$, to obtain a $\mathbf{B}$ -basis that includes the

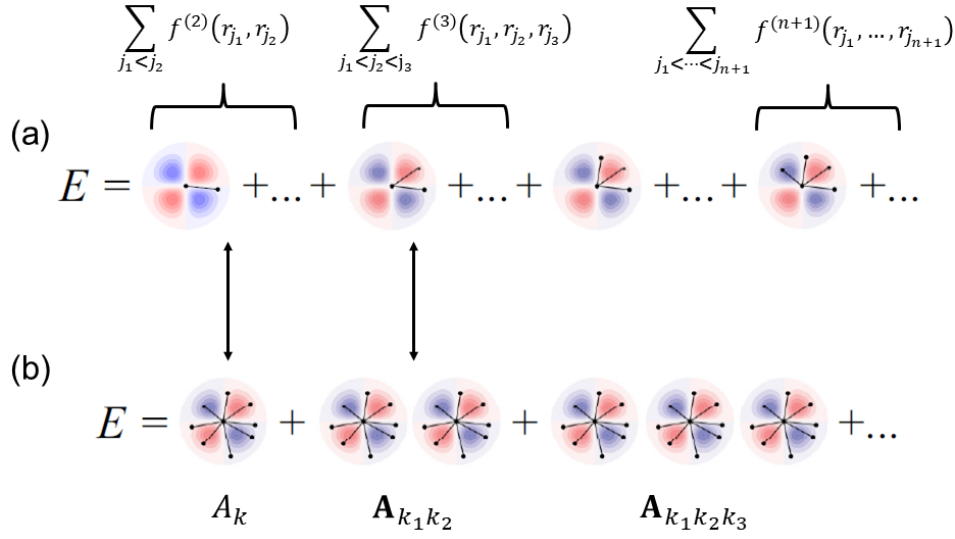


Fig. 3.3 Analogy between two formulations of the atomic cluster expansion: (a) ACE as a sum of pair, three-body, four-body, etc. (Equation 3.2), (b) ACE as a polynomial of the *atomic basis functions* (Equation 3.7). Adapted from [113].

rotational symmetry invariance [115]. At the end, similarly to Equation 3.7, the energy can be expressed as a combination of the contributions of the **B**-basis.

The strength of this approach is the possibility of building descriptors able to account for complex many-body interactions while being computationally efficient. Moreover, it represents a systematic scheme that allows for increasing the accuracy, including high-order terms in the expansion. Consequently, a large number of MLIPs employ the ACE basis descriptor [113, 116]. Building on this success, the ACE framework has been incorporated into more sophisticated neural network architectures to construct learnable descriptor representations, thereby yielding more expressive and flexible feature mappings [115].

3.1.3 Architectures of the ML models

The architecture of an ML model, i.e., its structural design, is a key factor that differentiates various ML potentials. In most cases, the architecture defines the regression component of the MLIP, which learns to approximate the PES by predicting energies and forces from input features, namely descriptors extracted from atomic configurations. While traditional models rely on predefined descriptors, deep learning approaches go further by learning both the regression function and the

optimal representations of the descriptors directly from data. The complexity of the architecture and the number of trainable parameters influence the accuracy, training speed, and inference performance of different MLIP models. In this work, we do not aim to provide an exhaustive review of all ML architectures used in MLIPs; instead, we focus on outlining the main characteristics of the specific models employed in this thesis.

Neural networks

Common architectures employed to develop MLIPs are based on artificial neural networks (ANNs or, more simply, NNs) [52]. The basic building block of an NN is a dense (or “fully-connected”) layer of neurons implementing a linear transformation of an input vector $\mathbf{x} \in \mathbb{R}^{n_{in}}$ to an output vector $\mathbf{y} \in \mathbb{R}^{n_{out}}$ via:

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}, \quad (3.8)$$

where the weight matrix $\mathbf{W} \in \mathbb{R}^{n_{out} \times n_{in}}$ and the bias vector $\mathbf{b} \in \mathbb{R}^{n_{out}}$ are composed of learnable parameters (see Figure 3.4a). The output of each neuron is then processed with a non-linear activation function σ (e.g., the hyperbolic tangent or rectified linear units - RELU):

$$\mathbf{h} = \sigma(\mathbf{y}) = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}). \quad (3.9)$$

These “fully-connected” layers can be stacked in series such that the output vector $\mathbf{h}$ of a layer becomes the input vector of the subsequent neuron layer (see Figure 3.4b). By stacking more layers and using nonlinear activation functions, it is possible to approximate any mapping between the input and the output of the NN model. In practice, it is common to stack multiple layers rather than a few layers with lots of neurons; these “deep” NN architectures are characterised by a superior accuracy in approximating any arbitrary relationship in the data, and they are more parameter-efficient.

In the case of ML models employed to fit the PES, the input of the first layer of the NN is a representation of the chemical structures (contained in the dataset or a new configuration to infer) cast through descriptors (input features), while the output of the last layer is the prediction of some target properties (e.g., energy and/or forces). The learning process of the NN (i.e., finding the optimal values of the model parameters of $\mathbf{W}$ and $\mathbf{b}$ for each neurons) is based on the minimisation of the error

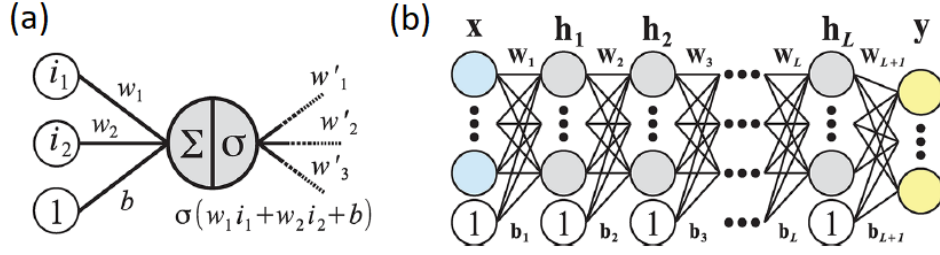


Fig. 3.4 Main characteristics of a "fully-connected" NN: (a) scheme of a single neuron of an NN processing the weighted summation (Σ) of the input via an activation function σ , (b) scheme of a deep NN with L hidden layers (in grey). Reproduced from [52].

between the prediction outputs and the true values of the properties, as computed at the *ab initio* level. For this scope, a loss function $\mathcal{L}$ is defined to measure the difference between the predicted and the ground-truth values. A typical multitask loss function for MLIPs to learn both energies and forces is the Mean Squared Error (MSE):

$$\mathcal{L} = \frac{w_E}{M} \sum_{i=1}^M \left(E_i^{\text{pred}} - E_i^{\text{true}} \right)^2 + \frac{w_F}{3MN} \sum_{i=1}^M \sum_{j=1}^N \left\| \mathbf{F}_{ij}^{\text{pred}} - \mathbf{F}_{ij}^{\text{true}} \right\|^2, \quad (3.10)$$

where:

- E_i^{pred} and $\mathbf{F}_{ij}^{\text{pred}}$ are the predicted total energy and force on atom j in sample i ,
- E_i^{true} and $\mathbf{F}_{ij}^{\text{true}}$ are the corresponding total energy and forces computed at the *ab initio* level (the labels of the data),
- M and N are the number of training samples and the number of atoms in each configuration, respectively,
- w_E and w_F are non-learnable parameters and are the weights of the energy and force terms. They are used to drive the learning process to prioritise the match of energy or forces. During the training, a commonly employed strategy tends to optimise first the prediction of the forces while improving the learning of the energy in a later stage of the training. For example, at the beginning of the training, one can set $w_F > w_E$ (e.g., $w_F = 100$ and $w_E = 1$, prioritising the forces) to switch in a second stage with $w_F < w_E$ (e.g., $w_F = 100$ and $w_E = 1000$) to improve the learning of the energy.

Given the complexity of the NN, it is impossible to minimise the loss function in a closed form, but the weights and the biases of the model are progressively updated to minimise $\mathcal{L}$ through optimisation algorithms based on gradient descent. The dependence of $\mathcal{L}$ with respect to the different learning parameters is calculated via derivative, and the trainable parameters are updated via a back-propagation algorithm. For example, the following update rule is applied for the weights of the neurons:

$$w_{j,k}^l \leftarrow w_{j,k}^l - \eta \frac{\partial \mathcal{L}}{\partial w_{j,k}^l} \quad (3.11)$$

where $w_{j,k}^l$ is a trainable weight from the k -th neuron in the $l - 1$ -th layer to the j -th neuron in the l -th layer, and η is the learning rate used to control the step size during the optimisation process. A similar relation is also applied for the elements of the bias vector $\mathbf{b}$ of each layer.

Message-passing neural networks

Message-passing neural networks (MPNNs) are a typical example of deep learning in which both the descriptor and the regression part of the MLIP are implemented with a ML model. MPNNs expand the expressiveness of traditional neural networks. Differently from the latters, whose inputs are fixed-size vectors, MPNNs are a class of graph neural network (GNN). Here the atomic configurations are mapped in a graph where atoms and their reciprocal interactions (within a cut-off radius) are represented as nodes and connections (edges), respectively [117]. The representation of the atoms (i.e., the descriptors of the local environments associated with each atom) evolves within the MPNN architecture by being processed by different interaction layers in sequence (see Figure 3.5a) [118]. The state of the node i at the interaction layer t is represented by:

$$\sigma_i^{(t)} = (\mathbf{r}_i, z_i, \mathbf{h}_i^{(t)}), \quad (3.12)$$

where $\mathbf{r}_i$ and z_i are the position and the chemical species of atom i , while $\mathbf{h}_i^{(t)}$ are its learnable features. This state is updated in the subsequent layer collecting information (i.e., the message) from its neighbour nodes. Specifically, a message $m_i^{(t)}$ is constructed summing contributions from the neighbour atoms j :

$$m_i^{(t)} = \sum_{j \in \text{Neighbour}(i)} M_t(\sigma_i^{(t)}, \sigma_j^{(t)}), \quad (3.13)$$

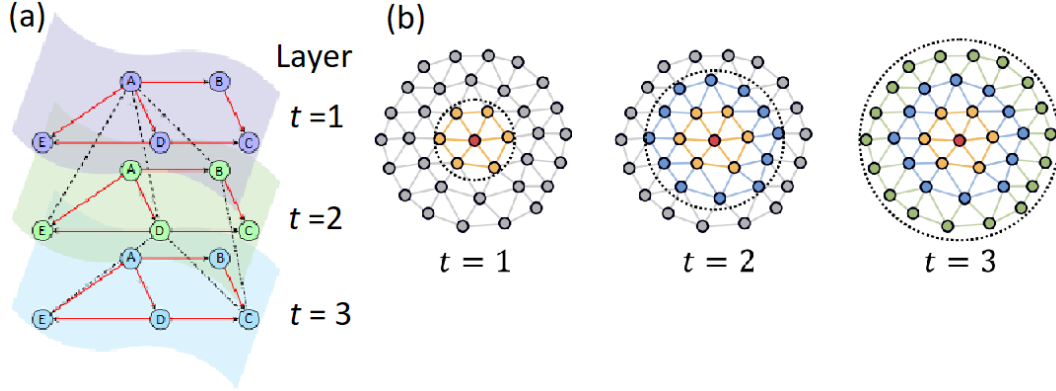


Fig. 3.5 Characteristics of an MPNN: (a) evolution of the graph network along three interaction layers. The edges of the graph are depicted in red, while the message propagation to the neighbour nodes is depicted with black dashed arrows. Thanks to the message propagation, the information of node A can reach node C in the interaction layer $t = 3$, even if A and C are not neighbours. Adapted from [120]. (b) Expansion of the receptive field of the central atom in red after three successive interaction layers. Reproduced from [121].

where M_t is a learnable message function. Then, in the next layer $t + 1$, the learnable features of the node i become:

$$h_i^{(t+1)} = U_t(\sigma_i^{(t)}, m_i^{(t)}), \quad (3.14)$$

where U_t is a learnable update function. Since the node representation $h_i^{(t+1)}$ depends on both its previous state $h_i^{(t)}$ and the states of its neighbours $h_j^{(t)}$, each layer enables an atom to exchange information with its immediate neighbours. By stacking multiple layers, this local exchange accumulates, allowing information to reach atoms that are farther away. Consequently, non-local information can be captured through a sequence of layers, increasing the receptive field of the model (see Figure 3.5b). Given T interaction layers, in which the neighbour atoms are defined within a region with a cut-off radius r_c , the receptive field of the model is expanded up to $T \times r_c$ [119].

After T layers, a target property (e.g., the local energy of atom i) is predicted, using a learnable readout function R_t and summing the contributions of each layer t :

$$E_i = \sum_{t=1}^T R_t(\sigma_i^{(t)}). \quad (3.15)$$

While the interaction layers composed of message propagation and update steps

are meant to implement the descriptors of the model, the last block associated with the readout function is the regression part of the MPNN. The message function M , the update function U and the readout function R can be implemented through any ML model, including "fully connected" NNs. In the following we present MACE [115, 118], a specific implementation of an MPNN employed in this thesis.

The MACE descriptors are based on the ACE formalism (cf. Section 3.1.2), but here ACE is integrated into a message-passing scheme, expanding the expressivity of its original formulation. At each interaction layer t the node features are described with respect to the basis functions of the ACE formalism: the *one-particle basis functions* $\phi_k^{(t)}(r_{ij})$ at the interaction layer t , not only include radial and angular contributions (eq. 3.3), but also the learnable features $h_i^{(t)}$. In this way, the final ACE many-body permutationally and rotationally invariant **B**-basis is defined for each interaction layer, and its representation is improved at each interaction passage through the update of $\phi_k^{(t)}$ via the learnable feature $h_i^{(t)}$. Thanks to the efficiency of ACE in expressing complex many-body terms, MACE constructs computationally effective high-order messages. Moreover, MACE parametrises the message in terms of invariant or equivariant ACE models. Finally, the update, message, and readout functions are implemented not with complex NNs but as linear combinations, except for the readout in the final interaction layer, which uses a "fully connected" NN architecture. All these characteristics make MACE more computationally and data efficient than other MPNN models while being accurate and having appreciable in-domain but also out-of-domain predictions. More details on the MACE implementation and performance can be found in the original article [118].

Kernel-based methods

Kernel-based regression methods represent another popular category used to develop ML potentials. The basic idea behind kernel methods is approximating the relation between the input data of the model $\mathbf{x} \in \mathbb{R}^{in}$ and the output $y \in \mathbb{R}$ in the form [52]:

$$y = f(\mathbf{x}) \approx \sum_{n=1}^M \alpha_n K(\mathbf{x}_n, \mathbf{x}), \quad (3.16)$$

where $\{\mathbf{x}_n\}$ is a set of M training data, $\mathbf{x}$ is new input data for which we want to predict the associated property y , $\{\alpha_n\}$ is a set of trainable parameters, and K is a symmetric and semi-positive defined kernel function. The kernel function performs

a similarity measurement between training data $\mathbf{x}_n$ and a new point $\mathbf{x}$. However, this measurement of similarity is not performed in the $\mathbb{R}^{in}$ -space of the data, but they are mapped in a high-dimensional Hilbert space $\mathcal{H}$ via the following transformation:

$$\phi : \mathbb{R}^{in} \rightarrow \mathcal{H}. \quad (3.17)$$

The kernel function measures the similarity of the data in this latent space via an inner product:

$$K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}'). \quad (3.18)$$

Following the so-called kernel trick strategy, it is possible to define some kernel functions with known properties without explicitly knowing the mapping function ϕ to the high-dimensional space. A common choice for the kernel function is the Gaussian kernel:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right). \quad (3.19)$$

Once the kernel function is selected, several methods (such as Gaussian processes, kernel ridge regression, etc.) can be employed to determine the optimal trainable parameters α_n in equation 3.16. These parameters are chosen to minimise a loss function, which is typically defined in terms of the discrepancy between the predicted and the true outputs [52].

An example of a kernel-based approach is the Gaussian process (GP) regression method. This strategy determines the optimal α_n parameters, supposing that the reference data $\{(y_n, \mathbf{x}_n)\}_{n=1}^M$ are drawn from a multivariate Gaussian distribution. For simplicity, the mathematical analysis below assumes that this distribution has a mean of zero, since any other mean can be obtained by adding an appropriate constant term. Under this assumption it is possible to rigorously prove (cf. Ref [122]) that the posterior probability of observing the value y_* for the new input data $\mathbf{x}_*$, given the data $\{(y_n, \mathbf{x}_n)\}_{n=1}^M$ in the training set (i.e., $p(y_* | \mathbf{x}_*, \mathbf{y}, \mathbf{x})$) follows a Gaussian distribution with the following mean value and variance:

$$\bar{y}_* = \mathbf{K}_* \mathbf{K}^{-1} \mathbf{y}, \quad (3.20)$$

$$\text{var}(y_*) = K(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^T, \quad (3.21)$$

where $\mathbf{K}_* = [K(\mathbf{x}_*, \mathbf{x}_1) \dots (\mathbf{x}_*, \mathbf{x}_M)]$ and $\mathbf{K}$ is the $M \times M$ kernel matrix with entries $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$. At this point, the mean value $\bar{y}_*$ is supposed to be the most likely approximation for y_* . Equating equations 3.16 and 3.20, it is possible to find a closed-form solution to determine the learnable parameter α_i for the linear combination of the kernels [122]:

$$\alpha = \mathbf{K}^{-1} \mathbf{y}. \quad (3.22)$$

The peculiarity of GPs is the possibility to predict not only the target output properties of new data (equation 3.20), but also the variance associated with this prediction (equation 3.21). The variance can be considered as a measure of the uncertainty of the new data, i.e., how different they are with respect to the training ones. The importance of this characteristic of the GPs will be discussed in Section 3.2.

An example of ML potential code that employs a GP is FLARE (Fast Learning of Atomistic Rare Events) [116, 123]. The local environments in FLARE are characterised by a set of invariant descriptors obtained as the product of equivariant ones based on the ACE formalism. These descriptors are then the input features $\mathbf{x}$ of the GP model. More details on the FLARE implementation are available in the original articles [116, 123].

The main drawback of the kernel-based approaches is associated with the computational cost of training and new data inference when the dataset is large [124]. The problem is intrinsic in equation 3.16, since the prediction of a label y requires the coupling inside the kernel expression of the new data $\mathbf{x}$ with the data $\{\mathbf{x}_n\}$ in the training dataset. Formally, given a dataset of M training points, the training cost usually scales as $O(M^3)$, while predicting new data costs $O(Md)$, with d the number of input features. For these reasons, kernel-based methods are not effective in the presence of large datasets.

A large-scale alternative to exact kernel methods is represented by the random feature (RF) strategy [124]. The rationale of the RF method is to perform a Monte Carlo approximation of the kernel function by randomly sampled feature maps. Given a vector $\mathbf{h}(\mathbf{R}_i) \in \mathbb{R}^d$ of features for the i -th atom, rather than directly defining a functional form for the kernel function (i.e., the conventional approach of kernel methods), an "inverse" strategy of the kernel trick is followed. The kernel is expressed as a scalar product in a finite-dimensional feature space as expressed in equation 3.18. Since ϕ is not known in most cases, it is possible to prove that the

kernel can be approximated as:

$$K(\mathbf{h}, \mathbf{h}') \approx \tilde{\phi}(\mathbf{h})^T \tilde{\phi}(\mathbf{h}'), \quad (3.23)$$

where $\tilde{\phi}(\mathbf{h}(\mathbf{R})) : \mathbb{R}^d \rightarrow \mathbb{R}^D$ is a random Fourier features map and D is the number of RFs controlling the quality of the kernel approximation. The use of a random Fourier features map in Equation 3.23 holds since any positively defined and shift-invariant kernel can be computed as a Fourier transform of a probability distribution [125]. Hence, the idea of the RF approximation is the construction of the kernel starting from its Fourier transform. For example, since the Fourier transform of a Gaussian is another Gaussian, the Gaussian kernel (eq. 3.19) can be approximated by sampling from a normal distribution and using the following randomised feature maps:

$$\tilde{\phi}(\mathbf{h}) = \sqrt{2} \cos \left(\frac{\mathbf{W} \cdot \mathbf{h}}{\sigma} + \mathbf{b} \right), \quad (3.24)$$

where each row of $\mathbf{W} \in \mathbb{R}^{D \times d}$ is extracted from a standard multivariate normal distribution, and the offset vector $\mathbf{b} \in \mathbb{R}^D$ has components sampled uniformly from $[0, 2\pi)$ [126]. It is important to note that the entries of $\mathbf{W}$ and $\mathbf{b}$ are not trainable parameters; instead, they are drawn once from a standard multivariate normal distribution and a uniform distribution, respectively, and subsequently kept fixed during learning. In this sense, the number of RFs D represents the number of samples extracted from the probability distribution and defines the model complexity.

An example of ML potential code that employs RFs is Franken [126]. In this scheme, the descriptors $\mathbf{h}(\mathbf{R})$ are extracted from the node features of a pre-trained GNN (see section 3.2.3), while the regression is implemented via a pooled scalar product. For example, the energy associated with the local environment of an atom is computed as:

$$E_{loc} = \sum_n \alpha_n K(\mathbf{h}_n(\mathbf{R}), \mathbf{h}(\mathbf{R})) \approx \sum_n \alpha_n \tilde{\phi}(\mathbf{h}_n)^T \tilde{\phi}(\mathbf{h}) = \mathbf{w}^T \tilde{\phi}(\mathbf{h}), \quad (3.25)$$

where $\mathbf{w}$ is a learnable vector of coefficients. The total energy is then obtained by summing together the atomic contributions, while the forces are calculated as the gradient of the total energy. From Equation 3.25 the advantages of the RF formulation are evident: rather than pass through the direct kernel evaluation coupling the configurations of the training set (red part) with a new configuration (blue term), the information of the data in the training set is lumped together with the fitting parameters α_n in a new trainable vector of parameters $\mathbf{w}$. Once $\mathbf{w}$ is fitted, it is

possible to evaluate new data without comparing it each time with all the data in the training set, resulting in an inference cost independent from the size of the dataset but only dependent on the number of RFs. This allows us to overcome the limits of standard kernel approaches in the case of large datasets. Moreover, it is possible to prove that the minimisation of a loss function to determine the trainable parameters is a convex problem with such an expression for energy and forces [126]. In this way it is possible to have a closed-form expression that globally minimises the loss function without performing the gradient descent algorithm, as performed for NNs, speeding up the training of the ML potential. Hence, the Franken architecture allows for fast training and inference typical of kernel methods but, notably, irrespective of the dataset size.

3.2 Approaches for the development of ML potentials

3.2.1 Quantification of the uncertainty

An important characteristic of an efficient scheme to develop MLIPs is the quantification of the uncertainty in the predictions on new data. A low error on the ML predictions with respect to the DFT labels for the configurations in the dataset is indeed necessary but not sufficient to grant a comparable accuracy on new configurations encountered along an MD trajectory. Consequently, having an uncertainty proxy to estimate the error with respect to the *ab initio* ground truth, without the necessity to compute the DFT labels, is crucial. In this way it is possible to identify new configurations that are not well described by the ML model and eventually improve the MLIP by adding these new data to the dataset after labelling them at the DFT level (see section 3.2.2).

Depending on the selected ML architecture, there are different strategies to quantify the uncertainty. NN and GNN models natively provide only a prediction of the target properties, without providing any additional information on the confidence level in the prediction. For this reason, a common approach involves the development of an ensemble of NN models (a committee) trained on different partitions or permutations of the dataset such that the predictions of the models are independent [127]. The variability in the prediction of the models on the same configuration is considered as an uncertainty proxy that highlights that this geometry is not well represented in the

dataset, and each model extrapolates in a different way. Commonly the variability in the predictions is quantified by computing the standard deviation on the predicted forces among a committee composed of at least four NN models. For each force component α , the uncertainty $\sigma_{i,\alpha}$ of an atom i is computed as follows:

$$\sigma_{i,\alpha} = \sqrt{\frac{1}{N_{\text{NN}}} \sum_{k=1}^{N_{\text{NN}}} \left(f_{i,\alpha}^{(k)} - \bar{f}_{i,\alpha} \right)^2}, \quad (3.26)$$

where $f_{i,\alpha}^{(k)}$ is the prediction of the force component α of the k -th model in the committee, $\bar{f}_{i,\alpha}$ is the corresponding average value among the predictions of the ensemble, and N_{NN} is the number of models in the committee. The uncertainty proxy used to identify the whole configuration is then the maximum of $\sigma_{i,\alpha}$ among the force components and the atoms present in the configuration.

Differently from NNs, statistical methods such as Gaussian process regression approaches do not only provide a prediction (the mean value, eq. 3.20), but also the associated variance (eq. 3.21): the latter is considered as a measure of the confidence level of the prediction. If uncertainty estimates are available for each individual atomic environment, they can provide insight into the most critical interaction types to be learned by the MLIP. The whole configuration is then eventually selected for labelling, relying on the maximum uncertainty among its different local environments. It is important to notice that, differently from the committee approach where the uncertainty is based on the predictions of the models (see eq. 3.26), the uncertainty for the GPs relies on a measure of similarity between a new point and the geometries in the dataset (see eq. 3.21) and it does not directly involve the predictions. This is an important consideration in view of the following discussion on the data-efficient active learning strategy.

3.2.2 Active learning

As anticipated in section 3.1.1, a reliable ML model should be trained on a dataset that exhaustively covers all the relevant regions of the PES to avoid risky out-of-domain extrapolations. However, sampling all the relevant configurations in a "one-shot" approach is challenging because it is not possible to determine which configurations are pertinent beforehand. For this reason, modern strategies are based on active learning paradigms based on iterative refinements of the ML model

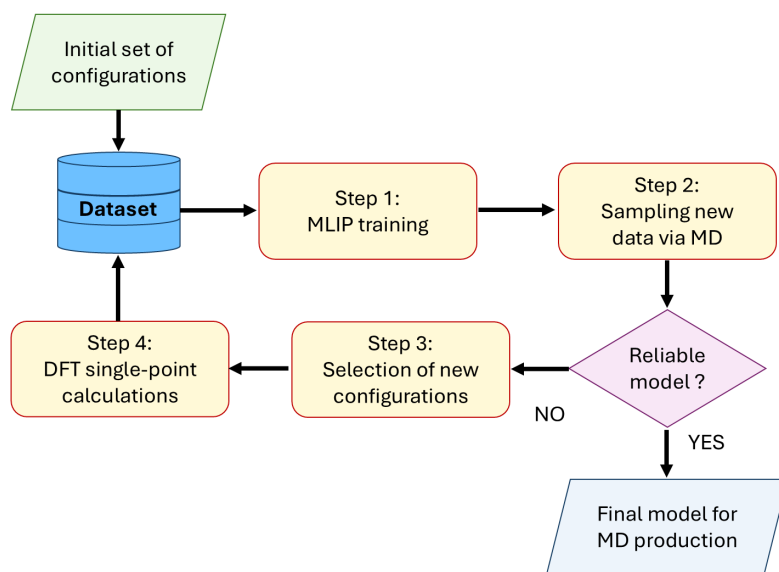


Fig. 3.6 Scheme for the state-of-the-art active learning paradigm for the development of reliable MLIP models.

through subsequent expansions of the dataset once new relevant configurations are encountered [52]. Starting from an initial dataset, a preliminary ML model is trained and is used to generate new structures via MD. A subset of the MD configurations is selected following different criteria (see following sections) and labelled with single-point DFT calculations to expand the dataset. As reproduced in Figure 3.6, this scheme (training, sampling new configurations, selection of the geometries to label, and single-point calculations) is repeated until the convergence of the ML model performance is reached. The advantage of the active learning scheme is the capability of expanding the dataset following the evolution of the MD trajectories towards pertinent and thermodynamically accessible regions of the PES. This minimises the risk of including high-energy geometries or those that are not relevant. In the case of MLIPs for reactive events, the active learning schemes can employ enhanced sampling techniques (Section 2.3) to drive the refinement of the ML model towards other basins of the PESs, including configurations for the transition states.

In the following we discuss in more detail each step of the active learning to present a commonly employed scheme to generate MLIPs, focusing on the case of NN (and GNN) architectures. Later in Chapter 4 we present how we have improved this scheme for the study of electrochemical interfaces at applied external potentials.

Preliminary dataset

The active learning cycle starts from a preliminary dataset that is used to train the first model. It is not necessary for this dataset to be fully complete and exhaustive; however, it is useful to roughly map out the main regions of the PES so that the first ML model can already provide reasonable predictions. Employing *ab initio* MD is the most intuitive approach to collect these preliminary configurations. Nevertheless, solely relying on AIMD is computationally expensive and inefficient. Geometries along the MD trajectory are highly correlated, requiring the discarding of many intermediate snapshots to obtain uncorrelated configurations. Selecting configurations along the MD trajectory every few picoseconds implies that at least $60 \div 70$ % of generated configurations via AIMD are not included in the dataset. Hence, different approaches can be employed to speed up the creation of the first dataset. For example, configurations can still be sparsely sampled along an MD trajectory but using a cheaper solution for propagating the dynamics with respect to DFT. These solutions include AIMD, but with softer convergence criteria, cheaper yet accurate Hamiltonian (e.g., semi-empirical methods) and, more recently, ML potentials trained on large material databases able to reasonably reproduce the properties of the different chemical species among variegated systems (i.e., universal models [128]). This allows for the collection of a large set of uncorrelated configurations at a much lower computational cost with respect to AIMD. When configurations are extracted from MD trajectories obtained through semi-empirical methods or universal MLIPs (i.e., approaches not as accurate as AIMD), and the sampled PES cannot be easily and directly mapped to the corresponding *ab initio* PES, it is advantageous to sample a broader region of the PES, e.g., performing MD at higher temperatures or using enhanced sampling techniques. This enhances the robustness of the model and increases the likelihood that the relevant parts of the PES are covered. A fraction of these configurations (e.g., evenly spaced along the trajectory) not generated via accurate AIMD then have to be labelled at the target *ab initio* level to ensure the required accuracy in the dataset.

Active learning cycle

After collecting this first dataset, the active learning cycle starts. In each iteration, an ensemble of models (usually four) is trained on different partitions of the dataset

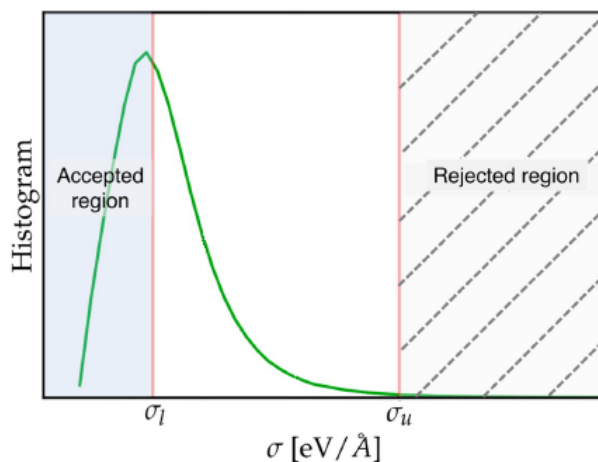


Fig. 3.7 Typical distribution of the uncertainty of NN force predictions using a committee of models. A common choice of the different regions used to select the configuration is reported.

to generate a committee of NNs, as discussed in Section 3.2.1. The trained models are then used to sample new configurations along MD trajectories. At this point a crucial aspect is the selection of a few but relevant geometries among the large amount of data generated by MD. The idea is to select only the configurations that are not well described by the ML model yet, i.e., for the regions of the PES where there is room for prediction improvement. A common approach for selecting these configurations is based on *query-by-committee*. Each configuration is characterised by an estimation of the uncertainty σ via the maximum standard deviation of force predictions (see Section 3.2.1). Analysing the distribution of σ (see an example in Figure 3.7), it is possible to identify an optimal range for σ in which the configurations will be selected. Typically a lower bound (σ_l) and an upper bound (σ_u) for the uncertainty are defined. The first is usually set slightly above the maximum of the distribution, and all the configurations with a lower uncertainty are considered sufficiently well described by the model, and there is no need to include them in the dataset. Whereas the upper bound is defined to reject the configurations in the tail of the distribution for which the uncertainty is large and they are probably not physical. All the configurations with σ between the two bounds are considered not sufficiently well described but still physically plausible, so they are candidates to be selected and labelled at the DFT level.

Nonetheless, the number of configurations in this selection region can still be large, and it may not be feasible to compute all these geometries at the DFT level. For

this reason, further selection criteria are required, above all when modelling reaction paths. In the latter case, there is an imbalance among the number of configurations belonging to the different basins, and usually only a few configurations are associated with transition states. Hence, selection criteria able to discriminate the configurations along the reaction path should be employed. In this framework, we will follow the DEAL protocol to maximise the effectiveness of the selection.

DEAL protocol: Recently, the “Data Efficient Active Learning” (DEAL) scheme was proposed to provide a systematic protocol to ensure a uniform coverage of the relevant configurational space while dramatically reducing the number of selected configurations to be labelled at the DFT level [129]. After the *query-by-committee* selection, DEAL evaluates the similarity among the local environments of the sampled configurations to identify a reduced subset of geometries, minimising the redundancy of the data. This filtering step is based on the Bayesian predictive variance of a sparse Gaussian process model (see eq. 3.21), using the FLARE code [116, 123]. The GP model is trained *on-the-fly* on the sequence of pre-filtered configurations and uses the analytical variance of the predictions to quantify the diversity of the local atomic environments, selecting only those configurations that are different from those already selected. As observed in section 3.2.1, the estimation of uncertainty of a GP is based on a similarity between local environments, not on the predictions of the model. Hence, in DEAL the GP is trained on the energy and forces, not computed at the DFT level but with the ML model used to generate the configurations (e.g., an NN model). These energies and forces used to train the GP might not be accurate with respect to the DFT ground truth, but the GP model is trained not to perform MD runs but only as a tool to measure the similarity of the configurations and reduce redundancy. In these terms, the uncertainty of the GP based on environment similarity and not on energy/force prediction is an essential element.

The effectiveness of the DEAL selection scheme has been extensively proved [129], showing a reduction of the number of required DFT evaluations by more than a factor of 20, without compromising the accuracy of the resulting MLIP. A specific example of the effectiveness of DEAL in structure selection in a metal/water interface, a system that will be central to the discussion in the following chapters, is reported in Figure 3.8. The peculiarity of recognising and distinguishing the various local environments in the innermost part of interfacial water will be a key point in the strategy proposed in Chapter 4. Indeed, relying solely on the *query-by-committee*

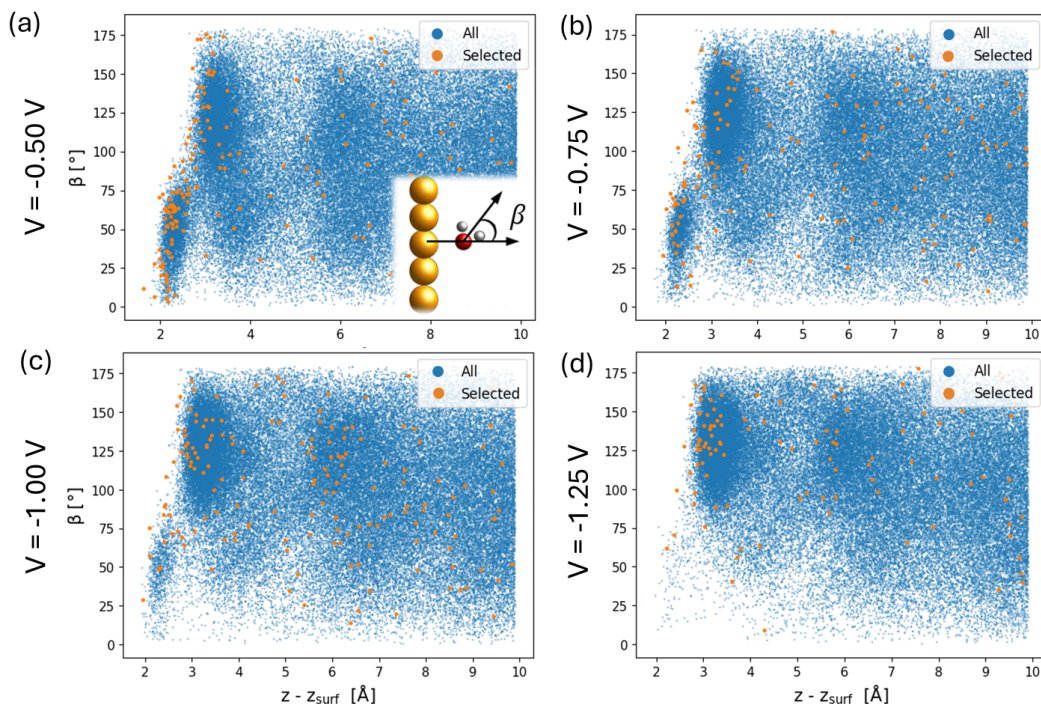


Fig. 3.8 Effectiveness of DEAL in identifying the most relevant and non-redundant local environments for a representative Cu (111)/water interface under different applied potentials (V vs SHE). Scatter plots show the variation of water dipole orientation (see the inset for the geometrical definition of the β angle) as a function of the distance from the metal surface (located at z_{surf}). Blue points correspond to orientations from an unfiltered series of MD snapshots, while orange points represent molecules with local environments selected by DEAL. Although the bulk water region is oversampled, DEAL prioritises the most representative and diverse environments near the surface, distinguishing between majority and minority dipole orientations of interfacial water and ensuring a complete, non-redundant selection of local environments.

selection would create an imbalance between the portion of the PES that is most frequently sampled, whose configurations are likely just above the σ_l threshold, and the more sparsely sampled regions, which are probably under-represented and less accurately described in the dataset. This is more detrimental in the case of the reaction paths, where the basins are more sampled than the transition states, but the latter are usually the configurations with higher uncertainty. DEAL overcomes this problem by reducing the redundancy in the sampling of the basins and valorising the few points associated with transition states. Indeed, even if DEAL selection does not explicitly consider the collective variables of the reaction path, it is able to distinguish the different states and provides a uniform sampling along all the reaction

paths. Being able to identify the transition states at higher uncertainty, DEAL focuses on the minimisation of the error for these geometries, achieving a uniform accuracy between reagent/product and reactive configurations.

After the filtering via GP, the standard active learning cycle carries on with the *ab initio* labelling for the selected configurations. This iterative procedure (training, sampling, data-efficient selection and DFT labelling) is repeated until a reliable potential is obtained.

3.2.3 Transfer learning

Until now, we have assumed that developing an MLIP requires assembling a dataset and training a new model from scratch. However, building an ML potential for a specific system is often a complex and computationally demanding task, especially when the underlying DFT reference data is expensive to generate. At the same time, the growing popularity of ML potentials and the availability of large material databases (e.g., Materials Project) for training general-purpose models have enabled new paradigms, such as transfer learning. Instead of starting from scratch, transfer learning adapts an existing ML model as the foundation for a new one [130–132]. By leveraging knowledge already captured by a pre-trained model, the learning process becomes faster and requires significantly less data. The new dataset is used primarily to tailor the inherited knowledge to specific materials, phases, or thermodynamic conditions that share similarities with the original training domain. This approach is particularly valuable in scenarios of data scarcity, where generating and labelling configurations at the *ab initio* level is costly.

Under the domain of transfer learning, various strategies exist for building MLIPs, including the fine-tuning of general-purpose models [115], the transferring from models trained on classical data [133] or large open-access *ab initio* databases [134, 135], and the adapting across chemical elements [136] or different DFT functional approximations [137, 138]. For example, the recently proposed Franken scheme [126] performs transfer learning from GNN models trained on large datasets, i.e., universal MLIPs. Its core idea is to extract descriptors (also called 'backbone') by concatenating node features up to a chosen interaction layer from a pre-trained GNN. These descriptors serve as input to a kernel-based regression block (using the RF approximation; see Section 3.1.3). This design combines the expressivity of deep neural network descriptors with the data efficiency of kernel regression. Importantly,

the backbone is inherited without retraining, and only the kernel regression block is optimised, drastically reducing the amount of data required. The success of this approach depends on the quality of the GNN used as the backbone. In the original work [126], Franken employs MPNN models from the MACE-MP0 family, which are general-purpose potentials trained on the MPTraj database [139]. This dataset includes numerous static calculations and structural optimisation trajectories from the Materials Project, primarily small periodic unit cells of inorganic crystals with some molecular components, computed at the DFT level using the PBE functional. MACE-MP0 models already exhibit strong generalisation, enabling stable MD simulations for diverse systems and reproducing chemically sensible structures, reactions, and transformations. Transfer learning allows Franken to inherit these capabilities while adapting to specific systems using a much smaller dataset. For instance, with only a few tens of configurations, Franken achieves stable and accurate MD simulations of bulk water and metal/water interfaces, outperforming the original MACE-MP0 model. This exceptional data efficiency makes Franken a cornerstone of the workflow proposed in Chapter 4.

In summary, in this chapter we presented the main concepts and practices that will be employed to generate machine learning interatomic potentials in the following chapters. We will develop ML potentials using MACE, a message-passing neural network whose descriptors are based on the atomic cluster expansion (ACE) formalism. We will not only train ML models from scratch, but we will also follow transfer learning strategies as proposed in the Franken method. This approach merges together the expressiveness of pre-trained descriptors built upon the message-passing architecture of MACE with the data efficiency of kernel regression methods. The limitations of kernel approaches in the case of large datasets are overcome in Franken through the random feature approximation approach. Given the previously outlined relevance of data within the ML framework and the challenges posed by the construction of comprehensive datasets, we will place particular emphasis on active learning strategies. Specifically, we will adopt the DEAL protocol, which integrates a *query-by-committee* pre-filtering stage with a non-redundant configuration selection driven by the Bayesian variance of a Gaussian process, as implemented in the FLARE code. All the elements discussed in this chapter will be leveraged in TRECI, our proposed workflow to generate ML potentials for electrochemical interfaces, as presented in Chapter 4.

Chapter 4

TRECI: transfer learning for electrochemical interfaces

The development of ML potentials, here referred to as ML force fields (ML-FFs) to avoid confusion with the electrode potential, forms the core of this thesis. While ML-FFs are an established tool in thermocatalysis [140–142], their application to electrochemical systems [143] is still in its infancy and has only recently gained significant attention.

Early ML-FFs for electrochemistry largely ignored the effect of applied bias, focusing instead on the potential of zero charge (PZC), i.e., the electrode potential in the absence of excess surface charge [144–146]. Such models fail to capture the key driving force behind electrochemical processes, making them unrepresentative of experimental conditions. To address this, later approaches introduced constant-charge schemes [147], followed by constant-potential ML models employing hybrid explicit–implicit solvent representations [148–150]. These efforts primarily focused on adapting ML architectures to incorporate explicit bias-related inputs (e.g., work function, interfacial electric field, or excess electrons). However, these studies rely on DFT setups, fixed ion numbers, implicit or hybrid explicit-implicit solvation, and low-level functionals such as PBE or rPBE+D3, similar to those used in AIMD. Since these setups are computationally affordable, they are the predominant choice for AIMD, even if it is known that they might introduce inaccuracies and artifacts (cf. Section 1.3.3). Consequently, ML-FFs trained on such ground truth inherit these limitations already present in AIMD studies.

In this thesis, we adopt a fundamentally different paradigm. To bridge the gap

between experimental observations and atomistic simulations and to fully exploit ML-FFs beyond AIMD constraints, we employ more advanced and accurate (though computationally expensive) *ab initio* setups: fully explicit solvent models, meta-GGA functionals, and rigorous constant-potential schemes. Here, ML-FFs serve not only to extend spatial and temporal scales with respect to AIMD but also to enable high-fidelity descriptions of electrochemical interfaces not affordable with previous methods. This challenge, using costly yet more accurate ground truth, motivates the development of a novel, data-efficient strategy rather than relying on existing ML schemes. The outcome is TRECI (TRansfer Learning for ElectroChemical Interfaces), a framework that combines transfer learning and active learning to construct ML-FFs for electrified metal/water interfaces with unprecedented accuracy and efficiency.

In this chapter, we discuss the main elements of the TRECI strategy (Section 4.1). Then, in Section 4.2 we show how to apply this scheme to develop ML-FFs at different applied potentials for a copper/water interface, a representative electrode system used in many electrocatalytic applications [19, 151]. The physical insights enabled by TRECI-based molecular dynamics will be discussed in Chapter 5. The results presented here are detailed in our submitted publication “Electrochemical Interfaces at Constant Potential: Data-Efficient Transfer Learning for Machine-Learning-Based Molecular Dynamics” [34], while the developed code supporting the results of this chapter is freely available at <https://github.com/michelegiovannibianchi/TRECI>.

4.1 Data-efficient workflow for electrochemical interfaces

The workflow of TRECI to provide a high-fidelity description of electrified interfaces in a data-efficient manner relies on three core pillars: 1) a high-quality quantum mechanical description of the interface, 2) an innovative transfer learning framework for building ML-FFs and 3) an effective active learning strategy (Fig. 4.1).

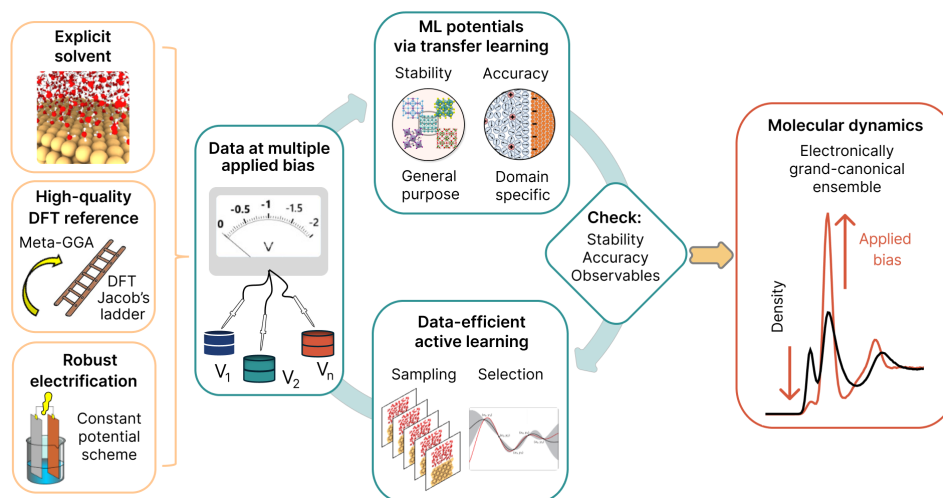


Fig. 4.1 General scheme of TRECI for electronically grand-canonical machine learning force fields: the pillars of the methods are a high-quality dataset at multiple applied bias values, a data-efficient ML architecture via transfer learning and an effective active learning approach. Reproduced from [34].

4.1.1 High-quality description at multiple applied bias values

A high-quality dataset and rigorous labelling are essential elements to provide an accurate description of electrified interfaces via ML-FFs. Any artifact in the solvent representation or ambiguities in the definition of the electrode potential can introduce inconsistencies in the labels, which become particularly detrimental when working with the small datasets employed in this work.

TRECI achieves a high-quality modelling of the interfaces via a) a comprehensive description of the solvent dynamics, b) an accurate treatment of the liquid water network, and c) a robust electrification scheme (cf. left side of Fig. 4.1).

The comprehensive description of the solvent is achieved by following a fully explicit solvent approach for capturing in the dataset both the overall dielectric response of the solvent and the localised interactions between molecules and the surface (cf. Section 1.3.3). Our approach carefully avoids schemes in which an artificial region of vacuum or an implicit solvent is on top of the explicit solvent, as, even if commonly employed in other ML-FF works, they could introduce artifacts at this interface between two domains.

A high-quality description of the liquid water network implies the choice of advanced DFT functionals, due to the well-known shortcomings of the common PBE-GGA approximation in water-based systems [79, 152]. GGA functionals usually tend to

overstructure the network of liquid water since they do not properly describe hydrogen bonding and dispersion forces, resulting in a too rigid structure. Simulating water at room temperature at the PBE level reproduces a behaviour more similar to an ice-like phase rather than a liquid and highly dynamic system. A common "workaround" is to increase the simulation temperature above ambient conditions (e.g., 400 K), but this solution is not viable in electrocatalysis, since any process will be activated by a combination of the increased temperature and the electric field, not reproducing the experimental conditions. For this reason, we consider the SCAN functional as the DFT reference due to its capability of correctly reproducing the liquid water behaviours (cf. Section 2.1.2).

The inclusion of the effects of the applied bias in the framework of ML-FFs represents the most crucial element. This is realised within an electronically grand-canonical ensemble using the "double reference method" [66]. By employing this scheme, TRECI constructs multiple datasets corresponding to distinct PESs, each associated with a well-defined and macroscopically accessible quantity, the bias value. The "double reference method" determines the electrode potential by using a series of potential references computed in auxiliary systems. This approach eliminates artifacts from solvent/vacuum interfaces and avoids inconsistencies between the energies computed in the systems with and without extra electrons (see Appendix A and the developed *Python* code at <https://github.com/michelegiovannibianchi/DoubleReferenceMethod-FCP-calculator> for more details on this method and how it is practically integrated into the TRECI workflow). As a result, it provides a consistent connection between surface charge and electrode potential, removes ambiguities in data labelling, and guarantees that each PES is uniquely identified. Such a rigorous electrification scheme represents the main step ahead with respect to other ML works.

To the best of our knowledge, a single framework integrating all these three modelling choices has not been realised in prior studies, largely due to the high computational cost associated with each component: SCAN is approximately 2.5 times more expensive than PBE, and a simulation with the "double reference method" is up to 10 times more costly than a calculation without external bias due to the need to evaluate the auxiliary systems for each configuration. When used together, these methods become prohibitively expensive, calling for highly data-efficient approaches.

4.1.2 ML force-fields via transfer learning

Given the high cost of generating such data across multiple potentials, training independent ML-FFs for each applied bias value would be highly inefficient. TRECI instead adopts a transfer learning paradigm (cf. Section 3.2.3) that reuses information from models trained in the absence of an electric field to efficiently learn the corresponding PESs at constant applied bias.

Specifically, we draw on the recently introduced Franken framework [126]. As outlined in Chapter 3, the strength of Franken lies in combining the data efficiency of kernel regression methods with the expressivity of pre-trained descriptors from MPNN models. Within TRECI, we extend this idea into a *multi-head* architecture, where a shared set of GNN-based descriptors is reused to simultaneously learn several PESs corresponding to different applied biases. Each head is trained to address a specific bias value (Fig. 4.2). Because only a small regression block is bias-specific, the overall data requirement is drastically reduced. We are able to train seven RF-based readouts for potentials ranging from -0.50 V to -2.00 V vs SHE (with a spacing of 0.25 V). This approach is, however, inherently scalable, enabling the simultaneous generation of a greater number of regression blocks that encompass an even broader spectrum of applied biases. In this way, although each readout can handle only a single potential case, the overall model spans a broad range of possible values and can be readily adjusted to a new electrical bias value with only a small amount of additional data.

Furthermore, TRECI generalises the Franken concept beyond fine-tuning of general-purpose ML-FFs: it also leverages domain-specific models trained for the same system at the PZC, in the absence of an applied field. While this approach requires training an additional model from scratch, the computational cost of labelling configurations at the PZC is significantly lower (approximately 10 to 20 times less than performing full constant potential calculations across seven target bias values). Thus, the gain in accuracy achieved with fewer grand-canonical DFT evaluations when employing a domain-specific backbone (cf. Section 4.2.3) provides sufficient justification for the additional effort involved in developing the PZC model. In the following we refer to models constructed using general-purpose descriptors from the MACE-MP0 family and domain-specific ones as Franken-MP0 and Franken-PZC, respectively. A recap table of the different backbones employed in this thesis is also reported in Appendix B. As detailed in Section 4.2.3, this innovative combination of

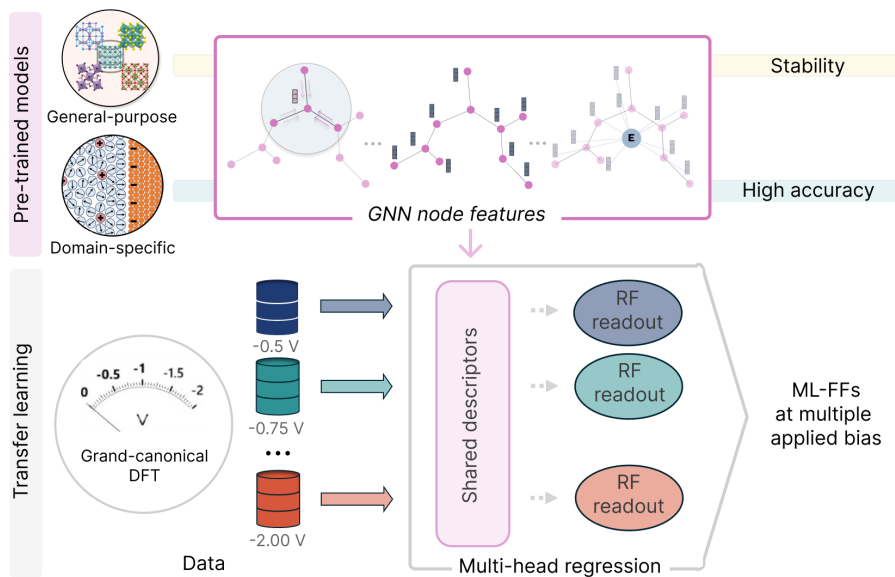


Fig. 4.2 TRECI strategy for the training of the constant potential ML-FFs via transfer learning: the node features of a pre-trained general-purpose or domain-specific GNNs are used as descriptors for a multi-head Franken model. Each readout block implements a large-scale kernel regression model targeting an applied bias value. Kernel functions are approximated by means of Random Fourier Features (RF). Reproduced from [34].

general-purpose and domain-specific models provides a route from stable exploratory simulations to accurate production runs.

4.1.3 Data-efficient active learning

The efficiency of TRECI ultimately hinges on identifying the few thermodynamically relevant configurations at each target bias. Simply re-labelling equilibrium structures obtained in the absence of an electric field is insufficient, as the applied potential can stabilise entirely different interfacial arrangements. To address this, TRECI adopts an iterative data acquisition strategy based on active learning (cf. Section 3.2.2). The dataset is progressively expanded by exploring the PESs with preliminary ML-FFs and selecting a limited number of new configurations for grand-canonical DFT labelling.

Specifically, for the selection, we exploit the capabilities of the DEAL protocol [129]. Differently from the DEAL approach described in Section 3.2.2, a two-step DEAL selection is performed (Figure 4.3). First, configurations are filtered independently for each applied bias; then, the selected configurations are merged across all bias

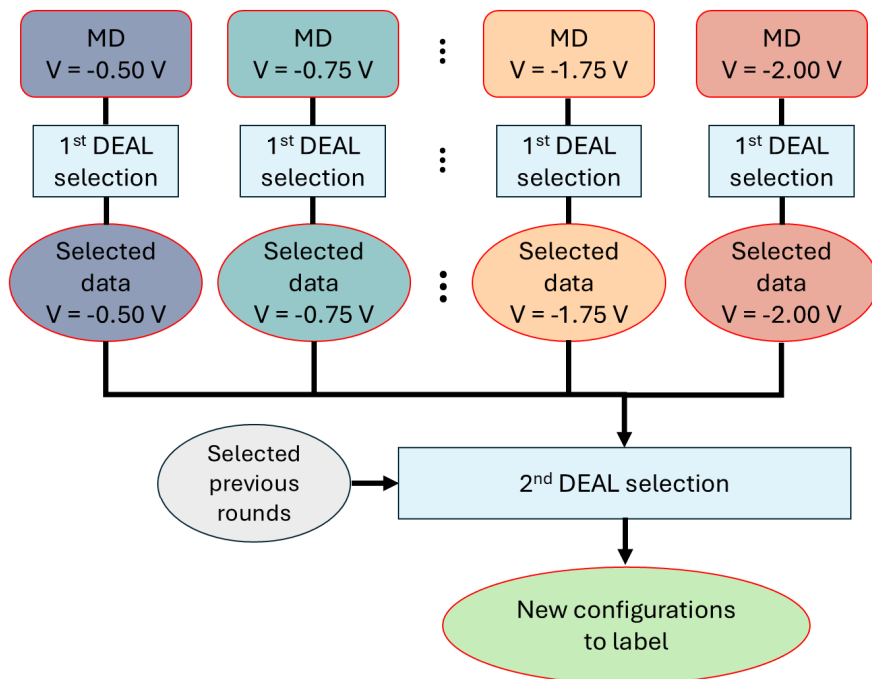


Fig. 4.3 Scheme of the two-step DEAL selection employed in the TRECI workflow during the active learning strategy for constant potential ML-FFs. The dataset of the previous rounds is composed of all the geometries for all the potential values collected in the previous loops of active learning.

values and screened again alongside the existing dataset. This second DEAL selection reduces redundancy among configurations sampled at the different biases and with respect to the existing dataset. This approach yields a compact and non-redundant dataset, minimising the number of expensive DFT calculations at constant potential, while ensuring efficient coverage of the relevant configurational spaces.

4.1.4 Complete strategy

Here we summarise the steps of TRECI to construct ML-FFs at constant potential. The workflow begins with the construction of an ML-FF trained for the PZC case using the MACE architecture (cf. Section 4.2.1). The PZC model is developed to have domain-specific descriptors useful for the transfer learning. In addition, this gives a starting dataset for the ML-FFs with applied bias and serves as a baseline for the results in the absence of an electric field.

An iterative strategy composed of four steps is then applied (Figure 4.4). The cycle

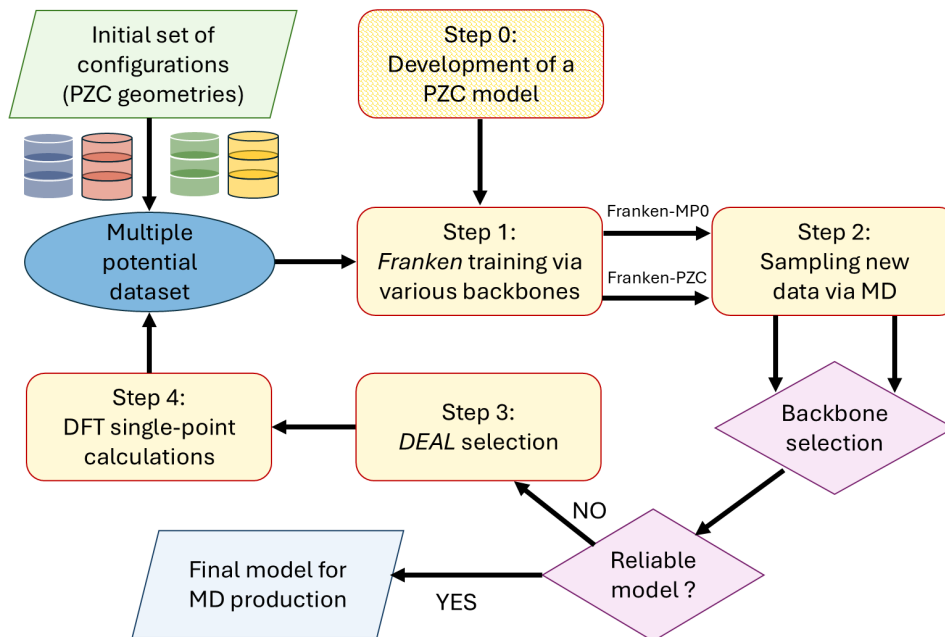


Fig. 4.4 Scheme of the TRECI workflow for constant potential ML-FFs. The development of a model for the PZC case can be considered as a preliminary step. The double arrows connecting the blocks identify the steps for which models with different backbones (general-purpose Franken-MP0 vs domain-specific Franken-PZC) are considered at the same time, before the selection of the optimal one at each round. Backbone selection is based upon the criteria of prediction accuracy vs MD stability. A model is considered reliable if accuracy in the prediction as well as physical observables reaches convergence.

starts with a dataset at multiple applied potentials computed via grand-canonical DFT calculations. Using this data, ML-FFs are optimised for the corresponding constant bias values using Franken transfer learning, employing both general-purpose (from the MACE-MP family) and domain-specific descriptors (from the model trained at the PZC). The models are then benchmarked by employing them in explorative MD runs to sample the target PESs. Models with different backbones are compared considering stability (i.e., the capability of generating stable MD trajectories) and accuracy of the predictions. If domain-specific descriptors fail to yield stable MD trajectories, the workflow proceeds using the general-purpose ones. Instead, once enough data are available and Franken-PZC models become robust, they are preferred due to their superior accuracy. Then, new atomic configurations from MD trajectories are selected through a data-efficient active learning protocol. The selected configurations are labelled with grand-canonical DFT calculations, and the dataset is expanded with new data. The cycle continues until both the accuracy

of energy and force predictions, as well as relevant physical observables, reach convergence. Only at this point, the ML-FFs are considered sufficiently reliable for use in production-level MD simulations.

4.2 Validation of the workflow

We validate the workflow of TRECI by developing ML-FFs for a copper-water interface at different negative potential values. Notice that this system represents a real-case scenario, as it is a typical cathodic system used in electrochemical applications such as the CO₂ electroreduction reaction. We will show that following the presented strategy, after a few active learning iterations and the collection of one thousand structures at constant potentials, TRECI achieves stable, accurate and converged results. Stability is evaluated by inspecting the MD trajectory for a duration of at least 1 ns and confirming that no unphysical phenomena occur. Accuracy is quantified using force predictions evaluated on an independent dataset, while convergence is assessed by analysing the behaviour of a representative physical observable, such as the density of interfacial water, and verifying that it has reached a stable, reproducible profile. As illustrated in Figure 4.5 and further discussed in Chapter 5, the presence of an interface and a surface charge alters the water density, leading to a modulation of the spatial profile characterised by pronounced peaks. At least two main regions can be distinguished: region I, where molecules are chemisorbed and the O atoms share electrons with the surface, and region II, where water is physisorbed and the partially positive H atoms are orientated toward the metal. A more extensive analysis of the properties of these regions is presented in Chapter 5. At this stage, it is crucial to recognise that the strength of these two types of interactions varies with the surface charge, with Cu···H electrostatic attraction increasing and Cu···O repulsion becoming stronger as the negative surface charge grows. Consequently, the peak height of the O density in these regions is a highly sensitive physical observable for characterising the metal–water interaction that the ML-FFs have to learn.

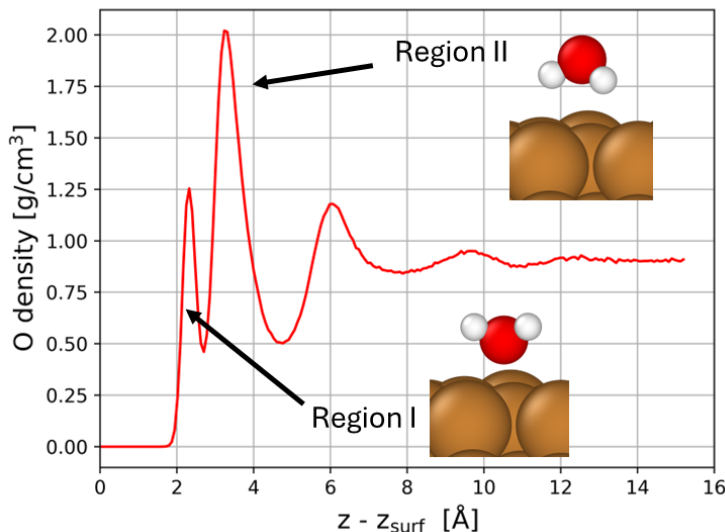


Fig. 4.5 Oxygen density profile of interfacial water with the main regions in the solvent highlighted. The insets show the typical orientation of water molecules in these peaks.

4.2.1 Preliminary step: model for the PZC

The development of an ML-FF for the PZC can be considered as a preliminary step for the models at constant potential. For the PZC, we follow the state-of-the-art active learning strategy discussed in Section 3.2.2 to generate an ML-FF model from scratch. The preliminary dataset on which the active learning cycle for the PZC starts is assembled following different strategies (use of an accurate pre-existing dataset at SCAN level for pure bulk water, AIMD at low PBE accuracy level and universal ML-FFs for Cu/water systems). In all these cases, AIMD at the SCAN level is never performed due to its computational cost and inefficiency in sampling uncorrelated structure (cf. discussion in Section 3.2.2).

Starting from this initial pool of configurations, the training dataset is iteratively refined through the DEAL scheme to ensure a broad and uniform coverage of the relevant configuration space while dramatically reducing the number of costly *ab initio* calculations. To access a broader portion of the configurational space, we additionally carry out a series of MetaDynamics simulations (see OPES-Explore in Section 2.3.5), in which the biased CVs correspond to the number of water molecules present in each region of the interfacial solvent. This approach allows us to probe a range of densities in the different interfacial regions and thus reliably identify the optimal density associated with the most thermodynamically stable state.

The ML-FF for the PZC is developed via the MPNN MACE architecture (cf. Section 3.1.3). During the construction of the ML-FF, an architecture with equivariant messages is employed to better optimise the stability of the model in case of a limited-size database. Once the dataset is complete, the final ML-FF is trained with invariant messages, resulting in a reasonable balance between accuracy and computational efficiency. At the end, a stable and accurate model for the Cu (111)/water interface at the PZC is achieved with only about 5300 configurations, including bulk water systems, bare Cu surfaces and Cu/H₂O structures.

4.2.2 Starting point for the constant potential models

Developing the model with an applied bias becomes more challenging, primarily because of the increased computational cost of labelling and, even more critically, the difficulty of constructing an initial dataset for constant-potential conditions. The approaches previously used to sample the initial dataset at the PZC are no longer applicable in this context. To the best of our knowledge, there are currently no large datasets labelled at constant potential (particularly using the double reference method), nor are there any universal ML-FFs designed to operate under these conditions. In addition, AIMD simulations, even at the relatively inexpensive PBE level, are entirely impractical within the double reference framework.

For this reason, we start the model at constant potential by sampling approximately 250 non-redundant configurations from the PZC dataset using the GP method implemented in DEAL. These configurations are then relabelled at constant potential, forming the preliminary dataset for the active learning cycle. In this context, active learning becomes even more critical than in the PZC case. The reason is that the configurational space covered by the PZC dataset does not fully represent systems under applied potential. Rather than spending substantial computational resources to label a large number of PZC geometries at constant potential, it is more efficient to first develop stable, though not necessarily highly accurate, constant-potential ML-FFs. These initial models enable early exploration and sampling of the PES regions relevant to the applied potential conditions. However, training stable models with only a few hundred data points remains a significant challenge.

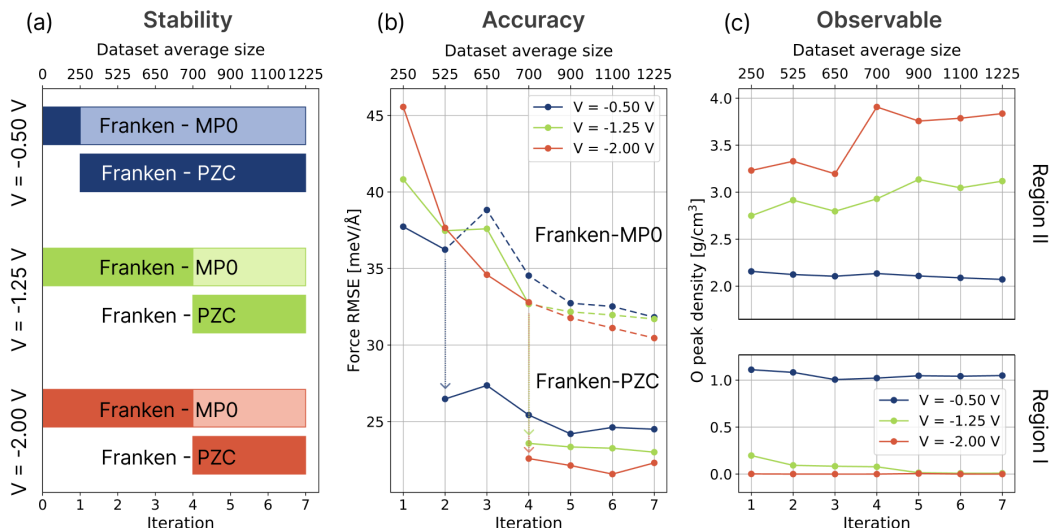


Fig. 4.6 Evolution of the learning process for TRECI models at representative target biases during the active learning. (a) Stability range of the different descriptors (Franken-MP0 vs Franken-PZC): coloured bars show the iterations at which the ML-FFs are stable; dark colours identify the descriptors employed for the sampling at the different iterations. (b) Force RMSE for Franken-MP0 and Franken-PZC. Solid lines identify the descriptors employed for the sampling. The accuracy is not reported when the models are not stable. (c) Peak value of the oxygen density in the different interfacial regions. Reproduced from [34].

4.2.3 From stability to accuracy

Starting from a few hundred configurations selected from the PZC dataset and directly targeting stable MD trajectories requires a careful choice of the ML architecture. This small initial dataset proves insufficient for training not only a standard MACE model (cf. Section 4.2.5), but also domain-specific Franken-PZC ML-FFs: trajectories generated with these models lack stability. In contrast, the Franken-MP0 consistently yields stable results across all applied potentials and throughout all active learning cycles (Figure 4.6a). For this reason, we start expanding the dataset in an active learning framework using configurations from MD trajectories generated with Franken-MP0 models. As the dataset is enlarged, Franken-PZC ML-FFs begin to produce stable results: at low applied bias ($V = -0.50$ V), stability is achieved after a single active learning cycle, while more negative potentials require additional iterations (Figure 4.6a). In particular among domain-specific models, we initially use equivariant backbones (Franken-PZC-L1), e.g., from a MACE model with equivariant messages ($L = 1$), trained exclusively on geometries at PZC. Equivariant descriptors are more expressive, and Franken models based on them deliver stable

results with smaller datasets compared to their invariant counterparts. However, equivariant architectures are computationally more demanding, increasing inference costs. For this reason, we then explored invariant descriptors trained on larger and more variegated datasets to expand the extrapolation capability of the model, making it more robust. One such case is a backbone from a MACE ML-FF with invariant messages ($L = 0$), trained on a combined dataset comprising PZC geometries and configurations compatible with applied potentials (Franken-PZC+Charged-L0). For the training of this backbone, geometries compatible with applied potentials were relabelled without extra charge to maintain consistency with the PZC dataset. This relabelling incurs no additional cost, as this DFT calculation is already required for the auxiliary systems in the “double reference method”. In this way, at each cycle of active learning, the new geometries, labelled both at the PZC and at constant potential, are used to improve the descriptor (dataset with PZC labels) as well as the regression part (dataset with constant potential labels) of the model. Indeed, the active learning process enables not only the progressive optimisation of the readout blocks but also the retraining of descriptors using the expanded dataset, which includes the newly acquired geometries associated with the applied potential from the previous iterations. We additionally direct the reader to Table B.1 in Appendix B, which provides a recap of the characteristics of the different domain-specific backbones that can be employed during the active learning cycles.

Once the Franken-PZC models have been proven to deliver stable MD trajectories, switching from general-purpose to domain-specific descriptors significantly improves performance, reducing prediction error by approximately 30% (lower RMSE) and ensuring consistent accuracy across all bias values (see Figure 4.6b and Table 4.1). Remarkably, the accuracy achieved using PZC descriptors is comparable to that of the PZC MACE model, yet TRECI attains these results with only 25% of the data used in the PZC dataset. Additional analyses validating the accuracy of TRECI constant-potential models and comparing them to the PZC MACE model are provided in Appendix B.

4.2.4 From accuracy to convergence

When using Franken-PZC models, the force accuracy rapidly converges to a plateau (Figure 4.6b). Despite this convergence, we perform a more rigorous check by analysing the evolution of key physical observables. As a representative quantity,

Table 4.1 Root-Mean-Squared-Error (RMSE) in $\text{meV}/\text{\AA}$ for the prediction of forces using TRECI models at different bias values, with descriptors from either MACE-MP0 (Franken-MP0) or our domain-specific PZC model (Franken-PZC). For reference, the corresponding RMSE for the PZC model trained from scratch with 5323 configurations is $17.4 \text{ meV}/\text{\AA}$.

Case [V vs SHE]	−0.50	−0.75	−1.00	−1.25	−1.50	−1.75	−2.00
# configurations	1202	1227	1173	1265	1305	1264	1260
Franken-MP0	31.8	30.9	32.2	31.7	29.8	29.9	30.5
Franken-PZC	24.5	23.7	24.3	23.0	21.3	22.1	22.3

we monitor the peak heights in the oxygen density profile at the interface, which are highly sensitive to structural changes induced by the applied bias. The analysis of these quantities during the active learning provides insights into the learning process (Figure 4.6c). As the models are progressively refined, the peak height of region I (chemisorbed molecules) decreases while the second one (physisorbed water) increases until both converge to stable values. This modulation of the peak density is a well-known effect of the applied potential, and these variations during the active learning cycles are an expected behaviour for a transfer learning process from an electric-field-agnostic model to ML-FFs in the presence of applied bias. By tracking these physical observables, we notice that even after the convergence of the force RMSE, approximately three additional active learning iterations are required for the peak densities to reach a plateau. At this point, we further validate the convergence of the solvent density profiles, comparing TRECI models built upon different descriptors (Figure 4.7). Even if the starting models for the transfer learning are different, the ML-FFs predict reciprocally consistent observables, confirming the effectiveness of the TRECI strategy to generate accurate and reliable models with small datasets. Interestingly, after completing the active learning, we observed on the final dataset that even domain-specific MACE descriptors (Franken-PZC-L0) with invariant messages, but trained only on a smaller dataset with respect to Franken-PZC+Charged-L0 (i.e., only the PZC geometries), can guarantee stable results. This demonstrates that stable performance can also be achieved with a thousand of data points without resorting to expensive equivariant models or continuously retraining the descriptors.

Concluding, based on the results of TRECI to develop the ML-FFs for the Cu/water interface, we recommend using general-purpose (Franken-MP0) first and equivariant domain-specific descriptors (Franken-PZC-L1), as soon as stable, during

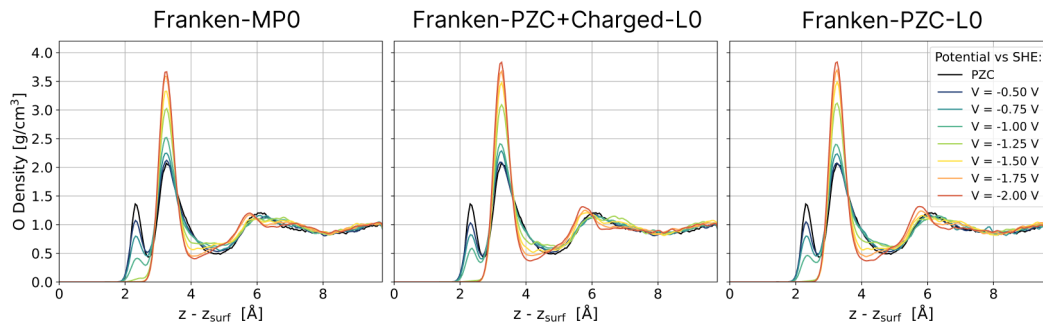


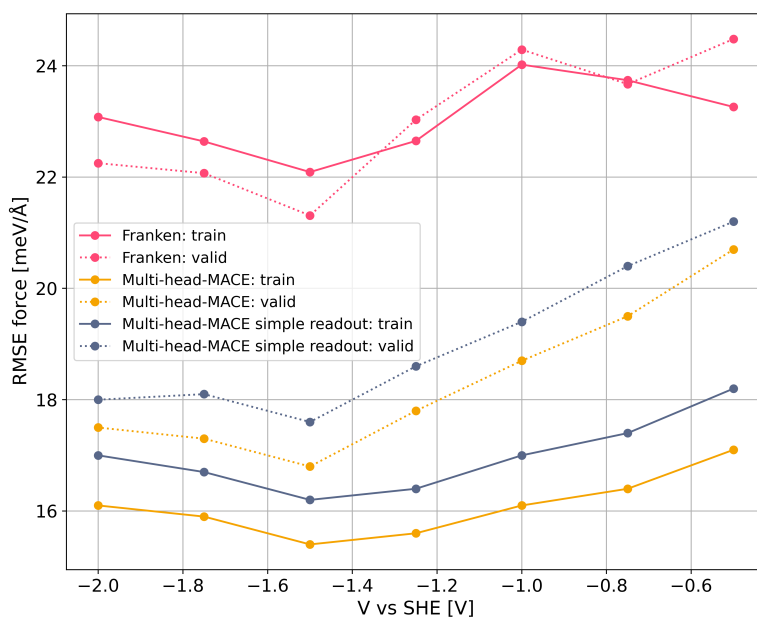
Fig. 4.7 Comparison of the oxygen density profiles of interfacial solvent at different applied potential values, as predicted by Franken models with different descriptors (cf. Appendix B for the characteristics of these descriptors). The same PZC profile computed with a standard MACE model is reported in each panel for reference. Reproduced from [34].

the active learning process when an early stability of the MD trajectory matters. Then we suggest transitioning to the invariant domain-specific backbones (Franken-PZC+Charged-L0 or even Franken-PZC-L0) at the last iterations of active learning and, above all, in the production runs due to their lower computational cost during inference. A recap of the suggested sequence of backbones to use during the active learning is also summarised in Table B.1 in Appendix B.

4.2.5 Comparison of different training approaches for the constant potential models

The success of TRECI is based on the Franken transfer learning strategy integrated in a multi-head architecture to enable multiple potential value predictions using a unique set of descriptors. We further benchmark this training strategy, comparing it with another approach based on transfer learning as well as multi-head architecture. This alternative method consists of the multi-head training implemented in the MACE suite. We use the multi-head MACE to train from scratch and simultaneously the PZC model and the 7 constant potential cases. Hence, 8 datasets are used to train the message-passing neural network implementing unique descriptors for all the models. If all the datasets are used to train the same descriptors, then each set is used to separately train a specific readout multi-layer perceptron (MLP, i.e., a head) for a total of 8 heads. Similarly to Franken, the strategy of the multi-head MACE focuses on maximising the use of the PZC dataset to construct descriptors, which are then utilised for the constant potential models, achieving a transfer learning effect

from the PZC to the applied potential cases. Nevertheless, the multi-head-MACE descriptors are also trained using each of the constant potential sets. The contribution of the constant potential data in the construction of the descriptors is associated with the relative dimensions of the applied bias datasets with respect to the PZC one. This contribution varies from iteration to iteration as new configurations with applied fields are collected. Another relevant difference of the multi-head-MACE



Potential	Iteration 1	Iteration 2	Iteration 3	...	Iteration 7
-0.50 V	✓	✓	✓	✓	✓
-0.75 V	✓	✓	✓	✓	✓
-1.00 V	✓	✓	✓	✓	✓
-1.25 V	✓	✗	✓	✓	✓
-1.50 V	✗	✗	✓	✓	✓
-1.75 V	✗	✗	✓	✓	✓
-2.00 V	✗	✗	✓	✓	✓

Fig. 4.8 Analysis of the performance of the multi-head-MACE training. (Top) Comparison of the force RMSE in the training and validation set for the final multi-head MACE model vs the corresponding Franken model. The “multi-head simple readout” is a model with the width of the layers of the radial MLP in the readout part reduced from 64 to 32. (Bottom) Analysis of the stability of the results for multi-head-MACE models during the different active learning iterations. Symbols ✓ and ✗ stand for stable and non-stable results, respectively. For cases identified with symbol ✓, the stability is achieved only with the “multi-head simple readout” model.

training that makes this approach less attractive than Franken is the training cost: the first takes about 96 GPU-hours for the full model since it also trains from scratch the PZC case, and globally it uses a large dataset (about 12600 configurations, including the PZC and the 7 constant potential sets). We stress this time refers to the training of a single architecture, implying that additional training is required to optimise all the hyperparameters of the GNN. On the contrary, seven constant-potential Franken models totally require about 24 GPU-hours with hyperparameter tuning corresponding to testing 16 different architectures. The benchmark of stability (Table 4.8) and accuracy (Figure 4.8) of the results for the multi-head MACE vs Franken also highlights the pros/cons of each approach. Regarding stability, as shown in Table 4.8, the multi-head-MACE approach is unable to grant stable results in the initial cycles of active learning, in particular, for the high negative potentials. A model with simpler readouts (i.e., fewer neurons in the MLP) can only partially mitigate the lack of sufficient data at the beginning of the active learning. Observe that, due to the limitations of the multi-head MACE method in achieving stable results at more negative potentials, it is especially difficult to provide new relevant configurations for these potential values, thus addressing the data scarcity and stability issues. On the contrary, Franken, as used in TRECI, is more versatile, enabling stable simulations from the first cycles of active learning by using general-purpose descriptors. Regarding accuracy, Figure 4.8 shows the RMSE on the training and validation set of the final Franken and multi-head MACE models. Even if the latter is characterised by a slightly smaller error ($\approx 5 \div 6$ meV), the Franken model presents less overfitting (difference between training and validation error). A model with a smaller number of neurons in the readouts is also considered without evident improvements. This proves that the kernel-based method of Franken is more data-efficient than the MLP of the multi-head MACE, which requires additional data to be more reliable.

From this comparison it is evident that Franken architecture is more suitable than the multi-head MACE one to generate the constant potential models with limited data due to the lower computational cost for the training, a minor tendency to overfitting and, above all, a superior stability in the results during the different iterations of active learning.

4.3 Conclusion

In this chapter, we presented TRECI, a general, data-efficient workflow for developing ML-FFs to investigate electrochemical interfaces under external electrical bias. This enables a high-fidelity description of these systems by combining a fully explicit solvent model, a DFT-accurate representation of the liquid water network via the SCAN functional, and a constant potential framework that avoids explicit-implicit solvent compromises.

To address the computational cost of this rigorous setup, TRECI leverages a data-efficient transfer learning algorithm for the ML-FF training and an effective active learning protocol. To ensure model stability in the early stages of the ML-FF development, we perform transfer learning from general-purpose potentials. In subsequent iterations, we refine the ML-FFs by transferring information from domain-specific models to achieve a superior accuracy. This dual strategy enables the construction of robust and precise ML-FFs across a wide range of potential values. Our workflow achieves stable MD trajectories with only a few hundred configurations, reaches the same accuracy as models trained from scratch, and converges physical observables with a final dataset of just one thousand configurations.

The validity of TRECI models will be shown in the next chapters, in which, by applying this workflow, we will develop ML-FFs for studying various phenomena at the electrified Cu/water interfaces. In Chapter 5 we will use the developed ML-FFs to analyse the effect of the applied potential on the structural and dynamical properties of water in contact with Cu (111) and Cu (100) surfaces. While in Chapter 6 we will show the effectiveness of TRECI in developing ML models targeting relevant reaction steps of common electrochemical reactions. In both cases, the accuracy and scalability of the developed ML-FFs will be used to reveal detailed properties of electrified metal/water interfaces, achieving a level of resolution rarely accessible with conventional approaches.

Aside from the specific applications in the following chapters, the relevance of TRECI has a broader impact. The applicability of the TRECI workflow is not constrained by our specific *ab initio* setup used for data labelling. Its modular design allows for easy customisation (in the choice of DFT functional, electrification scheme, or pre-trained model for transfer learning), making it adaptable to the requirements of diverse electrochemical interfaces. At the same time, the TRECI scheme is also able to face more complex scenarios, including surface defects, ions

and adsorbates to simulate specific *operando* conditions. While these more complex conditions introduce new challenges, such as accounting for long-range electrostatic interactions from charged species, the core strengths of TRECI remain unaffected. Its data-efficient and robust protocol provides a solid foundation for developing more advanced ML-FFs in which the TRECI models are a modular element within a more complex architecture. For these reasons, we believe that the TRECI strategy paves the way for high-quality, atomistic simulations of electrified interfaces.

Chapter 5

Properties of electrified copper/water interfaces via ML-accelerated MD

The capabilities of TRECI in developing ML-FFs for electrochemical interfaces, discussed in Chapter 4, are now employed to generate ML models tailored to describe via molecular dynamics simulations the interfacial properties of copper electrodes in contact with an aqueous solvent. As discussed in Section 1.2.3, copper electrocatalysts are commonly employed for CO₂RR applications to produce C₂₊ products; for this reason, in this chapter we will analyse the characteristics of water solvent in contact with Cu (111), the most stable surface termination, as well as Cu (100), which is the most promising one for the synthesis of C₂₊ species [19]. Since the CO₂ reduction requires the application of negative potentials at the cathode, we will focus on the effects of negative values of applied bias, considering the copper surface as the negatively charged electrode. Despite the standard electroreduction potential for water reduction in H₂ being -0.83 V vs SHE [153] and the possible formation of hydrogen adsorbates at negative potentials [154], the following analysis does not include the role of surface adsorbates and the effects of electrochemical reactions. Chapter 6 will include these elements in the discussion. Moreover, at this stage we consider only the effects of the electric field, while neglecting the presence of ions in the electrolyte. Hence, we are modelling the limit of a very diluted electrolyte. In this sense, the anisotropic nature of the solid electrode / liquid solvent interface as well as the additional surface charge due to the external potential are already responsible for relevant modifications of the structural and dynamical properties of the interfacial solvent, providing an interesting picture also without ions. The

analysis of these variations of the properties of interfacial solvent has a pivotal role in understanding not only the dielectric response of the electrolyte but also the surface reactivity of the electrode. To assess the perturbations caused by the electrode, we will examine how the structural and dynamical properties of the solvent vary with the distance from the metal surface. These variations are compared to the characteristics of a bulk water system (i.e., an isotropic system with only water molecules). We first perform a detailed analysis of the potential of zero charge (PZC) (Section 5.2). Following this, we study the effects of the application of different electric potentials (Section 5.3). We explore values of reducing potentials from -0.5 V up to -2.0 V vs SHE, the typical range employed in electroreduction studies [19].

The reliability of the results presented in this chapter is corroborated by additional analyses reported in Appendix C, which demonstrate the capability of the selected SCAN exchange–correlation functional and the employed ML architecture to accurately describe all relevant interaction types present in the considered interfacial system. The results of this chapter are partially presented in a submitted publication, “Electrochemical Interfaces at Constant Potential: Data-Efficient Transfer Learning for Machine-Learning-Based Molecular Dynamics” [34], and will be partially included in a future article in drafting.

5.1 Computational details

Before entering into the detailed analysis of the results of the molecular dynamics simulations at the PZC and at constant applied potential, we report some key computational details for performed atomistic simulations and the developed ML-FF models.

5.1.1 MD simulations

All the results reported in this chapter are obtained from the analysis of MD trajectories generated with our ML-FFs developed following the TRECI workflow discussed in Chapter 4. Structural properties are obtained in an NVT ensemble, while the dynamical properties are computed in the microcanonical ensemble to remove any spurious contribution of the thermostat. In all the cases, an integration step of 0.5 fs and the deuterium mass for hydrogen atoms are employed to properly describe

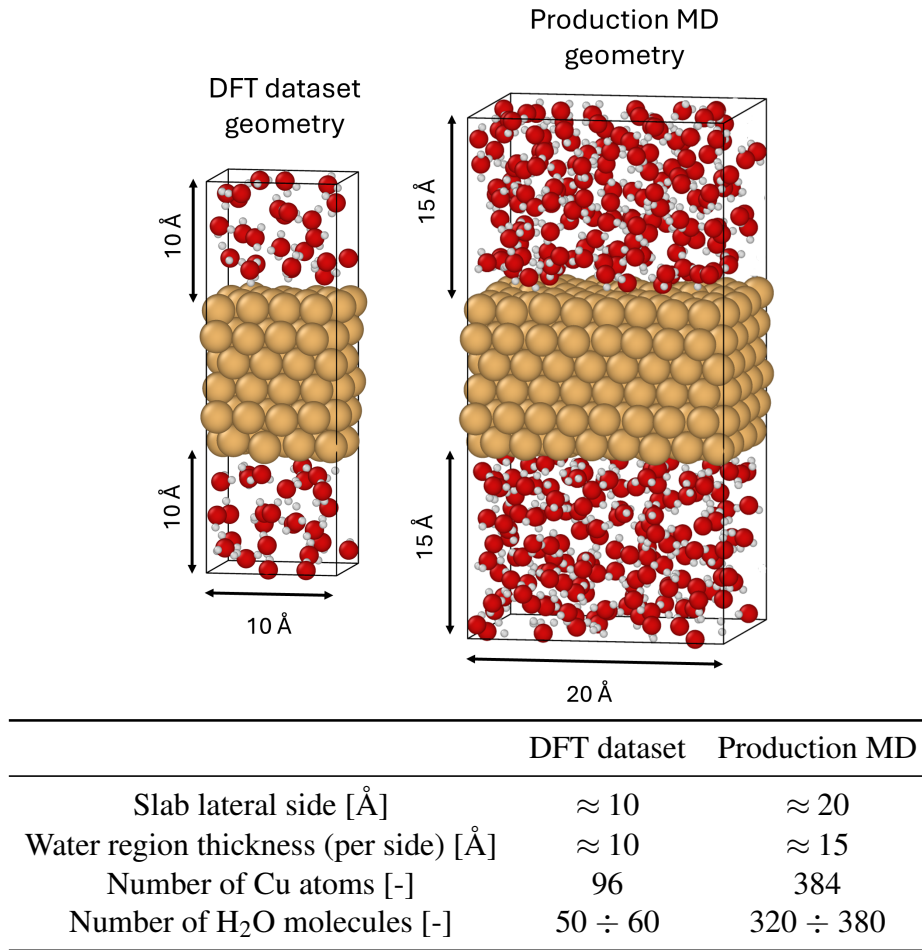


Fig. 5.1 System size comparison between the configurations of the dataset computed at the DFT level and the geometries of the production MD trajectories generated via ML-FFs. Main dimensions are quoted in the picture and reported in the table.

the dynamics of light atoms. MD runs are performed in a slab system composed of 6 layers of metal atoms, with the two innermost ones fixed into a symmetric double-interface slab. Periodic boundary conditions are imposed in all directions, and no regions of vacuum are present (cf. Figure 5.1). The lateral side of the slab is about 20 Å, corresponding to a (8×8) surface supercell. The metal slab is vertically separated by its periodic images by a region of 30 Å (15 Å per each side of the slab), filled with water molecules. It has already been proved that these system dimensions are sufficient to not cause finite-size effects [144], and they are markedly larger than the typical system size employed in our ML dataset and affordable at the DFT level (cf. Table 5.1). At the same time, the statistics for each observable discussed in this

chapter are gathered over a time window longer than 5 ns, which is significantly larger than the typical sampling timescale of AIMD.

5.1.2 Generation of the ML models

The ML-FF models are developed following the TRECI strategy reported in Chapter 4. The dataset for the ML models is collected at the SCAN exchange-correlation approximation level, considering a fully explicit solvent approach and applying the double reference method to reproduce the effects of the applied potentials. The ML model for the PZC case is developed from scratch using the MACE MPNN architecture: during the active learning, a model with equivariant features is considered for its superior expressivity, while for the final MD production runs are performed with an invariant model, resulting in a reasonable balance between accuracy and computational efficiency. The ML-FFs models at constant applied potential are trained following the Franken transfer learning strategy: during the active learning, general-purpose or domain-specific invariant backbones are employed, following the stability-accuracy criterion presented in Chapter 4. The final ML model has the domain-specific descriptors extracted from the MACE PZC model with invariant features (Franken-PZC-L0) and 16384 RFs. For both the PZC and the constant-potential cases, ML models with 2 interaction layers with a cut-off radius equal to 6 Å each are employed. A larger cut-off radius and/or a higher number of interaction layers are also tested, but the description of the Cu/water interface results is minimally influenced (cf. Appendix C.1). This appendix also presents an analysis of the impact of interactions extending beyond the receptive field of our ML-FFs (i.e., long-range interactions) since a growing number of ML-FF studies report the use of specific long-range schemes in electrochemical systems [147, 155]. In Appendix C.1 we identify the main source of long-range interactions in electrochemical interfaces: the ionic species. This appendix shows that systems without ions, such as the cases considered in this chapter, are characterised by an effective screening action of water molecules, minimising the effects of contributions outside the receptive field of the ML-FFs. This justifies why the results of this chapter are obtained without incorporating any explicit long-range interaction scheme into the ML model. Originally a model for the Cu (111) surface was developed both at the PZC and at constant applied potential. Then the ML models were generalised, adding to the dataset configurations for the Cu (100) case. With respect to the results reported in

Chapter 4 that refer only to the model for the Cu (111) surface, Table 5.1 reports the dataset size and the accuracy of the final models that are able to generalise among the two low-index surfaces.

Table 5.1 Root-Mean-Squared-Error (RMSE) in meV/Å for the prediction of forces using TRECI models for the Cu (111) and Cu (100) surfaces at different bias values [V vs SHE].

Case	PZC	−0.50	−0.75	−1.00	−1.25	−1.50	−1.75	−2.00
Data points	5912	1788	1983	2008	2120	2209	2143	1942
RMSE	15.7	23.2	22.6	22.9	22.6	22.2	23.0	23.25

5.2 Copper/water properties at PZC

In this section we report a detailed comparison of several structural as well as dynamical properties of Cu (111) and Cu (100) in contact with water at the PZC. The goal is the identification of the characteristics that distinguish the two surfaces.

5.2.1 Water density profile

The first information that is possible to examine is the density profile of the water solvent as a function of the distance from the surface (Figure 5.2). The contributions of oxygen and hydrogen atoms are reported separately. Looking at the O profile, it is possible to appreciate that there are at least three visible regions in the solvent, close to the interface. These regions are sequentially labelled as I, II and III, moving from the surface towards the water bulk. There are still some minimal fluctuations in the O density after 8 Å far from the surface, but they are completely smoothed out at 12 Å. Here, the solvent density is constant and equal to 1.02 g/cm³ (decomposed in oxygen and hydrogen contributions equal to 0.91 and 0.11 g/cm³, respectively). This value is similar to the densities (1.03 ÷ 1.05 g/cm³) computed at the SCAN exchange-correlation level for a pure water bulk [79, 156].

Focusing on the solvent regions close to the interface, it is possible to appreciate a higher degree of structuring in the solvent. This is evident as three peaks in the oxygen profile, where the density can even double its value with respect to the water bulk case. The specific values of this increase in the different regions for both

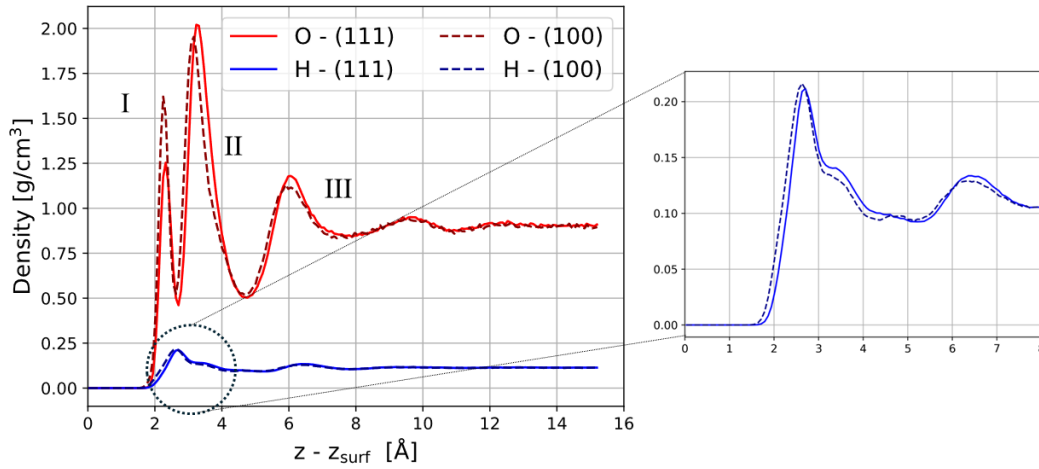


Fig. 5.2 Comparison of the water density profile at the PZC as a function of the distance from the Cu surface (placed at z_{surf}), decomposed into oxygen and hydrogen contributions. Solid and dashed lines refer to the Cu (111) and Cu (100) surfaces, respectively. The inset shows a magnification of the H density profile. Roman numerals are used to identify the three main interfacial regions.

surfaces are reported in Table 5.2. For both surfaces, the regions I and, above all, II are characterised by the highest increase in density. Nonetheless, these variations are still appreciable at 6 Å far from the surface, making it possible to distinguish a third interfacial region before reaching a bulk-like regime. We stress that standard AIMD studies do not report any comments on this third region, mainly due to the computational cost of considering water layers thicker than $6 \div 8$ Å. Nonetheless, it is important to recognise the presence of this transition region between the interfacial innermost water layers and more bulk-like regions far from the metal electrode. If the

Region	Surface	z_{peak} [Å]	z_{range} [Å]	Normalized O peak density [-]
I	(111)	2.32	$1.71 \div 2.71$	1.38
	(100)	2.25	$1.58 \div 2.62$	1.78
II	(111)	3.26	$2.71 \div 4.71$	2.22
	(100)	3.16	$2.62 \div 4.64$	2.15
III	(111)	6.03	$4.71 \div 7.79$	1.30
	(100)	5.98	$4.64 \div 7.55$	1.24

Table 5.2 Comparison of the structural properties of the different interfacial regions of water at PZC. Both the peak positions z_{peak} and the spatial ranges z_{range} of the regions are referred to the surface position z_{surf} . The O densities of the peaks are normalised with respect to the corresponding values of a reference bulk system.

O density in regions II and III is only slightly higher for the (111) surface with respect to the (100), the latter is characterised by a more noticeable peak I: the ratio between peak I and peak II heights turns from 0.62 in the (111) to 0.83 on the (100) surface. This is an expected feature for a surface with a lower coordination number, such as the (100), with respect to the more closely packed (111), making the surface/H₂O interaction more pronounced. The effects of a more open surface structure are also visible in the peak positions for the O as well as the H contributions: the first peak, and consequently the other ones, are slightly shifted towards the surface for the (100). In the subsequent sections we will more deeply investigate the interactions of H₂O molecules with surface atoms.

However, before proceeding to further analysis, it is essential to evaluate the reliability of the DFT ground truth obtained with the SCAN exchange–correlation functional in capturing this modulation of the water density induced by the interaction with the surface. Owing to the high computational cost associated with the SCAN functional, there are, to the best of our knowledge, no previous AIMD investigations of metal/explicit-water interfaces that compare the picture predicted by SCAN with that obtained from more extensively benchmarked functionals, such as rPBE-D3. Within this context, our study also provides an opportunity to compare different exchange–correlation functionals for metal/water interfaces in a consistent manner, decoupled from their different computational costs. As discussed in more detail in Appendix C.2, it seems that the rPBE-D3 functional tends to overestimate the metal/water interaction with respect to SCAN as a consequence of a stronger Cu $\cdots$ H interaction due to the D3 dispersion contributions. To eventually rule out any pitfalls of the SCAN functional in describing the van der Waals contributions in the metal/water interactions, we develop an ML model trained at the SCAN level with the addition of the non-local correlation functional rVV10. As shown in Appendix C.2, it turns out that the non-local rVV10 contributions have minimal effects on the interfacial structure of water, and the intermediate-range interactions, already captured by SCAN, are sufficient to reasonably describe the Cu/H₂O interface.

5.2.2 Water molecule orientation

After identifying the different interfacial regions in the solvent, here we better analyse the properties of water molecules, characterising their molecular orientation with respect to the surface plane (Figure 5.3). Specifically, the water dipole orientation

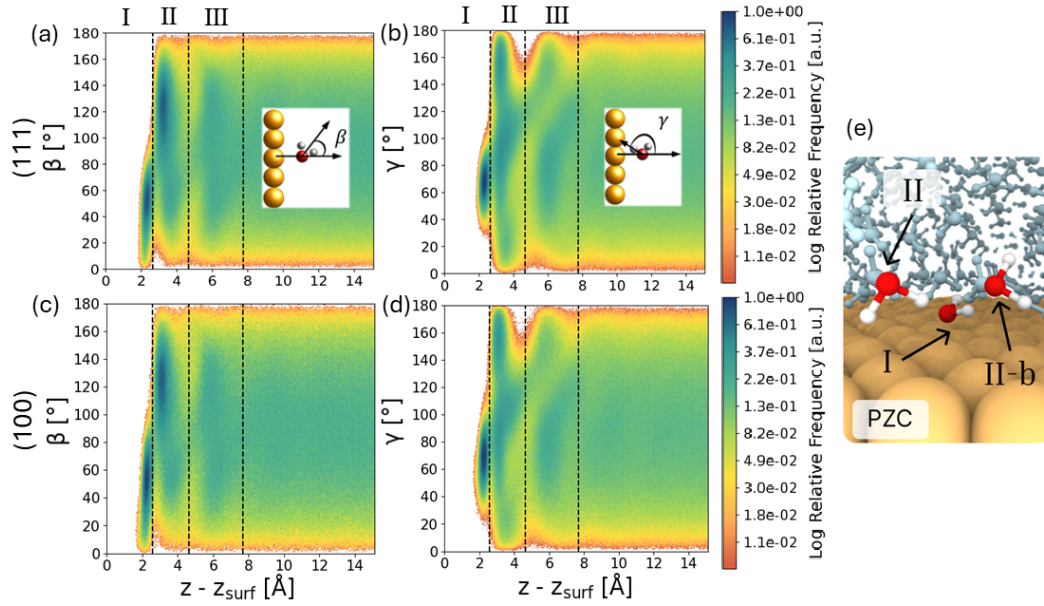


Fig. 5.3 Orientation of water molecules at the PZC for Cu (111) (top row) and Cu (100) (bottom row): colour maps for the distributions of the β dipole (a-c) and γ O-H-bond (b-d) angles; the insets show the geometrical definitions of the angles. The vertical dashed lines identify the limits of the different regions. (e) Configuration of molecules in region I, region II and between region II and III ("region II-b").

is studied considering the β angle. This is defined as the angle between the dipole vector of the molecule and the vector perpendicular to the surface (see inset of Figure 5.3a). While the O-H-bond direction is identified through the γ angle between the vector of the O-H covalent bond and the vector perpendicular to the surface (see inset of Figure 5.3b). In general, the angle distributions for the (111) and the (100) surfaces appear very similar with the same key features, proving that the different surface terminations can alter the number of H₂O molecules present in each region, but they do not evidently alter their orientation. Considering region I, the β distribution is peaked around $40^\circ \div 60^\circ$: the resulting molecular orientation has oxygen atoms exposed towards the surface and hydrogen atoms pointing away towards other water molecules (see also the orientation visible in the MD frame reported in Figure 5.3e). This adsorption configuration is commonly associated with a chemical interaction between H₂O and Cu, in which O shares its lone-pair electrons with the metal (cf. also Section 5.3.5 for further analysis on the strength of this bond vs other interactions present in the system). Although each molecule allows for two possible γ angles to be defined, the γ distribution in region I remains unimodal. This means that both H atoms behave in a similar manner, and they are

both at the same height with respect to the surface. These H atoms are associated with the unique H density peak in the separation area between regions I and II (see inset in Figure 5.2). The picture of H₂O orientation is completely different in region II. Here, the β distribution reaches its maximum value in the range $120^\circ \div 140^\circ$, meaning that the water dipole is mainly pointing towards the surface. The analysis of the γ distribution provides additional details. This distribution has the first lobe at about $140^\circ \div 170^\circ$, where the O-H-bond points towards the surface and H atoms weakly interact with the Cu atom through van der Waals forces. At the same time, there is another lobe in the γ distribution at about $80^\circ \div 110^\circ$. This is associated with the second O-H-bond that is almost parallel to the surface, and its H atom is interacting with other water molecules within region II. Despite "the water dipole pointing towards the surface" being the predominant picture in region II, a detailed analysis reveals a second weaker signal in the β and γ distribution. Looking at Figure 5.3a-c, it is possible to appreciate that there are few molecules with a β angle at about $40^\circ \div 70^\circ$. This occurs for a sub-region within region II, shifted towards region III (here named region II-b). This contribution is associated with water molecules with dipoles pointing away from the surface, towards region III (see MD frame in Figure 5.3e). This is also visible in the γ distribution with a contribution at around $90^\circ \div 110^\circ$ (O-H-bond parallel to the surface and H interacting with other molecules of region II) and at around $10^\circ \div 30^\circ$ (O-H-bond pointing towards region III and H interacting with molecules of region III). Notice that this detail is usually not accessible in standard AIMD due to the high signal-to-noise ratio of the statistics but provides interesting information: molecules passing from region II to III (and vice versa) have to flip the dipole from a β angle of $120^\circ \div 140^\circ$ to $40^\circ \div 70^\circ$. Finally, in region III, the β and γ distributions are more blurred with respect to previous regions, but they resemble the orientations of region II. We will show that this picture will be altered by the applied potential.

5.2.3 Hydrogen bond analysis

Information about how water molecules reciprocally interact can also be extracted from the analysis of the average number of hydrogen bonds per molecule (Figure 5.4). Following ref. [157], we define the presence of a hydrogen bond relying on geometrical criteria: the O donor-O acceptor distance should be smaller than 3.5 Å, the O donor-H atom distance should be smaller than 1.24 Å and the angle between

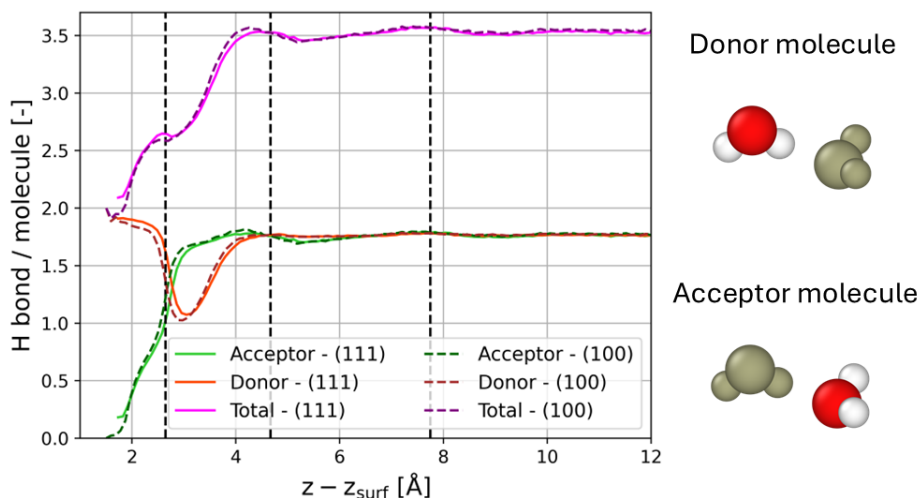


Fig. 5.4 Average number of H-bonds per molecule at the PZC, distinguishing the cases when the O atoms are donors or acceptors (see lateral insets for a graphical representation of donor and acceptor molecules). The distance of the molecules from the surface is evaluated considering the O atom position. The vertical dashed lines identify the limits of the different regions.

the O-O axis and the O-H-bond axis should be smaller than 30° . This definition of H-bond results in an average value of total H-bonds per molecule in water bulk equal to ≈ 3.54 . In general, the number of H-bonds per molecule is not influenced by the surface termination since the profiles are quite similar between the (111) and (100) cases. For the latter, the profiles are simply slightly shifted towards the surface in agreement with the shift in the density profile observed in Figure 5.2. For both surfaces, in region I, the number of acceptor oxygens (i.e., O atoms that employ their lone-pair electrons to establish H-bonds with other water molecules) is markedly smaller than 1. On the other hand, the number of donor oxygens (i.e., oxygens that have covalent bonds with H atoms, and the latter are involved in hydrogen bonds with other molecules) is close to two. This is in agreement with the picture previously described: in region I the lone-pair electrons are mainly involved in the interaction with Cu atoms and are not involved in H-bonds (i.e., low acceptor oxygen number), while both H atoms per molecule are interacting with O atoms of region II (i.e., high donor oxygen number). Notice that the number of donor oxygen atoms in region I is higher than the corresponding value in the water bulk far from the interface. This stands for a higher degree of order in the water structure induced by the surface. On the contrary, the number of donor O atoms drops to a value close to 1, and the

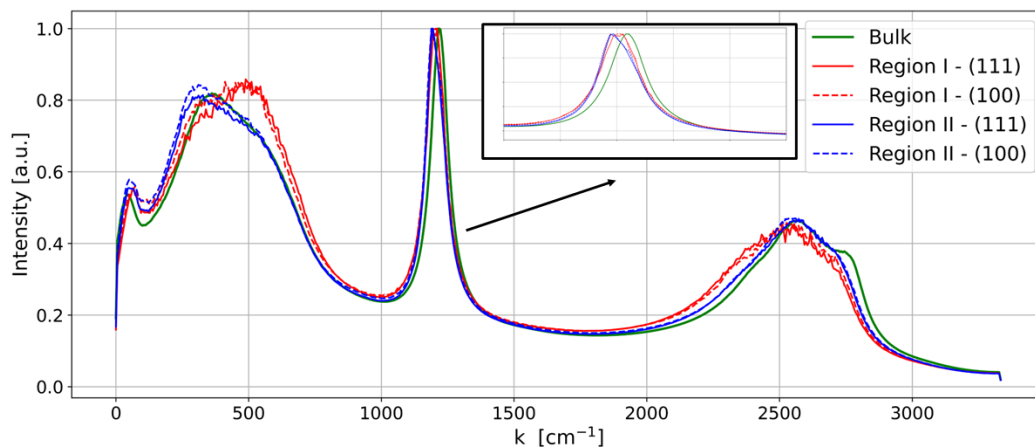


Fig. 5.5 Vibrational Density of States (VDOS) for hydrogen atoms in region I, region II of the Cu/water interface at PZC and in the water bulk. The spectra are obtained assuming the deuterium mass for the H species.

acceptor one increases in region II. This confirms that, for each molecule, one H atom is involved in H-bonds with other molecules within region II, while the other H is interacting with the metal atoms and not with the O atoms of region I.

5.2.4 Water vibrational density of states

A dynamical property that can provide further insights into the structure and strength of the hydrogen bond network is the vibrational density of states (VDOS). Here it is computed as the real part of the Fourier transform of the velocity autocorrelation function [158]. Specifically, the VDOS of the hydrogen atoms is considered since it allows to identify all the relevant vibrational modes of the water molecules [159]. Considering that the previous analysis shows that the most evident variations are associated with regions I and II, we evaluate the VDOS only for the water molecules belonging to these regions. The velocity autocorrelation function is evaluated by recording the velocity every 5 fs for a time window of 6 ps. The VDOS plots for hydrogen atoms in region I and region II are reported in Figure 5.5. For comparison, the VDOS for hydrogens in a water bulk is also reproduced. In all the cases, the spectra are dominated by four visible peaks. The peak at about $40\div50\text{ cm}^{-1}$ is not uniquely identified/classified [160, 161], but this low-frequency range of the spectrum is commonly associated with collective intermolecular H-bond modes; for example, Jin et al. report that it is associated with the hydrogen-bond bending,

perpendicular to the hydrogen bond [160]. The broad band between 150 and 700 cm^{-1} is attributed to the libration motion, while the sharp peak at about 1220 cm^{-1} refers to the H-O-H bending mode [156, 159]. Finally, the broad peak between 2300 and 2800 cm^{-1} can be decomposed into two contributions associated with O-H symmetric (lower k) and asymmetric (higher k) stretching modes [159, 162].

The presence of the Cu surface alters the strength of the covalent as well as hydrogen bonds of interfacial water, and these variations have an impact on the VDOS peak positions. We notice that the differences in the spectra for the two surface terminations are minimal, consistent with the analysis of the H-bond network previously reported. The peak at low frequencies ($\approx 40\div 50 \text{ cm}^{-1}$) associated with the collective H-bond modes slightly tends to blueshift in region I. These shifts are commonly observed in bulk water at decreasing temperatures and are correlated with a higher tetrahedral order of the structure [160, 161]. This element suggests that the interfacial water molecules in region I are characterised by a more "rigid" H-bond network with respect to molecules far from the interface. It is important to note that this constitutes an observation enabled by ML-FFs, extending beyond the conventional results typically reported in AIMD investigations, which generally do not examine the behaviour of this low-frequency peak. This primarily arises from the difficulty in resolving its intrinsically weak spectral intensity and from its proximity to the much more prominent librational band

The libration band in region II slightly tends to red-shift with respect to the water bulk profile, while it more evidently shifts at higher energy for region I. Notice that the libration motion is a measure of the stiffness of the rotational tendency of water molecules: a higher libration frequency is proof of a more restricted rotational motion [163]. Hence, the rotation of water molecules in region I is more hindered due to the $\text{Cu} \cdots \text{O}$ interaction and a higher number of H atoms involved in H-bonds (see Figure 5.4). On the contrary, a smaller number of total H-bonds in region II can explain the red-shift of the libration band in region II [159]. The red-shift of the H-O-H bending peak in both region I and region II is another feature of interfacial water [159, 164]. For water in region II, a lower total number of H-bonds (cf. Figure 5.4) results in a less constrained scissoring motion for water molecules, softening the H-O-H bending motion and red-shifting this peak [159]. The peak for region I also exhibits a red shift; in this situation, the shift is attributed not to a change in the number of H-bonds but instead to a perturbation of the electronic density caused by the stronger $\text{Cu} \cdots \text{O}$ interaction compared with the usual intermolecular interactions

between water molecules in the bulk phase [159, 165]. The different nature of the red-shift of the H-O-H bending band in peak I and peak II is also suggested by the average values of the intramolecular H-O-H angles of water molecules in the different regions. For chemisorbed molecules, this angle is higher than the typical value of water in a bulk region (105.87° in region I vs 105.28° far from the interface), suggesting a minor repulsion between bond and lone-pair electrons due to a major sharing of the latter with the surface. On the contrary, the average H-O-H angle in region II is smaller than the bulk case (105.09° in peak II). Even if these variations are small, they are consistently observed for both surfaces, identifying a common physical reason rather than statistical noise.

Regarding the shift of the O-H stretching peaks, this effect is commonly regarded as a proxy for the O-H covalent bond strength [162]. The stretching band in region II, but above all, in region I, red-shifts with respect to the water bulk case. This is commonly interpreted as a signature of a weakening of the covalent bonds as a consequence of stronger H-bonds [159, 162]. This is supported by the average O-H bond lengths that are progressively larger moving closer to the surface (0.978 , 0.979 and 0.984 Å in the bulk, in region II and in region I, respectively), suggesting a weaker covalent bond. The correlation between the increase of the strength of the H-bonds and a weakening of the covalent bonds is also supported by the analysis of the electronic density in the bond critical points, as shown in Section 5.3.5.

We emphasise that analysing the VDOS offers information that complements other analyses of the H-bond network and interaction strength. However, peak shifts in the VDOS can be quite subtle; in this regard, the enlarged statistical sampling afforded by ML-FFs is crucial for achieving a clearer and more reliable interpretation.

5.2.5 Ordering of the interfacial water structure

The occurrence of peaked regions in the solvent density with preferential orientations, together with a larger number of stronger hydrogen bonds and lengthened O-H covalent bonds, are all indicative of a structured system exhibiting ice-like characteristics. For this reason, we evaluate the ordering of the water structure, considering the in-plane O-O radial distribution function (RDF) within each region (Figure 5.6a-b). The RDF quantifies the spatial probability density of finding neighbouring atoms at a given distance from a reference atom. Pronounced peaks in the RDF indicate an enhanced likelihood of consistently locating neighbouring atoms at specific interatomic

distances, serving as a signature of increased structural ordering within the system. In the present analysis, the RDF is computed by considering exclusively in-plane interatomic distances. In addition to the O-O RDFs, the corresponding in-plane Cu-Cu distribution for the outermost layer of metal surface atoms is reported for references. Notice that the Cu-Cu distribution is reported as scaled down by a factor of 3, meaning that the liquid structure of water is always markedly less ordered than a solid phase. For both surfaces there is a clear correlation between the peak position

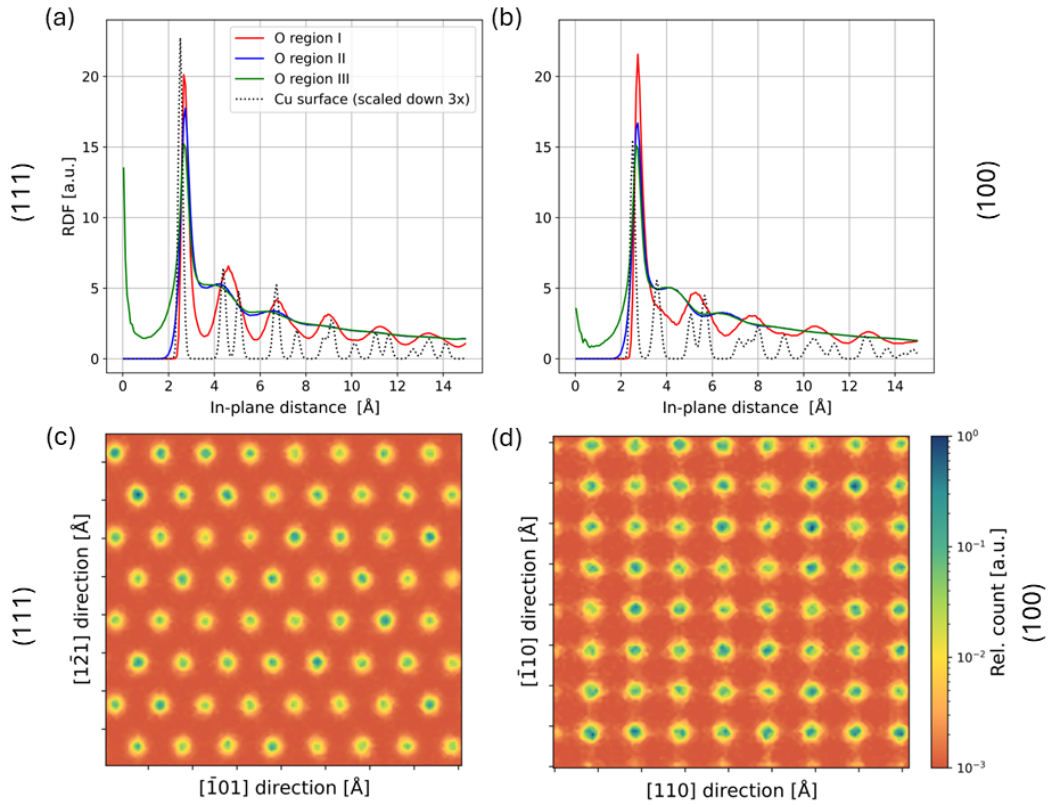


Fig. 5.6 In-plane ordering of water at the PZC for Cu (111) and Cu (100): (a-b) In-plane radial distribution function (RDF) for the O-O couple for the different interfacial regions. The corresponding Cu-Cu distribution for surface atoms is scaled down by a factor of 3. (c-d) Time-averaged in-plane spatial distribution for oxygen in region I. These maps provide information on the preferential sites of water and do not show the instantaneous adsorption pattern.

of the O-O RDF in region I and the Cu-Cu one, revealing an influence of the ordering of the surface on the spatial distribution of O atoms. This regular peak sequence in O-RDF is preserved also at long distances. Since at low coverage the preferential adsorption site is in the top one (cf. Appendix C.2), this correlation between O and

Cu periodicity suggests that water molecules are preferentially chemisorbed on the top Cu site also in our dynamic picture with explicit solvent. We will prove more rigorously this thesis in the following. Despite the regular profile, the O-O RDF in region I is relevantly smoother than the Cu one, proving that the water structure, even if more ordered than bulk water, is different from a highly ordered ice-like phase. Furthermore, the widely accepted theory of the "ice-like bilayer structure" [38] for interfacial water is not supported by the RDF analysis in region II. The main peak of the RDF in region II still inherits a correlation with the first RDF peak of region I, but it is significantly smoother. This indicates that the second layer of water molecules is less ordered than the first one, and they cannot be depicted as two layers with similar properties within an ice-like bilayer. Additionally, the RDF profile in region II is extremely smooth and almost coincides with the RDF of region III. Observe that, aside from the first peak, there is no correlation between the profile in regions II and III and the underlying periodicity of the surface: the bump in the RDF in regions II and III at approximately 4 Å and the subsequent one at 6.5 Å are the same for both surfaces, independent of the peak location in the Cu RDF.

The same periodicity of the peaks in the O RDF in region I and in the Cu RDF suggests a preferential adsorption site for the chemisorbed molecules of region I. For this reason, we also analyse the in-plane spatial distribution of O atoms of region I over the surface sites (Figures 5.6c-d). Notice that these maps report the time-averaged in-plane spatial distributions, providing information on the preferential sites of water with respect to the underlying pattern of Cu surface atoms, and do not show the instantaneous adsorption pattern. In region I, it is evident that O atoms tend to interact mainly on the top site of Cu, following a hexagonal pattern for the (111) and a square one for the (100). This is in agreement with experimental results that identify the top atom as the preferential interaction site for water molecules on metal FCC surfaces [166]. Considering the top Cu atom as the unique adsorption site, it is possible to evaluate the number of water molecules that are chemisorbed at the same time on the surface. We define the coverage θ of the adsorbed molecules in region I as $\theta = n_{H_2O}^I / n_{Cu_{top}}$, where $n_{H_2O}^I$ is the average number of water molecules in region I and $n_{Cu_{top}}$ is the number of Cu top sites. The resulting coverage of water molecules is 12.11 % and 16.13 % for the (111) and (100) surfaces, respectively. This means that Cu top atoms are often not covered by a water molecule (differently from the picture of an "ice-like bilayer structure"), but during the MD trajectory, each top atom can become a site in which a molecule is adsorbed.

5.2.6 Evaluation of the PZC value

Considering that the value PZC of the potential of zero charge is a unique characteristic that identifies an electrochemical interface and it is the result of different factors, including the nature of the surface and the interaction between the surface and the solvent molecules, we conclude this first section evaluating the value of the PZC.

Since the PZC value in simulations at constant charge is strongly influenced by the specific atomistic configuration [3], here we employ the final ML-FF at the PZC to generate a set of uncorrelated structures for both the surfaces in order to have a representative ensemble over the whole interface behaviour. These sampled configurations are then evaluated at the DFT level via the Double Reference Method to determine the value of the electrode potential in the absence of extra charge (i.e., the PZC). Following this approach, we estimate the average PZC value to be about $-0.1 \div -0.2$ V and $-0.5 \div -0.6$ V vs SHE, for the (111) and (100) surfaces, respectively. Notice that this is an average value since a constant potential is not forced, but the potential freely fluctuates to allow a zero surface charge. Although the PZC value varies across the different configurations, our average estimations are in very good agreement with experimental observations. Several studies, based on distinct experimental approaches, consistently indicate that the PZC is more positive for the (111) surface than for the (100) surface, and they report a difference of approximately 0.34 V between the two values [167, 168], in line with our predictions. Notably relevant is also the comparison with the results presented in Ref. [33]. In that work, the authors determine the potential of zero free charge (*pzfc*). This quantity generally differs from the potential of zero charge, also referred to as the potential of zero total charge (*pztc*), because the former accounts only for the surface charge, i.e., the actual electronic charge on the electrode, whereas the latter also includes the contribution from adsorbed species. In our situation, where no surface adsorbates are present, the total and free charges are identical; thus, the *pzfc* value can also be used to evaluate the PZC. By extrapolating the data reported in Ref. [33] from alkaline conditions to pH 7, we obtain a PZC value for Cu (111) of approximately -0.202 V vs SHE, which again agrees with our findings.

Aside from the experimental comparison, from an atomistic point of view, the more negative value of PZC for the (100) surface can be rationalised in terms of the influence of the dipole of water in proximity to the surface. Ref. [169] identifies the primary role of water molecules of region I in lowering the work function of the

interface, making the value of PZC more negative. Considering the higher density in region I for the (100) surface, the higher number of chemisorbed molecules has a larger role in lowering the PZC in the (100) vs the (111) system. To verify the effects of the different water structures on the electrode potential, namely, the work function, we evaluate the average surface charge density on the two surfaces at the PZC. For this reason, the charge on the electrode is computed at the *ab initio* level following a specific charge partition scheme. Notice that the accumulation of surface charge on the metal electrode is also due to the effect of the water molecule dipoles. For this reason we need to partition the charge density and assign the charge to the different atoms following a scheme that is able to properly reproduce the effects of water dipoles and be consistent with the electrostatic potential. Among the different charge partition schemes, we assign the charge to the different atoms following the DDEC6 (density-derived electrostatic and chemical) scheme proposed in Ref. [170, 171]. Integrating the charge on the Cu atoms of the two topmost surface atomic layers of the slab and averaging over the different configurations generated by the ML-FFs, we get the average surface charge densities reported in Table 5.3. We observe that the (100) surface can host a more negative surface charge compared to the (111) surface. An increased accumulation of electrons at the surface raises the Fermi level, thereby decreasing the work function and shifting the PZC toward more negative values. This behaviour confirms that the different screening characteristics of the interfacial water structure modulate the surface charge, lowering the PZC of the (100) facet relative to the (111) one. We also attempted to quantify the surface charge density using the commonly employed Bader charge partitioning schemes [172, 173]. However, these approaches proved inadequate for reproducing the expected correlation between surface charge and PZC, as they incorrectly predict a more positive surface charge for the (100) surface.

Even if the surface charge is evaluated at the *ab initio* level and the ML-FFs are employed exclusively to generate uncorrelated structures, a potential future extension

Surface	(111)	(100)
DDEC6 partition	-2.21	-3.19
Bader partition	+1.04	+2.14

Table 5.3 Comparison of the average surface charge density [$\mu\text{C}/\text{cm}^2$] for the Cu (111) and Cu (100) surfaces at the PZC obtained with different charge partition schemes.

of TRECI could involve the development of an ML model capable of predicting atomic charges based on an *ab initio* reference dataset. Within this framework, the scalability and data efficiency of TRECI would also be advantageous for enabling the construction of a charge dataset computed with the more computationally demanding DDEC6 partitioning scheme, rather than relying on the cheaper Bader approach.

5.3 Copper/water properties at constant potential

The structural and dynamical properties of interfacial water for Cu (111) and Cu (100), which are studied at PZC in Section 5.2, are now analysed when the system is subjected to a more negative (reducing) potential U than the potential of zero charge. Even though the two surfaces exhibit different PZC values, with the (100) surface being more negative, we use the same range of applied potentials for both cases, bearing in mind that the PZC of the (100) surface lies closer to the investigated applied bias window $[-0.5 \div -2.0 \text{ V vs SHE}]$.

5.3.1 Water density profile

Firstly, the density profiles of interfacial water (Figure 5.7a-b), and their decomposition in oxygen (Figure 5.7c-d) and hydrogen (Figure 5.7e-f) contributions are analysed at various values of external electric potential. The same profiles computed at PZC are also reported for reference. The notation for labelling the interfacial regions (i.e., I, II, and III) used earlier is also applied in this section.

The first evidence that it is possible to notice are the variations of the profile for the low potential cases with respect to the PZC: while the PZC line is well distinguishable from the $U = -0.5 \text{ V}$ case for the Cu (111) surface, the two profiles almost overlap for the (100) case. These results further confirm the correctness of our estimation of the PZC: it is more negative for the (100) surface with respect to the (111) one, and it is very close to -0.5 V vs SHE . The other evident signature of the applied potential is the variation in the magnitude of the water peak density in regions I and II (Figure 5.7). The density of peak I drastically decreases at increasing negative potentials, and it almost disappears below a threshold potential, U_{th} : for the (111), U_{th} is between -1.0 and -1.25 V vs SHE , while it is in the range $-1.25 \div -1.5 \text{ V vs SHE}$ for the (100) case. Notice that common AIMD studies often

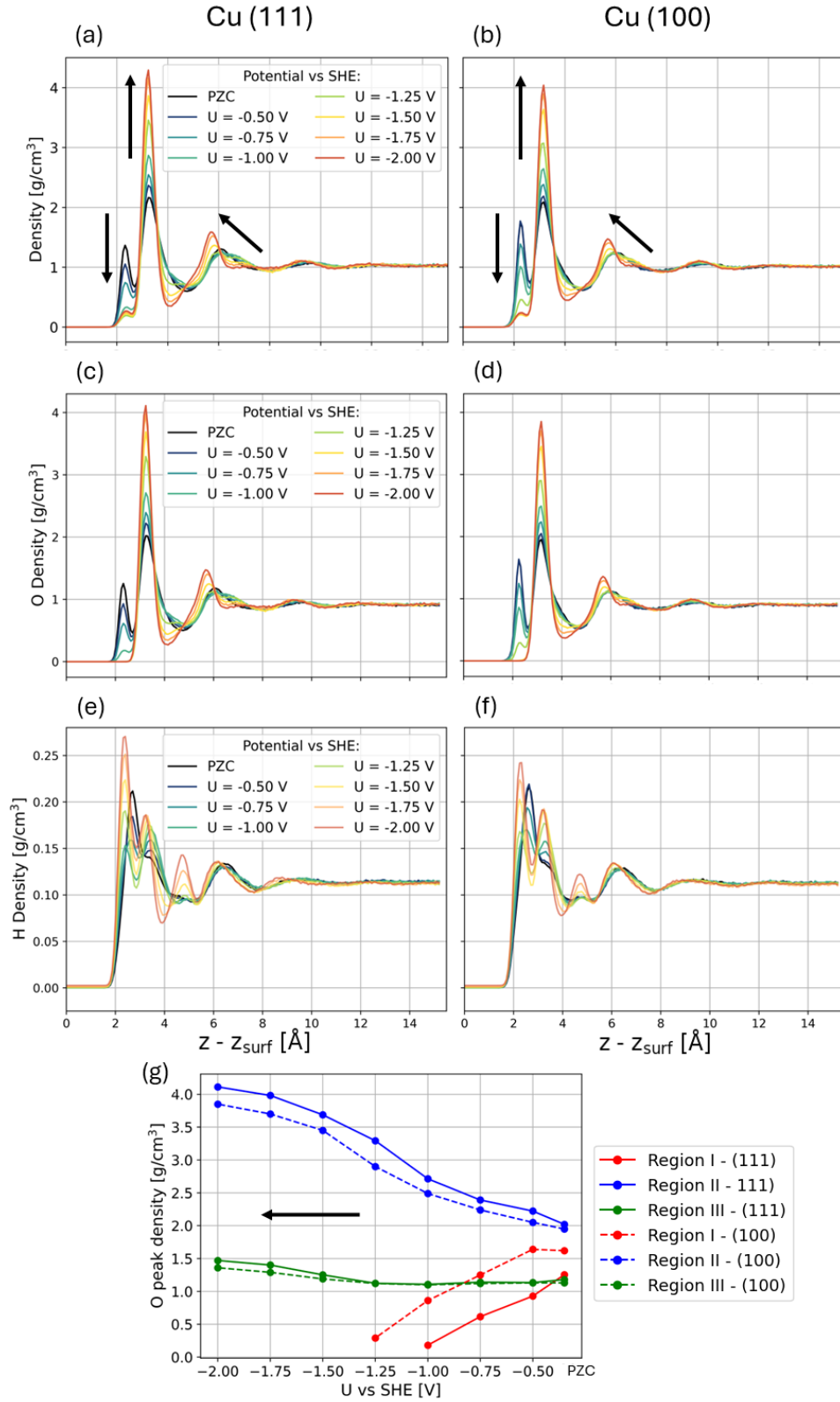


Fig. 5.7 Interfacial water density profile for Cu (111) (left) and (100) (right) surfaces at different values of potential as a function of the distance from the Cu surface (placed at z_{surf}): total density (a-b), oxygen (c-d) and hydrogen contributions (e-f); (g) summary of the O peak density in the interfacial regions vs the potential. Arrows emphasise trends moving towards more negative potentials

struggle with a precise identification of this threshold potential, mainly due to the complexity of considering, at the same time, more values of applied potential to get a fine discretisation of the potential window. On the contrary, the flexibility of the TRECI scheme to develop models in parallel for different applied bias cases allows tracking more closely the condition for which the first peak is suppressed. A more detailed analysis, focusing only on the oxygen role (Figures 5.7c-d and g), reveals that the O contribution in region I is completely suppressed for $U < U_{th}$: there are no more H₂O molecules in region I, and the residual contribution to the total density in region I is associated with hydrogens of region II molecules. This observation allows for a better interpretation of the different values of U_{th} in the two surfaces. Being the (100) system a less-densely packed surface with more water molecules interacting with the surface at the PZC, a larger external perturbation (i.e., a more negative potential value) is required to completely suppress the chemical interaction between the molecules of region I and the surface.

The application of the bias also causes the density in peak II to increase up to a value 4.5 times larger than the bulk water one in the considered range of potentials (Figure 5.2g). The increase of the magnitude of the second peak is similar for both surfaces, with the density for the (111) case always slightly larger than the corresponding (100) condition.

This phenomenon of modulations in the peak intensity under the application of an external potential was already predicted by atomistic simulations [67, 174] and also observed experimentally [40, 42, 44] for other metallic surfaces, and it can be rationalised considering the variations in the surface charge. At increasing negative potentials, the electrode surface gets more and more negatively charged. This alters the balance between the chemical (region I) and the electrostatic (region II) interactions among the Cu surface atoms and the H₂O molecules. The repulsion between the negative surface and the partially negative O atoms in region I becomes predominant over the dative Cu $\cdots$ O interaction, suppressing peak I. Whereas the negative surface charge enhances the electrostatic interaction with the H atoms of region II, resulting in a boost of peak II. The novelty of this study with respect to the reported results via AIMD approaches or experimental findings is the possibility of tracking, in an almost continuous way, this evolution of the density profile over a wide potential window and with a high signal-to-noise ratio not achievable with previous approaches.

The achievable level of detail allows one to capture a more complex scenario with

two newly identified elements for both surfaces: an evident transition in region I-II at U_{th} and a more complex behaviour of region III. As soon as peak I disappears, at $U < U_{th}$, there is an abrupt increase in the O density in region II, sharpening the peak. At the same time, this transition has also consequences in region III. Here, the O density is almost unperturbed at low voltages (Figure 5.7g), but peak III tends to increase at higher negative values. The bias influences not only the height of the peak but also its position: peak III tends to shift closer to the surface at more negative U ($\Delta z_{peak} \approx 0.2 \div 0.3 \text{ \AA}$ at $U = -2.00 \text{ V}$ for both surfaces). To the best of our knowledge, the effects of the applied potential in region III have never been discussed before. Indeed, this region and the complete transition to the water bulk are commonly not accessible via AIMD due to the computational cost of considering water layers thicker than $6 \div 8 \text{ \AA}$.

5.3.2 Water molecule orientation

To better catch the variations in the water structure discussed in the previous section, it is possible to analyse the dipole β and O-H-bond γ angle distributions (Figure 5.8) at varying potentials. The geometrical definitions for the β and γ angles are the same as reported in Figure 5.3. For the sake of simplicity, the angle distributions are only reported for $U = -0.75, -1.25$ and -1.75 V vs SHE (Figure 5.8) and only for the (111) surface. The behaviour for the (100) case is similar, aside from a shift of about 0.25 V in the trend due to a more negative U_{th} for this surface.

At low potentials ($U > U_{th}$), the dipole and the O-H-bond angle distributions in region I are similar to the PZC case (H_2O dipole pointing away from the surface and both H atoms almost at the same height, see Figure 5.8d). However, this contribution is progressively suppressed as the peak I is vanishing at more negative potentials. Whereas the contributions of molecules with $\beta > 90^\circ$ (dipole pointing towards the surface) in region II become more and more evident at negative potentials. Similarly to the PZC, there are still few molecules with dipoles pointing away from the surface in region II at low potentials (see configuration in Figure 5.8d, region II-bis), but these contributions disappear at more negative potentials. Indeed, for $U < U_{th}$, the β distribution is centred at $120^\circ \div 140^\circ$, while the γ distribution is peaked at $80^\circ \div 110^\circ$ (O-H-bond parallel to the surface and its H atom interacting with other water molecules within region II) and at $140^\circ \div 170^\circ$ (O-H-bond almost perpendicular to the surface). This orientation is weakly dependent on the applied potential value:

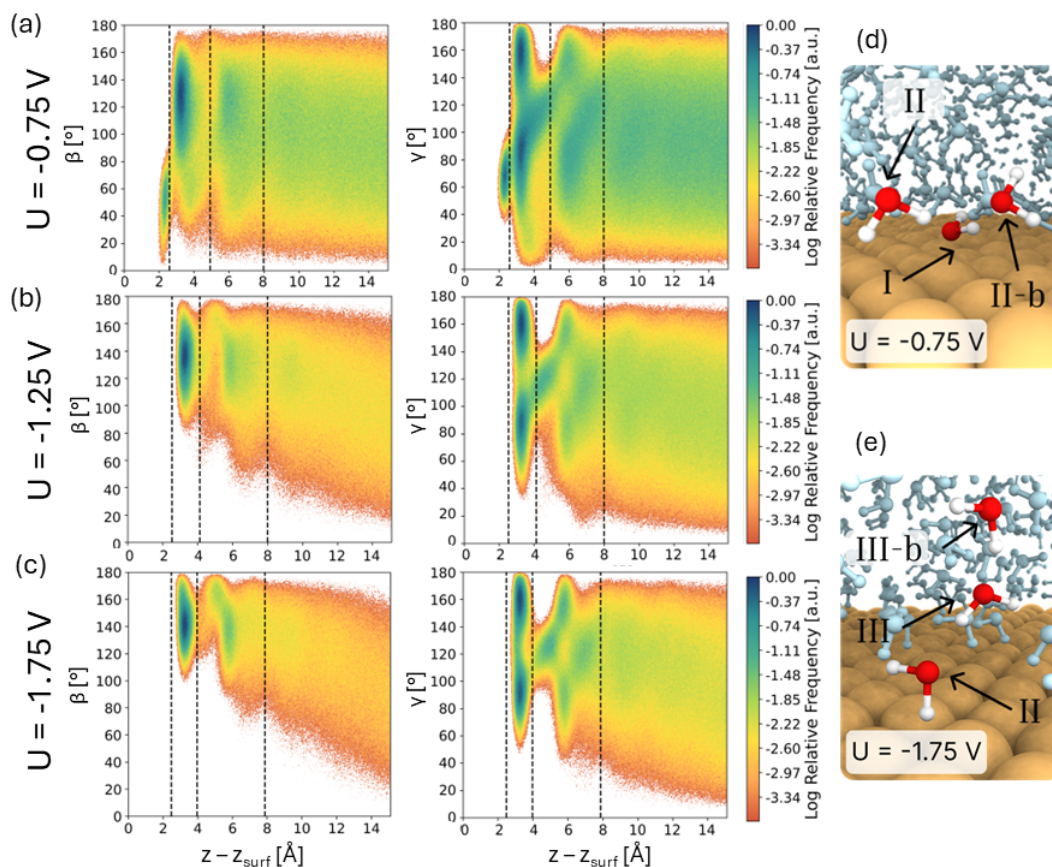


Fig. 5.8 Colour maps for the distributions of the β dipole and γ O-H-bond angle at -0.75 (a), -1.25 (b) and -1.75 (c) V vs SHE for the Cu (111) surface. The vertical dashed lines identify the limits of the different regions. Frames of the MD trajectory in which specific molecules are emphasised to show the water orientation: (d) configuration of molecules in region I, region II and between region II and III (“region II-b”) at low potentials; (e) configuration in region II, in region III close to region II (“region III”) and in the middle of region III (“Region III-b”) at high negative potentials. O and H atoms of emphasised molecules are depicted in red and white, respectively.

the peak of the β distribution only slightly shifts at higher angles at more negative potentials due to a stronger electrostatic interaction between hydrogens and the surface atoms. These variations are associated with a mechanism of dipole flipping, and they are compatible with the picture of the modifications in the O but also in the H density profile (Figures 5.7e-f). When there is the first O peak, H atoms of peak I and peak II are spatially overlapped in the separation area between regions I and II. These atoms contribute to the main peak in the H density at about 2.7 Å far from the surface. This peak progressively disappears at increasing negative potentials, and when the signal of chemisorbed molecules fades and the physisorbed molecule one

intensifies, the dual behaviour of H atoms of region II becomes evident. A distinct peak emerges at around 2.4 Å, indicating H within an O-H bond perpendicular to the surface, along with a broader peak at 3.2 to 3.5 Å representing H in an O-H bond parallel to the surface (see Figure 5.8e-region II to better visualise this orientation). Interestingly, we notice for the first time that the effects of high, negative potentials also have impacts on the orientation of water in region III. If there is no clearly predominant dipole direction at low potentials, it becomes more evident that there are at least two water orientations in region III at $U < U_{th}$. The dipole angle is about $160^\circ \div 180^\circ$, while a single spot for γ is about $110^\circ \div 130^\circ$ in region III in proximity to the edge with region II. These contributions are associated with water molecules with the dipole perpendicular and pointing towards the surface. Both hydrogen atoms are almost at the same height (single spot for γ) and are interacting with O of region II. This orientation can be visualised in the MD frame of Figure 5.8e, where this orientation is identified as “region III”. These effects are also visible in the H profile with a peak at about 4.75 Å far from the surface. On the contrary, the β and γ distributions in the middle of region III are more similar to the region II ones. An O-H bond is perpendicular to the surface (γ at $140^\circ \div 170^\circ$) and the other one is almost parallel to it (γ at $70^\circ \div 100^\circ$), as emphasised in Figure 5.8e, with the label “region III-b”. The resulting dipole is not exactly perpendicular to the surface (differently from region III close to region II) but β is peaked at $110^\circ \div 150^\circ$. We emphasise once more that these findings in region III cannot be obtained through conventional AIMD; however, they could play a crucial role in surface reactivity. At considerably negative potentials, molecules that interact on the surface (region II) are not directly in contact with an area resembling bulk water; rather, they can migrate in a zone that retains partial structural organisation (region III).

5.3.3 Hydrogen bond analysis

Similarly to the PZC case, the analysis of the H-bond number per molecule can provide additional insights on the structure of interfacial water and on the rigidity of the H-bond network at applied potential. Relying on the same geometrical definition of H-bond of Figure 5.4, Figures 5.9a,c,e show the average number of hydrogen bonds as a function of the distance from the surface and at different applied potentials. The picture at low potentials ($U > U_{th}$) is similar to the PZC condition since there is always the co-presence of chemisorbed and physisorbed molecules at the

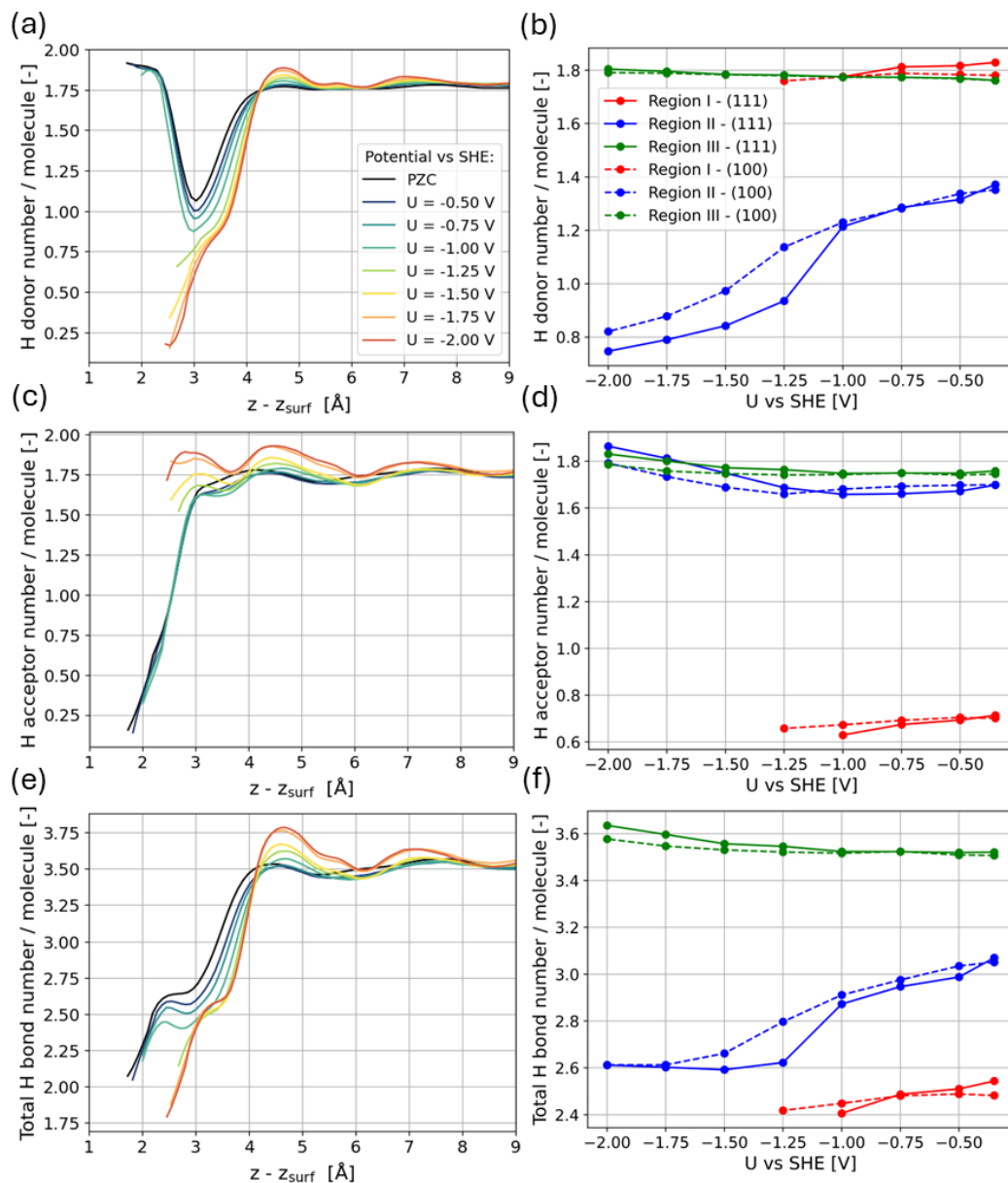


Fig. 5.9 Spatial profile of the average number of H-bonds per molecule for the (111) surface (left column) and the number of H-bonds averaged on block regions for the (111) and (100) surfaces (right column) at different applied potentials: (a-b) number of donor O atoms, (c-d) acceptor O atoms, and (e-f) total number of H-bonds. The distance of the molecules from the surface plane (z_{surf}) is evaluated considering the O atom position.

interface. Region I (O pointing towards the surface) is characterised by almost the maximum number of donor O atoms and no acceptor oxygens. On the contrary, the number of donor oxygens drops to 1, and the acceptor one increases in region II. At

more negative potentials ($U < U_{th}$) the trends in region II become even clearer. The average number of donor oxygens is progressively smaller than 1 (i.e., molecules in region II employ the H atoms to interact more with the surface rather than with other molecules). Whereas the acceptor number progressively increases up to reaching values larger than the water bulk condition (i.e., water molecules in region II receive many H-bond interactions from molecules of region III). Interestingly, the H-bond analysis also shows the effect of high negative potentials in the H-bond network in region III. Different from the low potential cases where region III has a network similar to the water bulk, both donor and acceptor numbers are higher than the bulk ones at $U < U_{th}$, proving a more structured condition also in region III.

To more clearly elucidate the differences in the H-bond networks at the two surfaces, we additionally report the number of H-bonds per water molecule, averaged separately over regions I, II, and III (Figures 5.9b,d,f). The overall trends are comparable for the (111) and (100) surfaces. In region I, the density of donor molecules is slightly higher than in bulk water, whereas the number of acceptor molecules remains low at small applied potentials. In region II, by contrast, the number of donor molecules generally decreases, while the number of acceptors increases at negative potentials. Nonetheless, owing to a more negative U_{th} , the (100) surface consistently exhibits a larger number of donors and a slightly smaller number of acceptors than the (111) surface at the same negative potential. Overall, within the potential window from -1.0 to -1.5 V vs SHE, i.e., between the two distinct U_{th} values of the two surfaces, the (100) interface has a higher total number of H-bonds. This finding is particularly significant in view of the key role played by the interfacial H-bond network in modulating surface reactivity, especially considering that the differences between the two systems emerge within a potential range commonly employed in electrocatalysis.

5.3.4 Water vibrational density of states

The analysis of the VDOS spectra of H atoms in water, alongside the previous section on the H-bond number, allows for a clearer identification of the features of the H-bond network at applied potential. For example, the relative strength of the H-bonds vs the O-H covalent bonds in water is correlated to O-H stretching vibration modes [159]. As emphasised with the black arrow in Figure 5.10b, the O-H stretching band in region II visibly red-shifts at negative potentials. This shift is the

typical signature of the weakening of the O-H covalent bond due to stronger H-bonds, and it is consistent with the bond critical point analysis (Section 5.3.5). In addition to the red-shift of this peak in region II, we also observe a flattening and broadening of the O-H stretching band. This is in agreement with experimental results: in strongly hydrogen-bonded networks, the conventional distinction between symmetric and asymmetric stretching modes is no longer applicable, and the stretching spectrum is more appropriately described in terms of two individual O-H oscillators [165, 175]. This might explain why the O-H stretching band at $U = -2.00$ V does not have a peaked profile but rather a quite flat apex resulting from the convolution of two O-H oscillator peaks with similar intensity.

If the stretching band reflects the strength of H-bonds through the relative motion of oxygen and hydrogen atoms along the covalent bond axis, the libration modes capture the restricted rotational motion of water molecules caused by H-bonding interactions [163]. As emphasised by the red arrows in Figure 5.10a-b, the libration band in regions I and II is altered by the applied bias. The shift at lower energies in region I is proof that molecules can more easily rotate in region I due to the weaker Cu-O interaction. Whereas, in region II, the high-energy shoulder of the libration band is more suppressed at increasing negative potentials. This red-shift can be interpreted as the signature of a less restricted rotational motion at increasing negative potentials due to a smaller number of H-bonds in region II (H atoms pointing towards the surface are not involved in H-bonds with other molecules). The red-shifts of the H-O-H peaks at negative potentials can also be rationalised in terms of a lower total number of H-bonds, as visible in Figure 5.9f.

These discussed trends are valid for both the (111) and (100) surfaces. A more detailed comparison (Figure 5.10c) can further highlight the difference of the H-bond network between the two cases. At $U = -1.25$ V, i.e., at a potential lying between the two distinct U_{th} values, the surface exhibits the greatest diversity in its behaviour. In region II, the libration band of the (111) is more red-shifted with respect to the (100) case, consistent with the lower number of H-bonds for the (111) observed in Figure 5.9f. At the same time, the O-H stretching band at -1.25 V as well as at -2.00 V is slightly more red-shifted for the (111) surface, suggesting a more evident weakening of the O-H covalent bond in favour of the H-bond strength for the closest-packed surface. Therefore, it can be concluded that the H-bond network exhibits distinct responses to the applied potential on the two surfaces, with the most

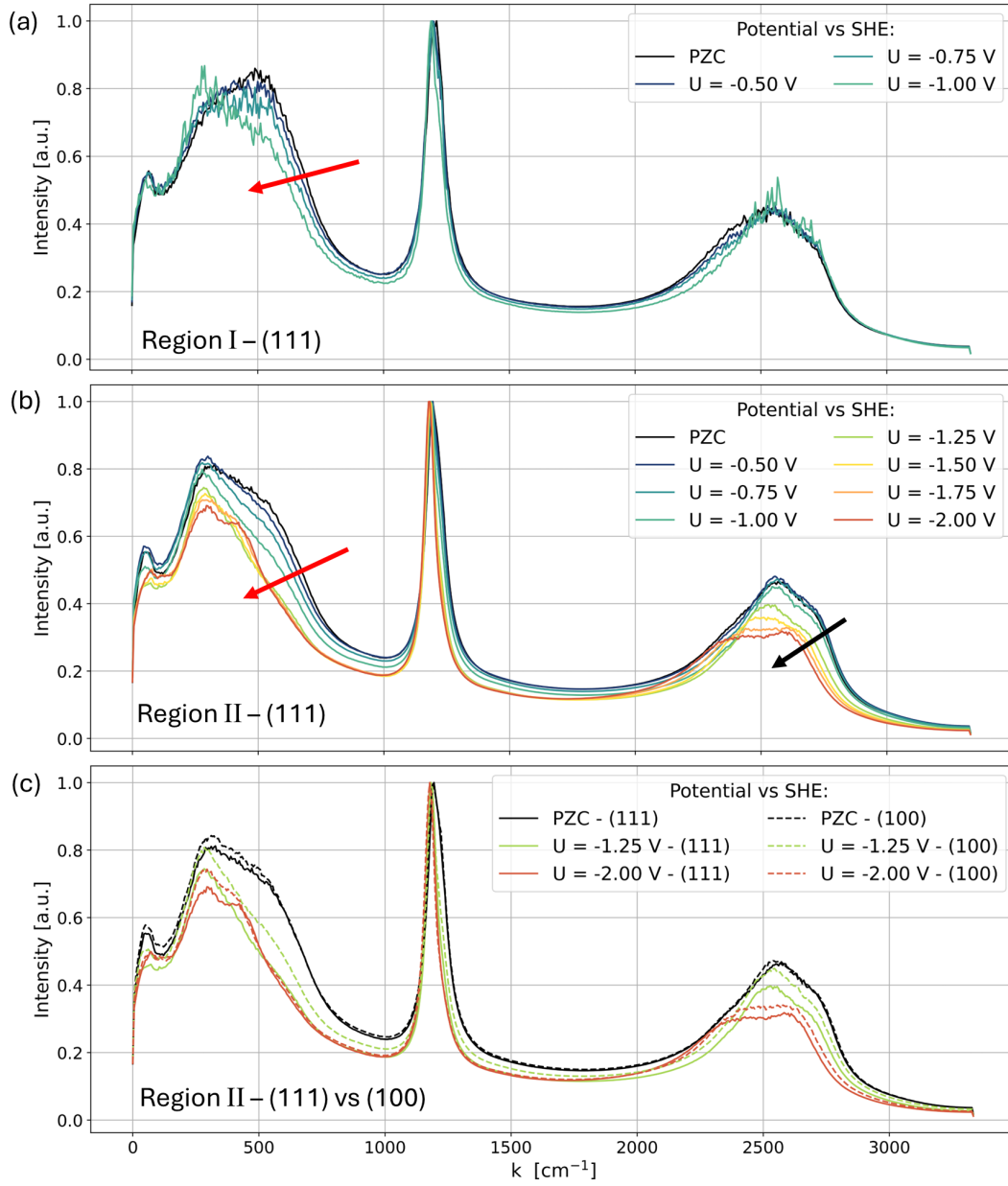


Fig. 5.10 VDOS spectra for hydrogen atoms in region I (a) and region II (b) of Cu (111)/H₂O interface at different applied potentials. (c) Comparison of the VDOS spectra for Cu (111) and Cu (100) in region II at selected potentials. The spectra are obtained assuming the deuterium mass for the H species.

pronounced differences occurring in the potential range from -1.00 V to -1.50 V, as previously discussed in Section 5.3.3.

5.3.5 Analysis of the bond strength via critical point analysis

The low computational cost of ML-potential-accelerated MD enables the generation of many uncorrelated structures that can be used to characterise, from the statistical point of view, not only the structural and dynamical properties previously discussed but also more subtle characteristics of the system, such as the strength of the metal/water and H-bond interactions. We perform this bond strength analysis in the framework of the “Quantum Theory of Atoms in Molecules” by Bader [172], considering the electron density of the bond critical points (BCPs). This quantity has been proved to be a good proxy for the strength of the bonds, above all for the hydrogen bond interactions [176, 177]. In this context, we selected a set of configurations generated by the ML potentials at PZC and at -2.00 V, where we performed a DFT-based analysis of the charge density to determine the BCPs using the *Critic2* program [178]. The results of the BCP analysis on the Cu (111) are reported in Figure 5.11 at the PZC (a) and at $U = -2.0$ V (b). The line of points with a density of about $0.3 \div 0.4$ a.u. is associated with the BCPs of the O-H covalent bonds. The underlying points with a density smaller than 0.1 a.u. are related to the intermolecular H-bonds and metal/water interactions. Focusing on the interactions in the interfacial region at PZC (Figure 5.11a), the BCPs with a density smaller than 0.1 a.u. and at a distance from the surface smaller than 2 Å are associated with interactions between the H₂O molecules and the metal surface, including both the Cu $\cdots$ O and the Cu $\cdots$ H interactions. The BCPs closer to the surface are associated with the Cu $\cdots$ O bonds, while the others are for the Cu $\cdots$ H cases. These two types of interactions can be distinguished by the magnitude of the density in the BCPs: the electron densities in the BCPs associated with the O atoms interacting with the surface (peak I) are approximately 0.5 ÷ 1 orders of magnitude bigger than the BCP density of the H atoms pointing towards the surface (region II). This proves the different strengths of the metal/water interaction of regions I and II, justifying the previous classification (chemisorbed molecules in region I vs physisorbed ones in region II). Nevertheless, the electron density at the Cu $\cdots$ O bond critical point remains lower than that at the O-H bond critical point and only marginally higher than that associated with the H-bonds. This means that the Cu $\cdots$ O interaction in region I has a strength that is intermediate between the intramolecular O-H covalent bond and the intermolecular H-bond.

Our BCP analysis also reveals how the Cu/H₂O interaction changes due to the ap-

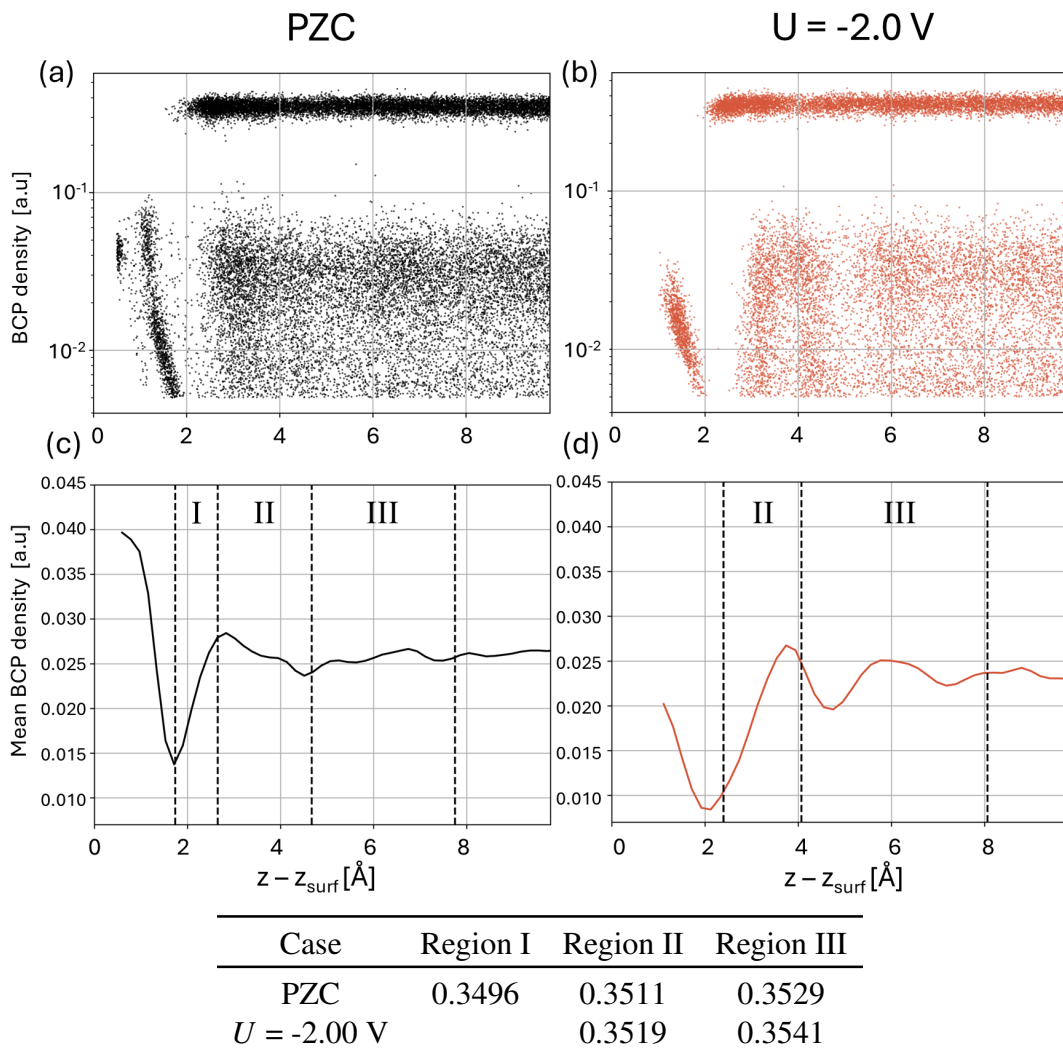


Fig. 5.11 Analysis of the electron density in the BCPs as a function of the distance from the Cu surface (placed at z_{surf}) at PZC (a) and at $U = -2.00$ V vs SHE (b). Spatial profile of the average density in the BCPs, considering only the intermolecular and the metal/water interactions at the PZC (c) and at $U = -2.00$ V vs SHE (d). The vertical dashed lines identify the limits of the different regions. Table: average density in the BCPs associated with the O-H covalent bonds in the different regions.

plied potential. At $U = -2.00$ V (Figure 5.11b), the density of the BCPs associated with the Cu/H₂O interaction markedly drops. A lower BCP density (i.e., a weaker Cu/H₂O interaction) at the interface is compatible with an exclusively electrostatic interaction between Cu and H, proving a more hydrophobic behaviour of the surface at negative potentials. Finally, the BCP analysis also provides interesting insight not only on the Cu/H₂O interaction, but also on the intramolecular and H-bonds

within the water region, evidencing the effects of the surface and of the applied potential. Indeed, the presence of the surface also alters the strength of the O-H covalent bonds (cf. Table 5.11): the average density in the BCPs of the covalent bonds tends to decrease close to the interface. This occurs at PZC and more evidently at $U = -2.00$ V. This implies that the O-H covalent bond is weakened at the interface, above all when a high negative potential is applied. These results are consistent with the analysis of the shifts in the VDOS peaks (red-shift of the stretching modes), as reported previously. At the same time, as also discussed in the VDOS sections, the weakening of the O-H covalent bonds is coupled with an increasing H-bond strength. The profile of the average BCP density (Figures 5.11c-d) also shows this: the density of the H-bond BCPs increases between regions I and II at the PZC and between regions II and III at $U = -2.00$ V, i.e., for the H-bonds between molecules directly interacting with the surface and the other ones that are a bit further away.

It is relevant to stress that all the conclusions about the BCP analysis are derived considering the wide statistics enabled by the uncorrelated configurations generated via ML-FFs. The reduced range of variations between the density of the interfacial regions and the bulk part, together with the large statistical variation due to thermal fluctuations, can easily hinder trends in a standard AIMD approach. To the best of our knowledge, our results are one of the first analyses of the BCPs in the MD framework for complex heterogeneous systems such as electrochemical interfaces.

5.3.6 Ordering of the interfacial water structure

Finally, similarly to Section 5.2.5, we evaluate the impact of the different surfaces and of the applied potential on the ordering of interfacial water, considering the in-plane radial distribution function of the oxygen atoms in the different regions (Figure 5.12). Region I remains the only region that exhibits pronounced, well-defined peaks in the RDF profiles, even at larger distances, with a clear correspondence to the sequence of peaks observed in the RDF of the surface atoms. It is worth noting that, for region I, the positions of the various RDF peaks, irrespective of the applied potential, change with the nature of the surface and follow the periodicity of the underlying surface lattice. This behaviour provides clear evidence that the structural ordering in region I is governed by the surface atomic pattern, since, in both cases, the chemisorbed molecules predominantly occupy the top sites of the lattice. Consequently, it is possible to quantify the coverage of water molecules associated with the top sites,

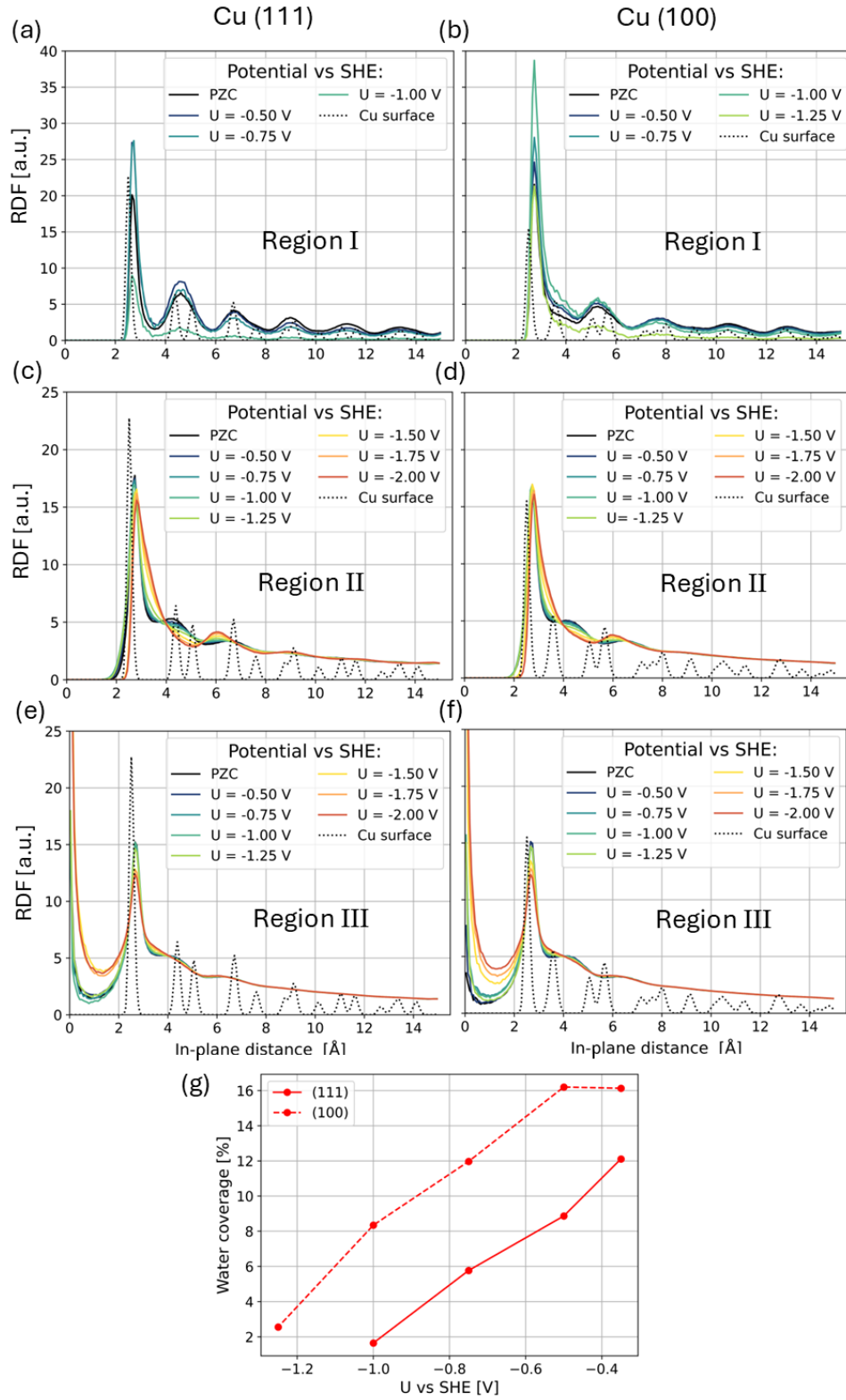


Fig. 5.12 In-plane radial distribution function (RDF) for oxygen species at applied potential for Cu (111) (left) and Cu (100) (right): (a-b) region I, (c-d) region II, (e-f) region III. The corresponding Cu-Cu distribution for surface atoms is scaled down by a factor of 3. (g) Coverage of H_2O molecules on the top Cu sites in region I.

as shown in Figure 5.12g. Interestingly, we observe that while surface coverage decreases (i.e., at negative potentials), the intensity of the first peak of the RDF in region I does not follow a monotonic trend. Indeed, it reaches a maximum at $-0.50 \div -0.75$ V and -1.00 V vs SHE for the (111) and (100) surfaces, respectively, before decreasing, getting closer to the corresponding U_{th} . For both surfaces, this occurs at a low coverage, around $9 \div 6$ %. Notice it is challenging to detect this clustering behaviour of water molecules at low coverage in systems of the size typically accessible to AIMD because such coverages usually involve fewer than 1–2 chemisorbed molecules. On the contrary, on the system size achievable with ML-FFs (see more details in Section 5.1), these low coverages still correspond to 4–6 adsorbed molecules. The ordering tendency of water molecules in region I progressively disappears in region II and, above all, in region III. In these cases, the profiles of the RDF between the two surfaces are very similar and smooth and do not correlate with the periodicity of the RDF peaks of the surface. In this sense the structure of water is no longer influenced by the surface pattern. Aside from the modulation of the intensity of the peak in the RDF, the main effect of the applied potential on these profiles is the position of the peaks in region II. At $U < U_{th}$, the first peak of the RDF tends to slightly shift at higher distances, while the second one moves from about $4.5 \text{ \AA}$ to $6 \text{ \AA}$. The increase in the in-plane O–O separation is consistent with the narrowing of the peak of the H_2O density in region II at negative potentials (Figure 5.7). As the vertical distribution of water molecules becomes more localised, reflected by the reduced width of the peak of the density in region II, the O atoms preferentially remain within a more well-defined plane. This enhanced confinement in the direction normal to the surface leads to a corresponding increase in their lateral spacing, which manifests as a larger in-plane separation between neighbouring water molecules.

5.4 Conclusion

In this chapter, we have fully leveraged the capabilities of the TRECI workflow to develop ML-FFs for electrochemical interfaces, performing ML-accelerated molecular dynamics simulations to characterise the properties of copper/water interfaces under electrochemical conditions. Our analysis focused on two application-relevant surfaces, the close-packed (111) and the less densely packed (100) surfaces, and

compared their interfacial behaviour at the potential of zero charge (PZC) and under applied negative potentials spanning -0.5 to -2.0 V vs SHE. We emphasise that the results reported in this chapter are obtained using ML-FFs that do not incorporate any explicit treatment of long-range interactions. This is not an oversight but a deliberate choice: we have verified that the systems considered here, which do not contain ions, are characterised by negligible interactions beyond the receptive field of our models. A combined interpretation of the structural and dynamical features of the interfaces reveals that the primary distinction between the two surfaces at the PZC originates from the balance between chemisorbed molecules in the innermost solvation layer (region I) and physisorbed molecules in region II. The (100) surface hosts a significantly larger number of chemisorbed water molecules in region I compared with (111), which results in a lower PZC and strongly influences the response of the system under constant potential conditions. For both surfaces, increasing negative bias enhances the interaction between the surface and water molecules of region II, while progressively suppressing the population of chemisorbed molecules in region I. This suppression occurs below a threshold potential U_{th} , which differs for the two facets: for (111), U_{th} lies between -1.0 and -1.25 V vs SHE, whereas for (100) it is shifted to more negative values, approximately -1.25 to -1.5 V vs SHE. This difference reflects the stronger chemical interaction of water with the less densely packed (100) surface. As a consequence, there exists a potential range ($\approx$ -1.0 to -1.5 V vs SHE) where the two surfaces exhibit distinct interfacial behaviours, including differences in preferential water dipole orientation, H-bond network structure, and interaction strength, features that may ultimately contribute to their different electrochemical reactivities. The reactivity of the surfaces for selected reaction steps will be discussed in the next chapter.

Beyond the specific insights into Cu/H₂O systems here analysed, this work highlights the significant advantages of ML-FF-accelerated simulations for studying electrochemical systems. The extended spatial and temporal scales accessible with ML-FFs enable revisiting extensively studied metal/water interfaces with unprecedented resolution, uncovering new phenomena, such as the emergence of structured water in region III at negative potentials, which deviates from bulk-like behaviour, and allowing property analyses (e.g., VDOS and bond-strength characterisation) with a statistical accuracy that remains beyond the reach of standard AIMD. Moreover, the scalability of the TRECI workflow makes it possible to generate ML-FFs across a finely spaced set of potentials, facilitating the identification of potential-dependent

transitions in interfacial structure and dynamics. In short, this chapter highlights the power of ML-FFs generated through the TRECI scheme, enabling high-level accuracy MD simulations able to provide a renewed and more comprehensive understanding of electrochemical interfaces.

Chapter 6

Electrochemical reaction steps elucidated via ML-accelerated MD

The importance of characterising the nanoscale processes occurring at electrified metal/water interfaces ultimately hinges on understanding the reactivity of these systems for common redox reactions. This chapter is developed upon the results reported in Chapter 5, where we studied the effects of applied reducing potentials on the properties of low-index copper surfaces in contact with a water solvent. Here we expand the picture, discussing the effects of the applied potential for selected electrochemical reactions. As discussed in Chapter 1, copper-based electrocatalysts are particularly appealing for their unique possibility of reducing CO_2 not only to simple C_1 added-value products but, above all, to more complex C_2 and C_{2+} species. While elucidating the whole reaction path for the CO_2RR in multi-carbon products is well beyond the aims of this chapter, here we focus on a reaction step that is commonly regarded as the main limiting process for the production of C_2 and C_{2+} species [19]: the CO dimerisation, i.e., the coupling of two $^*\text{CO}$ to form a C-C bond that is at the basis of any multi-carbon compounds (cf. Section 1.2.3). The analysis of the activation barriers for this step at different applied potentials is performed, comparing the Cu (111) and Cu (100) surfaces to understand the different reactivity of these interfaces (Section 6.2). In addition, we also consider the main side reaction that commonly limits the energy efficiency of the CO_2RR , i.e., the hydrogen evolution reaction. As discussed in Section 1.2.2, the HER reaction proceeds through a first proton transfer from the water solvent to the surface (Volmer step), followed by the coupling of this $^*\text{H}$ adsorbate with another proton from the solution (Heyrovsky

path) or one present on the surface as adsorbate (Tafel step). The elucidation of the full HER reaction mechanism is dedicated to future works, and here we focus on the main limiting step for HER. Its nature among the Volmer, Heyrovsky and Tafel processes depends on the catalyst material and the pH conditions in which the reaction occurs [18]. Commonly, the Volmer step is considered the limiting process for HER on Cu materials due to the weak binding energy of hydrogen on this electrocatalyst, making a proton transfer from the aqueous electrolyte to the surface less favourable [18, 179]. This is particularly true in neutral or alkaline conditions where the proton donor is a water molecule rather than a hydronium ion. For this reason, we will discuss the activation barriers for the Volmer step for the HER in Section 6.3. We consider this process in neutral/alkaline conditions since many studies on the CO₂RR avoid the use of acidic medium to minimise the competitive HER reaction.

The primary objective of this chapter is twofold: first, to demonstrate how the TRECI scheme presented in the previous chapters can be employed to construct ML-FFs capable of accurately describing reactive processes; and second, to show the enabling capability of high-accuracy, ML-accelerated MD simulations for the investigation of electrochemical reactions. The findings reported in this chapter represent preliminary results, and further investigation is ongoing to refine the ML framework and the model while deepening the analysis of the reaction mechanisms.

6.1 Computational details

Before entering into the detailed analysis of the results of the reactions, we report some key computational details for the developed ML-FF models and the performed MD simulations coupled with enhanced sampling techniques for rare events (cf. also Section 2.3).

6.1.1 Generation of the ML models

The ML-FFs models are developed following the TRECI strategy discussed in Chapter 4, with computational details similar to the information reported in Chapter 5. The main difference with respect to the previous approach employed to generate the ML-FFs without reactive events is the use of different backbones for the Franken

transfer learning strategy. As discussed in detail in Appendix B.3, the study of the reaction presents new challenges for transfer learning. The general-purpose descriptors based on MACE-MP0 (Franken-MP0), employed in the previous chapter, result in being more inaccurate when dealing with molecular processes, such as the reactions. At the same time, the biased MD simulations at constant applied potential more frequently push the ML-FFs to sample regions of the PES not explored by the PZC models, challenging the extrapolation capability of the domain-specific PZC descriptors and increasing the amount of data required to ensure the stability of Franken-PZC models. For this reason, to better bridge the performance gap between Franken-MP0 and Franken-PZC, we employ in this chapter new general-purpose descriptors built upon the Multi-Head model family (MACE-MH) that are more accurate than MACE-MP0 for the reactions while preserving the same stability. More considerations on models based upon the MACE-MH-0 (Franken-MH0), alongside a comparative table with the ML-FFs presented in other chapters, are reported in Appendix B.3. The results at constant applied bias reported in this chapter are obtained via Franken-MH0, with 8192 RFs, while the corresponding PZC model is trained from scratch with a MACE invariant architecture.

6.1.2 MD simulations with enhanced sampling techniques

The ML-FF models are employed in MD simulations with setups similar to the setting discussed in Chapter 5. Specifically, the study of the reactions via MD requires the use of enhanced sampling techniques to increase the probability of sampling configurations belonging to different reagent and product basins as well as the TSs. Here we follow the OPES strategy, as implemented in the Plumed package [180], and presented in Section 2.3.5. During the active learning phase, the OPES-Explore variant is employed (cf. Section 2.3.5) to quickly escape from the basins and more easily collect configurations of the different basins, sampling the PES in a more effective way to generate an exhaustive dataset. Multiple short MD runs with OPES-Flooding (cf. Section 2.3.5) are also performed during the active learning phase to selectively sample only configurations associated with TSs, following the example of ref. [129]. Whereas final results are obtained with the OPES-METAD flavour to address a more accurate estimation of the height of the activation barriers. Simple but well-established geometrical CVs based on the relative distance of atoms will be employed, as discussed in more detail in the following sections.

The results reported here are obtained on a symmetrical double-sided slab (cf. the geometry in Figure 5.1) with a lateral side of about 10 Å and a water region of about 10 Å per side for a total number of 50 ÷ 60 water molecules in a 20-Å-thick region. A single couple of adsorbed hydrogen - hydroxide ion in solution or a single couple of *CO are considered in each system, biasing the reactions on a single side of the slab. To better preserve the symmetry between the two surface sides, we add an *H or a *CO adsorbate on the opposite side of the slab. Further simulations will be performed in larger systems, similarly to the size employed in Chapter 5 to eventually rule out any finite-size-cell effects in the discussed results.

6.2 CO₂RR: the CO dimerisation step

As anticipated in the introduction of this chapter, here we only focus on the rate-limiting step for the conversion of the CO₂ in C₂ or C₂₊ products: the CO–CO dimerisation step. We implicitly suppose the availability of *CO adsorbates on the Cu surface; the full reaction path that results in the CO₂ adsorption and dissociation until the *CO formation will be addressed in future work.

6.2.1 CO diffusion

Before discussing the dimerisation process, we characterise the behaviour of *CO adsorbates on Cu (111) and Cu (100) in order to identify the preferential adsorption sites and their diffusion properties. For this reason, we performed MD simulations coupled with OPES-enhanced sampling techniques to bias the diffusion of a single *CO over the whole surface. The in-plane x-y coordinates of *CO are selected as biasing CVs. The resulting free energy surfaces of a single *CO diffusing on the two Cu surfaces at the PZC are reported in Figure 6.1. It is evident that, irrespective of the specific surface termination, the most probable adsorption site is the hollow one. This is in contrast with static DFT studies performed in a vacuum environment that identify surface-termination-dependency for the preferential adsorption site on these surfaces [181, 182]. This discrepancy highlights the importance of sampling the finite-temperature free-energy landscape in the presence of the solvent: dynamic fluctuations and adsorbate-solvent interactions can reshape the relative stability of adsorption sites compared to zero-temperature, vacuum calculations. While

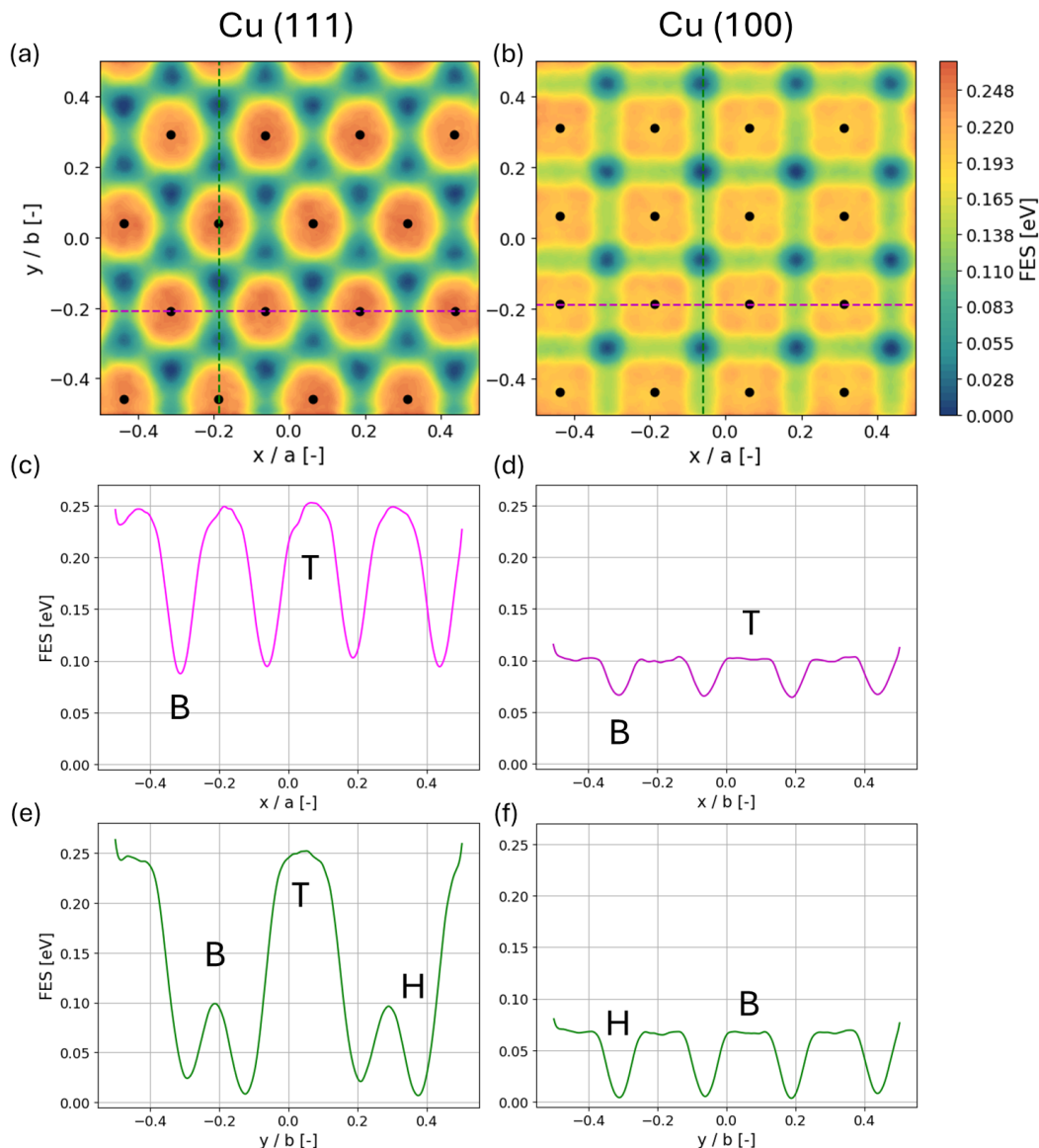


Fig. 6.1 FES of a single CO adsorbed on Cu (111) (left column) and Cu (100) (right column) surfaces at the PZC. In-plane $x - y$ positions are reported scaled with respect to the $a - b$ lattice vectors. Black spots represent the average positions of the Cu surface atoms. Diffusion barriers between main sites are reported along 1-D y -cut (c-d) and x -cut (e-f) plots and averaging over equivalent cuts. The 'H', 'B' and 'T' acronyms stand for 'hollow', 'bridge' and 'top' sites, respectively. The energy reference is set in the minimum energy basin.

these findings cannot be generalised to other terminations or electrocatalysts, they emphasise the need to benchmark against solvent effects and interfacial dynamics

to capture the realistic behaviour $^*\text{CO}$. Regarding the other lattice sites, the CO adsorptions on the top and the bridge ones in the (111) surface correspond to unstable states since they are an absolute maximum and a saddle point in the FES, respectively. The picture is different in the (100) case: even if the bridge and top sites are not absolute energy minima, the flatness or small concavity of the FES at these points suggests the possibility of observing the adsorption of CO also on these sites. Aside from these stability considerations, it is also interesting to observe the diffusion barriers for the $^*\text{CO}$. For the (111) surface (Figures 6.1c-e), the diffusion process between adjacent hollow sites through bridges is almost barrierless at room temperature (diffusion barrier < 0.1 eV), while the transition from a hollow to a top site requires 0.25 eV. Diffusion processes are even more barrier-less on the (100) surface since the energy differences among all surface sites are always smaller than 0.1 eV. In light of these results, it is possible to draw a picture in which the $^*\text{CO}$ are mainly in the hollow sites on the Cu (111) surface, with frequent transitions between adjacent sites mainly through bridges. Whereas, in the (100) case, $^*\text{CO}$ can diffuse across nearly all surface sites, with a non-negligible probability of temporary adsorption at top or bridge sites. For both surfaces, these low diffusion barriers facilitate the migration of $^*\text{CO}$ adsorbates that are independently generated at different locations on the electrode, increasing the likelihood that two $^*\text{CO}$ species encounter one another and thus promoting the conditions required for CO–CO dimerisation.

6.2.2 CO dimerisation: activation barriers and insights

After discussing the characteristics of $^*\text{CO}$ adsorbates, we exploit the ML-FFs developed via TRECI to describe the CO – CO dimerisation step on the Cu (111) as well as on the Cu (100) surface. In this case the process is simulated via two $^*\text{CO}$ placed on the same side of the Cu slab and biasing the dimerisation process via OPES-enhanced sampling. The C-C distance of the two adsorbates is employed as geometrical CV for the bias potential. To restrict the configurational space sampled during the biased simulation, an upper bound of 4 Å is imposed on this CV to prevent the two $^*\text{CO}$ species from excessively separating from one another. This distance is larger than the separation between two adjacent hollow sites on both surfaces, allowing a full reconstruction of the incoming motion of the adsorbates. The resulting activation barriers for the process on the two surfaces at different

applied potentials, alongside some representative snapshots of the dimer formation along the reaction path, are reported in Figure 6.2. The main difference in the dimerisation process between the surfaces, regardless of the applied potential value, is the range of the activation barrier for the formation of the dimer. In the analysed cases, these are always larger than 0.9 eV for the (111) surface, while they are lower than 0.8 eV for the (100) case. A preferential dimerisation on the Cu (100) surface relative to the Cu (111) case has been consistently reported in both experimental and computational studies [19, 183, 184]. Our analysis further clarifies this result, providing a clear picture of the condition in which the *OCCO TSs are formed. As visible in Figure 6.2e, the geometrical centre of the C-C dimer in the TSs can be in a hollow as well as in a bridge site on the Cu (111). This is also visible in the snapshot in Figure 6.2c, where it is evident that a hollow site cannot host both *CO species, and one is partially in a hollow site, while the other is on a bridge. The result of this triangular geometry of the hollow site is a distorted and highly unstable configuration in which two *CO are not face-to-face, preventing effective dimer formation. In general, the product of the dimerisation on the Cu (111) is also a high-energy configuration with respect to two independent *CO adsorbates, since the dimer is always partially hosted between hollow and bridge sites. On the contrary, the squared lattice geometry of the (100) case favours the formation of the *OCCO TSs. As visible in Figure 6.2f, the centre of the C-C couple during the TSs is always within a unique hollow site. The two *CO species passing from the bridge sites reach the same hollow site, where they face one another at the same height, resulting in an optimal C-C coupling condition. The final dimer is also fully hosted in a unique hollow site, stabilising the product species and increasing the lifetime of this state. Indeed, the reverse barriers of the dimerisation (i.e., the barriers that prevent the dimer from breaking and generating two independent *CO) are relevantly higher for the (100) surface, while they are almost absent for the (111) case, above all at low potentials. High reverse barriers are important since they increase the lifetime of the dimer, allowing it to undergo further reaction steps to produce C₂ and C₂₊ species. In this framework the low or absent reverse barriers for the Cu (111) further explain the scarcity of C₂ products that can be obtained on this surface.

Even if the activation barriers for the Cu (100) are lower than the corresponding one on the Cu (111), we observe that they are systematically higher than the common values reported in *ab initio* studies based on static DFT calculations [185, 186]. A systematic investigation comprising a detailed comparison of the influence of the

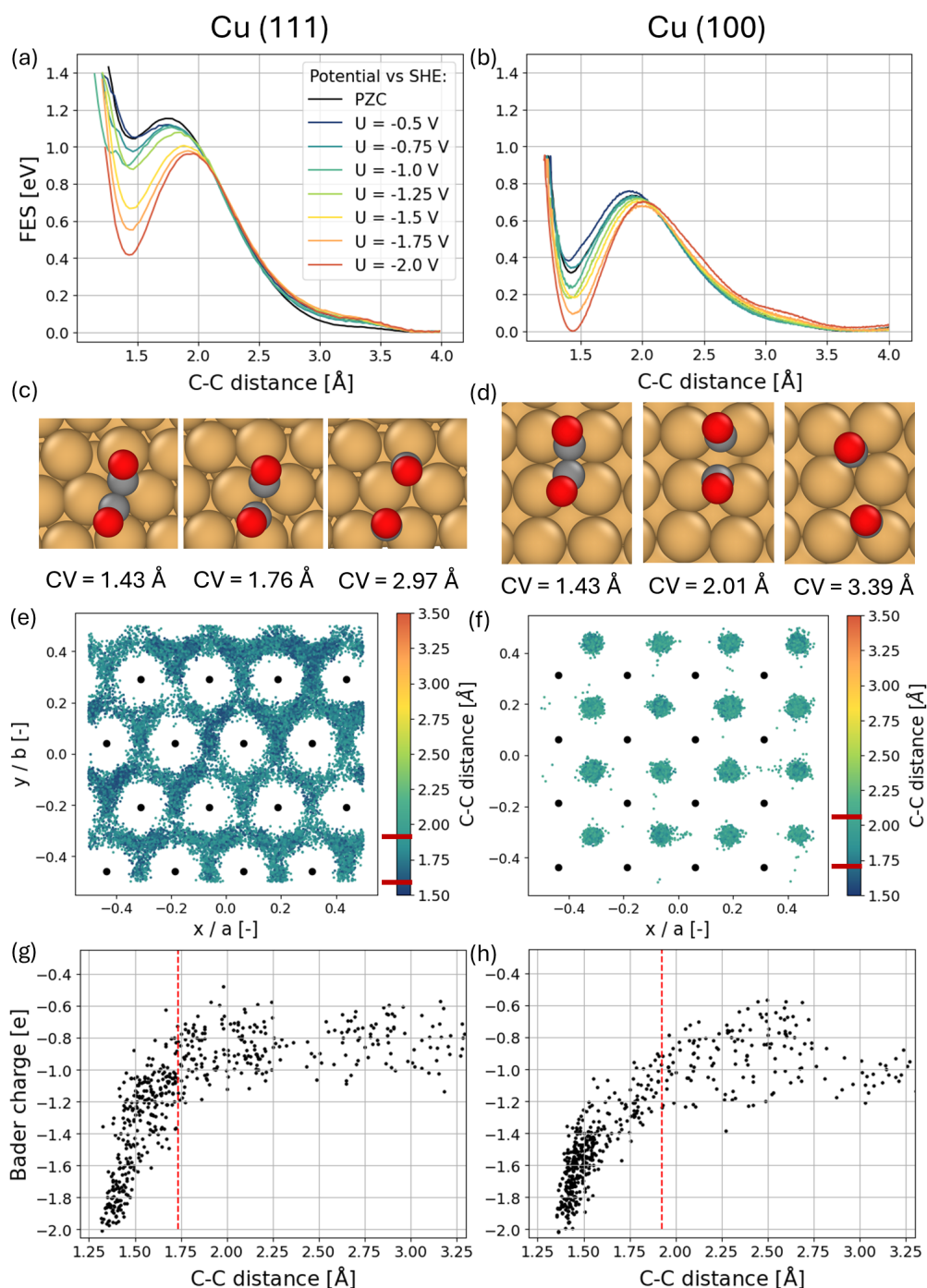


Fig. 6.2 CO – CO dimerisation step on Cu (111) (left column) and Cu (100) (right column) surfaces. (a-b) Activation barriers at different applied potentials vs the C-C distance CV. (c-d) Representative MD snapshots along the reaction path, depicting (from left to right) product, transition, and reactant states. Water molecules are omitted for visual clarity. (e-f) Mapping of the geometrical centre of the dimer on the surface lattice. Contributions of configurations with a C-C distance value in an interval (identified by the red ticks on the colour bars) around the TSs are only reported. The coordinates are scaled by the in-plane $a - b$ lattice vectors. Black spots are the averaged positions of the Cu surface atoms. (g-h) Sum of the Bader charge of both adsorbates vs the C-C distance. Red vertical lines represent the location of the TSs. The results of panels c-h are reported only for the PZC as representative of all the other cases.

various exchange–correlation functionals, solvent models, electrification schemes, and static vs dynamic methodologies employed across the different studies would be required to clearly identify the origin of the discrepancies between our results and those reported in the literature. Such an analysis, however, lies beyond the scope of the present thesis. At this level we also point out that static DFT approaches commonly evaluate the dimerisation barrier with respect to an initial state where the two $^*\text{CO}$ adsorbates are face-to-face on bridge sites. Even if we observe the $^*\text{CO}$ species reaching the hollow site for the dimer formation passing through the bridges (Figure 6.2d), we allow the $^*\text{CO}$ adsorbates to more freely move on the surface, identifying two $^*\text{CO}$ in adjacent hollow sites (or even at higher distances) as species of the reagent basin. This difference in the identification of the reagent states is a further element to be considered to explain the different heights of the barriers.

Focusing on the dependence of the activation barriers with respect to the applied potential, they are extremely high at low potentials ($U \geq -1.25$ V vs SHE) for the (111) surface. Only at highly negative potentials ($U < -1.5$ V), the barriers start to visibly decrease and the dimer is stabilised. Nonetheless, the activation energy for the process on the Cu (111) remains high ($\approx 0.9 \div 1.0$ eV). As discussed in the following Section 6.3.2, the Volmer barriers for the HER are largely decreased in the same potential window, enhancing the competitive HER reaction with respect to the CO dimerisation. The role of the applied potential is more convoluted for the (100) surface. The primary effect of applying negative biases is the stabilisation of the products, as evidenced by the reduction in the energy of the dimer basin, which becomes thermodynamically comparable to the energy of two independent $^*\text{CO}$ species. On the contrary, the applied potential does not relevantly reduce the activation barriers; they only tend to slightly decrease at negative bias without any abrupt variations. To better identify why the dimer basin is largely stabilised by the applied potential, while the TSs, and consequently the height of the barriers, are minimally influenced by the extra surface charge, we perform a Bader analysis evaluating the sum of the atomic charges of both $^*\text{CO}$ species along the reaction path (Figures 6.2g-h). For both surfaces, the atomic charges of the adsorbates in proximity of the TSs (indicated by red vertical lines) are comparable to those of two isolated $^*\text{CO}$ species. In contrast, the $^*\text{OCCO}$ adsorbate begins to accumulate additional negative charge in the product basins, as the C–C distance decreases below that at the TSs, until acquiring nearly one extra electron. This can explain the different responses of products and TSs to the applied potential: the accumulation of negative

surface charge induced by the potentials stabilises the dimer products since their formation requires the transfer of one electron from the surface to the adsorbates. On the contrary, the TSs do not acquire extra electrons, resulting in being minimally influenced by the different charge available on the surface.

To further understand the process of dimerisation, we performed a more detailed analysis of the properties of the adsorbates along the reaction path at the different applied potentials (Figure 6.3). As the two *CO get closer, the C-O bond is no longer perpendicular to the surface, but it starts to be laterally tilted, as reported in Figure 6.3a-b. When the C-C distance decreases below about 2.5 Å, the C-O bond angle, measured with respect to the vector perpendicular to the surface, tends to rapidly increase to enhance the C-C interaction while maximising the O-O separation. This trend is valid for both surfaces and for all applied potentials; nonetheless, the tilting is more marked for the (111) surface as a consequence of the distorted configuration where the C atoms of the dimer are between the hollow and bridge sites. The trends associated with the C-O bond lengths are even more interesting: the formation of the C-C bond results in a weakening of the C-O bond. This is clearly visible in Figure 6.3c-d where the C-O bond is progressively more elongated as the C-C distance decreases. This elongation is larger in the (100) case, consistently with a stronger and more stable dimer on this surface. Notably, the applied potential has an important effect on the C-O bond length. Negative applied potentials weaken the C-O bond, resulting in a larger C-O distance and favouring the C-C dimerisation. The correlation between the lowering of the activation barriers and the elongation of the *CO adsorbates with the applied bias is markedly evident on the (111) surface, where the decrease of the barriers between -1.25 V and -1.50 V vs SHE is coupled with a rigid shift over the C-O length along all reaction paths. Notice that independent *CO adsorbates, before dimerisation, already present a larger C-O bond length at highly negative potentials, confirming that the applied potential already partially destabilises the *CO, resulting in more reactive adsorbates for the C-C dimerisation. The weakening of the C-O bond is also the result of the surrounding water molecule influence. We observe this effect by evaluating the average number of H-bonds that each O atom of the *CO establishes with the solvent molecules (Figures 6.3e-f). The same geometrical definition of H-bonds discussed in ref. [157] and already employed in Chapter 5 is considered here. This analysis clearly proves the importance of the H-bond network in stabilising the dimer: as the C-C distance decreases, the number of H-bonds for each *CO increases, passing

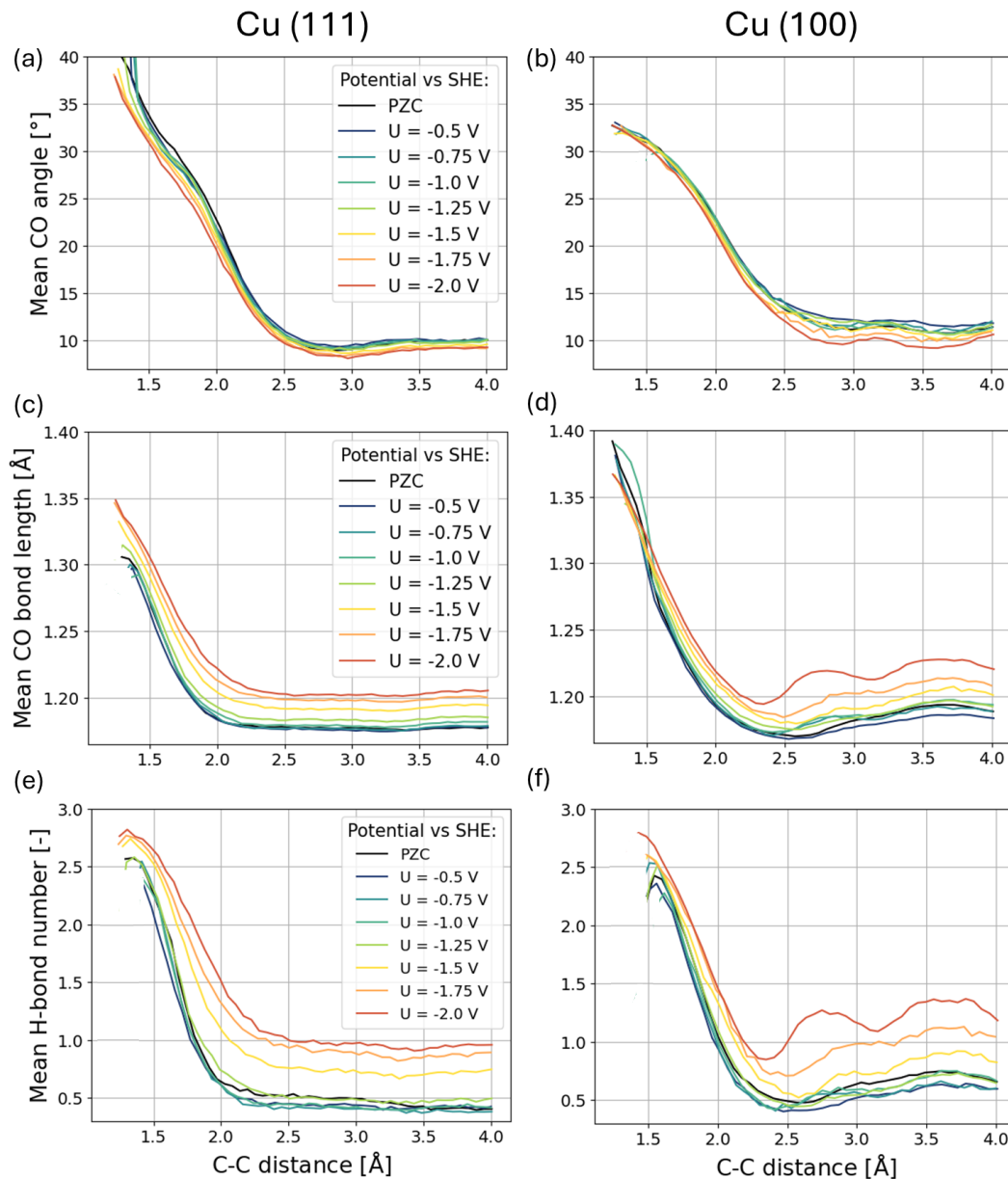


Fig. 6.3 Analysis of the $^*\text{CO}$ adsorbate properties on Cu (111) (left column) and Cu (100) (right column) surfaces during the CO dimerisation at different applied potentials. (a-b) average tilting of the $^*\text{CO}$ species measured as the angle between the C-O bond vector and the vector perpendicular to the surface, (c-d) mean C-O bond length and (e-f) mean number of H-bonds per $^*\text{CO}$. Results are determined by averaging between the two $^*\text{CO}$ species.

from less than 1 for the independent adsorbates to more than 2 for the final dimer product. Consistently with the variations of the C-O length, the highly negative applied potentials also increase the number of H-bonds of the dimer as well as the

single $^*\text{CO}$ species. The dipole flip of the H_2O molecules with a dipole pointing towards the surface, induced by the negative bias, facilitates the interaction of the H atoms with the O of the $^*\text{CO}$. Overall, a slightly higher number of H-bonds is observed for the Cu (100) case, indicating that, on this surface, the adsorbates have a slightly larger effective dipole moment, which enhances the hydrogen-bonding network between the adsorbates and the surrounding water molecules. We stress that these effects are visible only by explicitly including in the simulation the role of the water molecules as well as of the applied potential, calling for approaches such as TREC I able to include multiple physical elements in a unified atomistic picture.

Combining the different insights discussed in this section, we can summarise the effects of the applied potential on this process in the following picture. The applied bias, through the accumulation of negative surface charge, has a direct impact on stabilising the final dimer product since it exists as a negatively charged species, but it does not have a direct influence on the TSs since their charge is the same of two independent $^*\text{CO}$. Nonetheless, both the dimer and the TSs indirectly depend on the applied potential through the action of the surrounding water molecules. The water dipole flip induced by the applied potential increases the number of H-bonds that can be established between the $^*\text{CO}$ and water molecules, weakening the C-O bond and favouring the C-C formation. This partial stabilisation of the TSs resulting from the dipole reorientation is more evident on the Cu (111) surface, where the TSs are highly energetic and a more pronounced decrease of the activation barrier is observed between -1.25 V and -1.50 V vs SHE.

6.3 HER: the Volmer step

For a more complete view of Cu electrochemical interfaces, we include in the picture the presence of H adsorbates and the effects of water dissociation, focusing on the Volmer step of HER.

6.3.1 H adatom diffusion

Prior to studying the Volmer step of the HER, we discuss the characteristics of $^*\text{H}$ adsorbates, i.e., product species of the Volmer process. In this sense, we study the behaviour of a single $^*\text{H}$ adatom on the Cu surface in order to identify the prefer-

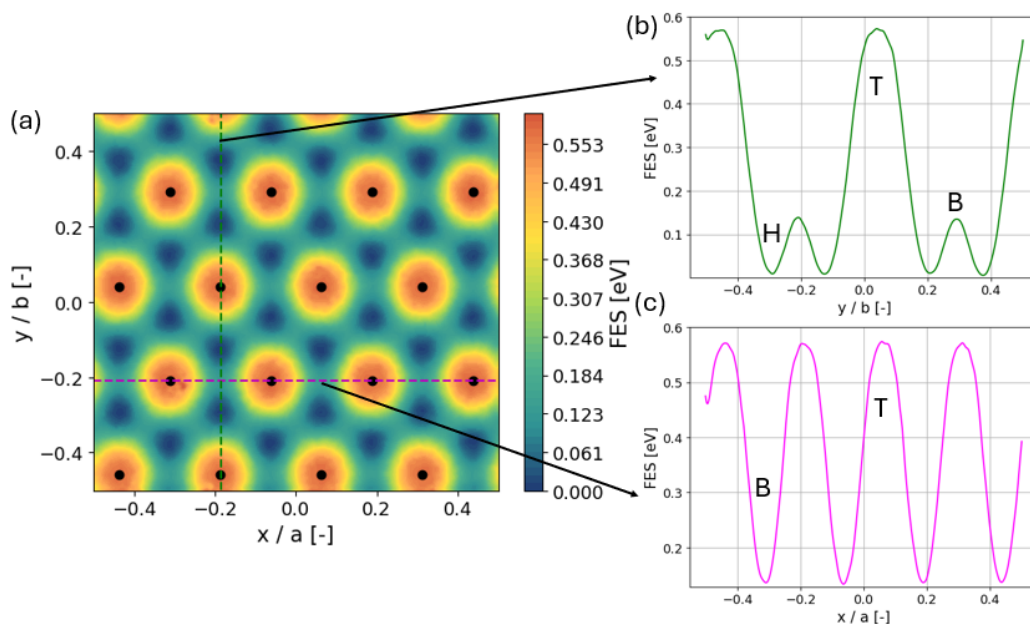


Fig. 6.4 FES of a single H adsorbed on a Cu (111) surface at the PZC. In-plane $x - y$ coordinates are reported scaled with respect to the $a - b$ lattice vectors. Black spots represent the average positions of the Cu surface atoms. Diffusion barriers between main sites are reported along 1-D x -cut (b) and y -cut (c) plots and averaging over equivalent cuts. The 'H', 'B' and 'T' acronyms stand for 'hollow', 'bridge' and 'top' sites, respectively.

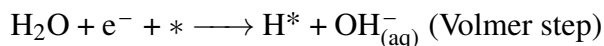
ential adsorption sites and the diffusion barriers among different surface sites. The investigation of the adsorption of individual hydrogen adsorbates (i.e., under low hydrogen coverage conditions) is consistent with the well-established evidence of the weak binding of hydrogen on Cu surfaces, which makes the occurrence of high hydrogen coverages highly unlikely [18, 179]. Consistent with the methodology employed in the previous chapters, explicit solvent molecules are present above the surface in order to accurately capture the varying interaction strengths of $^*\text{H}$ adatoms at hollow sites relative to those at more exposed positions (e.g., top sites). Enhanced-sampling MD simulations are performed to evaluate the diffusion barriers, placing a single $^*\text{H}$ adsorbate and biasing its x - y in-plane coordinates to explore all the sites of the surfaces. Figure 6.4a reports the free energy surface of the $^*\text{H}$ adatom over the (111) surface at the PZC. It is evident that the most favourable sites for hydrogen adsorption are the hollow ones, with no evident difference between fcc and hcp hollow sites. This is in line with results reported for other FCC metals [67, 146] at low H coverage, and it is confirmed by experimental evidence [187, 188]. More interesting are the evident energy differences (≈ 0.57 eV, see Figure 6.4b-c) between

the hollow and top sites, with the latter resulting unstable. We stress that previous static DFT studies that do not account for the solvent role predict a smaller energy difference between the two sites [189], proving the importance of properly accounting for the active water role as well as its dynamic effects to identify the correct relative stability of the different adsorption sites.

While the hollow sites are by far the most stable ones, the energy difference between the hollow and bridge is minimal (< 0.15 eV, see Figure 6.4b), resulting in diffusion barriers between adjacent hollow sites through bridge ones being easily overcome at room temperature. From this analysis, a scenario emerges in which $^*\text{H}$ adatoms exhibit high mobility on the Cu (111) surface, diffusing readily between adjacent hollow sites through bridge sites while only rarely occupying the top ones. The instability of $^*\text{H}$ on top sites at low coverage is an important element with consequences for the Volmer step, as discussed in the following section.

6.3.2 Volmer step: activation barriers and insights

In the previous section we supposed a $^*\text{H}$ adatom was already present on the surface without discussing the origin of this adsorbate. Here we address this point, focusing on the reaction of proton transfer from a water molecule to the Cu surface:



The geometrical CV defined to enhance the sampling of the proton transfer process is given by the relative difference between (i) the distance d_{OH} between the transferring hydrogen atom and the oxygen atom of the donor water molecule and (ii) the vertical distance $dz_{\text{H-surface}}$ between the hydrogen atom and the surface plane, i.e., $d_{\text{OH}} - dz_{\text{H-surface}}$. A graphical representation of the distances involved in the definition of this CV is reported in the lateral inset of Figure 6.5.

As visible in Figure 6.5, the free energy difference between the reagent and product basins is evident at low potentials ($U \geq -1.0$ V vs SHE), with the adsorbed hydrogen atom less thermodynamically favourable. Nevertheless, the picture changes at increasing negative surface charge: for $U \leq -1.5$ V vs SHE, the product basin progressively deepens, and at highly negative potentials, the $^*\text{H}$ adsorbates are in the lowest energy state, and with high backward barriers, that makes the back recombination to the water molecule unfeasible.

In addition to these thermodynamic considerations, the dependence of the activation

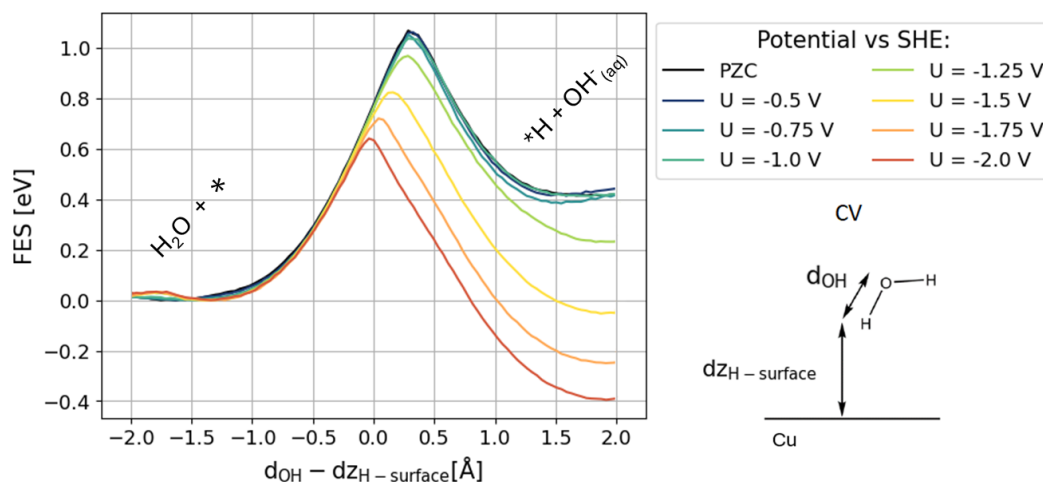


Fig. 6.5 Free energy profile for the Volmer step of HER at different values of applied potential. The left basin is associated with the reagents (undissociated water molecules on a clean surface), while the right one is related to the products (*H adsorbate and a hydroxide ion in solution). The lateral inset shows a sketch of the distances involved in the definition of the biased geometrical CV.

barrier on the applied potential is more interesting. For $U > -1.25$ V, i.e., for potentials more positive than the U_{th} value for the (111) surface, the barrier for the proton transfer from the solvent to the surface is high (>1 eV) and, above all, independent from the applied potential. In this condition the Volmer step and, consequently, the full HER reaction are highly unfeasible, confirming that the Cu electrocatalyst, above all in a neutral/alkaline environment (i.e., H_2O as the proton donor), is not an effective material for HER with respect to other noble metals such as Pt. This picture starts to change when the peak of the region I of the solvent is suppressed ($U \approx U_{th} = -1.25$ V vs SHE, cf. Section 5.3): here the barriers start to decrease as the potential gets more and more negative. Nonetheless, more negative potentials ($-1.75 \div -2.0$ V) are required to achieve lower barriers ($0.6 \div 0.7$ eV). Only in these conditions it is reasonable to assume that the proton transfer from the solvent toward the surface can be observed in typical operating conditions. This result is in agreement with experimental evidence that identifies highly negative potentials required to observe HER on Cu catalyst in alkaline conditions [22, 190]. Notice that the activation barriers for the Volmer step are lower than the CO dimerisation ones in this range of highly negative potentials (cf. Section 6.3.2), making more probable the production of molecular hydrogen rather than C_2 species on Cu (111).

Aside from this specific result for Cu (111) electrodes, it is interesting to investi-

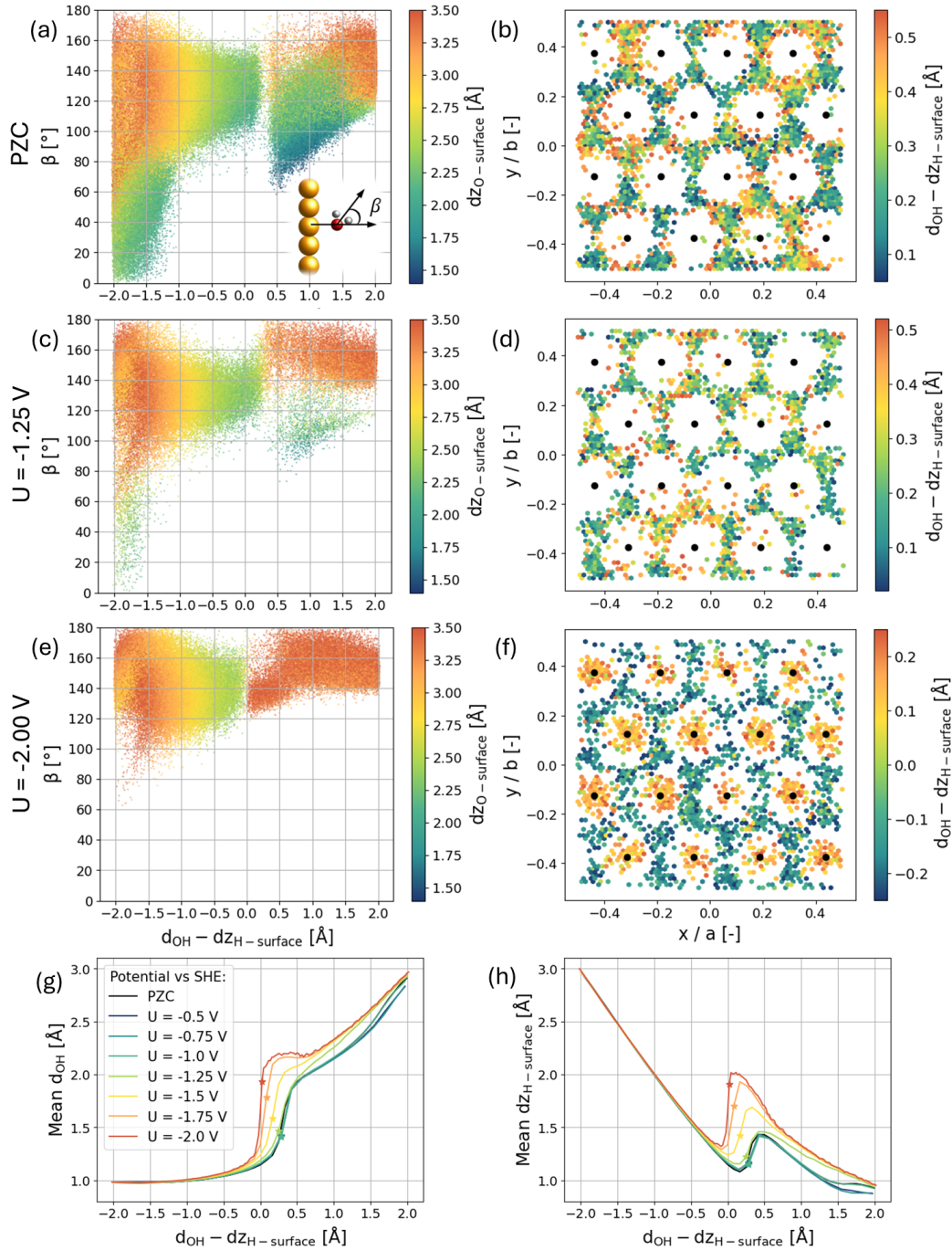


Fig. 6.6 Correlation between the β dipole angle distribution for the involved H_2O , the distance $dz_{\text{O-surface}}$ of the biased O atom and the biased CV ($d_{\text{OH}} - dz_{\text{H-surface}}$) at PZC (a), $U = -1.25$ V (c) and -2.00 V (e). The inset shows the geometrical definition of the dipole angle β . Scatter plots of the in-plane distribution of the transferred proton vs the biased CV at PZC (b), $U = -1.25$ V (d) and -2.00 V (f). Contributions of configurations with a biased CV value in an interval of ± 0.25 Å around the TSs are only reported. H coordinates are scaled by the in-plane $a - b$ lattice vectors. Black spots are the averaged positions of the Cu (111) surface atoms. Correlation between the mean O-H distance (g), the mean H-surface distance (h) and the biased CV for the different values of potential. The locations of the TSs are identified with a marker.

gate the correlation between the decrease of the Volmer activation barriers and the characteristic of interfacial water discussed in Chapter 5, such as the suppression of the first peak of the density. Figures 6.6a,c,e reproduce the H₂O dipole distribution and the distance of the biased O atoms along the reaction path at selected values of applied potential. It is evident that the proton transfer occurs only when the H₂O molecule has a dipole vector pointing towards the surface (i.e., H atoms exposed towards the surface, β angle $> 90^\circ$), regardless of the value of applied potentials. Hence, at PZC and low potentials (Figure 6.6a), the proton transfer requires having a water molecule with a dipole pointing towards the surface and at $2 \div 2.25$ Å from the surface. This implies a molecule in region I with hydrogen exposed towards the surface, but this orientation is highly unlikely in region I, as discussed in Chapter 5. On the contrary, for $U < U_{th}$ (Figures 6.6c,e), when the contribution of region I is suppressed, there are only H₂O molecules with already the optimal dipole orientation for supporting the proton transfer favouring the Volmer reaction. This explains the correlation between the decrease of the activation barrier and the suppression of region I.

To better understand the conditions that favour this reaction, we analyse the reaction sites over which the proton transfer occurs (Figures 6.6b,d,f). Notice these maps do not represent an instantaneous picture of the spatial position of the proton involved in the reaction but the cumulative contributions of all the proton transfers that occurred during the MD trajectory. To focus on the reaction events, we report in these plots only the contributions in a CV interval around the TSs. Consequently, the positions of the H atoms in the TSs do not necessarily represent the stable sites for the *H adatom in the product basin. At PZC and at low potentials (Figure 6.6b), the O-H bond cleavage occurs in correspondence of the hollow or bridge sites, and the transferred proton directly occupies these sites. On the contrary, at more negative potentials, when the Volmer reaction is more favoured, the transferred proton has to pass through the top sites prior to reaching the hollow/bridge one. This is clearly visible in Figure 6.6f, where for the CV values slightly larger than the TSs (reddish range of the colour palette), the proton position corresponds to the top site. This insight is aligned with the results reported for other more HER-active electrocatalysts [67, 146], which highlight the importance of having H on top sites to be involved in the HER. We also observe the importance of the proton on the top site during the Volmer step, but this state is unstable with respect to the hollow ones on Cu (111), cf. Section 6.3.1, making this electrocatalyst less effective for HER with respect to other noble metals.

In addition to the modulation of the barriers, the applied potential also shifts the positions of the TSs along the reaction coordinate, resulting in smaller values of $d_{OH} - dz_{H-surface}$. We investigate this shift correlating the biased CV with the average O-H bond length and the average H-surface distance for the TSs (Figures 6.6g and h, respectively). The mean O-H bond distance for the TSs is progressively longer at negative potentials, in line with the observations reported in Chapter 5 that show a more elongated O-H bond already without reactions due to the weakening of the covalent bonds due to the surface and, above all, the negative applied potential. At the same time the distance between the H and the surface for the TSs is also larger at negative potentials due to a more negative surface charge, enhancing the electrostatic attraction of H atoms at larger distances. However, comparing Figures 6.6g and h, it is visible that the increase of $dz_{H-surface}$ is larger at negative potentials with respect to d_{OH} at the TSs. This results in a shift of the TSs to smaller values of $d_{OH} - dz_{H-surface}$. We emphasise that all insights derived from Figure 6.6 rely on the enlarged statistical ensemble afforded by the ML-FF-based simulations. In contrast, conventional AIMD, even when combined with enhanced sampling techniques, would be unable to sample a sufficiently large set of configurations to robustly characterize the TSs.,

6.4 Conclusion

In this section we applied the TRECI scheme to develop ML-FFs able to include in the atomistic picture of electrochemical interfaces the description of some common reaction steps occurring on Cu surfaces in contact with water at reducing negative potentials. Representative reaction steps that are commonly regarded as limiting processes for the reactivity of Cu surfaces are selected as a benchmark to show the effectiveness of TRECI in describing reactions in the advanced atomistic setup presented in the previous chapters. Specifically, we study the CO dimerisation step along the CO₂RR to convert CO₂ to C₂ or C₂₊ products and the Volmer step in alkaline conditions for the HER.

From the ML perspective, the description of reactions represents an additional challenge since the TRECI scheme is coupled with enhanced sampling techniques such as OPES to properly sample TSs, different basins and get an exhaustive dataset along the whole reaction path. A more stringent balance between stability and

accuracy motivates a reconsideration of the dualism between robust, general-purpose descriptors and highly accurate, domain-specific backbones in the context of transfer learning. In this case, we investigate universal backbones that are also trained on reaction datasets, enabling an enhancement in the accuracy of general-purpose Franken ML-FFs with respect to models based on MACE-MP0, without incurring in any loss in stability.

With these improvements, we show the effectiveness of the TRECI scheme in delivering a detailed picture of the selected reaction steps in advanced simulation setups, getting beyond hybrid explicit-implicit solvent schemes and static DFT calculations that represent the state-of-the-art approach to evaluate the activation barriers for electrochemical reactions. Our results indicate that only highly negative potentials lead to a substantial reduction of the activation barriers for the Volmer step and the CO dimerisation on Cu (111), while the activation of the CO dimerisation is weakly dependent on the applied bias on the (100) surface, raising possible issues on the selectivity of pure copper electrocatalysts. Nonetheless, the detailed analysis enabled by TRECI provides mechanistic insights into the reactivity and may inform the rational design and screening of more promising catalytic materials. For the CO dimerisation, we highlight the effects of the different dimerisation sites in determining the reactivity of the surfaces. We also emphasise the dual role of the applied potential. On the one hand, the increase in negative surface charge induced by the applied bias directly stabilises the dimer products as negatively charged species. On the other hand, only via the enhancement of the H-bond network between the adsorbates and the water molecules, arising from the reorientation (dipole flipping) of H₂O at negative potentials, the applied bias partially stabilises the transition states, thereby weakly lowering the activation barriers. For the Volmer step of HER, we correlate the decrease of the activation barrier with the suppression of the first peak of the solvent density due to the negative potentials, and we highlight the role of the top Cu site in mediating an effective proton transfer. Comparing the effects of the applied potential on the activation of the discussed reaction steps, we recognise that the activation of the Volmer step of HER is directly influenced by the increase of the applied bias with respect to CO dimerisation, where the action of the applied potential is more convoluted. This indicates that the straightforward strategy of applying increasingly negative potentials to promote the reactions is ineffective for enhancing CO dimerisation, and it might favour other competitive processes such as HER on Cu (111). These considerations highlight the necessity of accounting for all

competing reaction pathways when assessing the performance of an electrocatalyst, because different processes can exhibit distinct dependencies on the applied potential. Consequently, this diversity in potential response further complicates the identification of an optimal potential window for selectively promoting a given reaction. In this framework, the versatility of TRECI in facing different reactions represents a solid starting point not only for a more advanced comprehension of the limiting reaction steps on Cu-based electrocatalysis but also for a wider inclusion of other reaction steps for CO₂RR, alongside a systematic study of competing processes. More generally, this example of reaction study enabled by TRECI models can inspire other works to employ a similar scheme to consider other reaction paths as well as other electrocatalyst materials, resulting in a promising route to advance the comprehension of electrochemical reactions from the atomistic point of view.

Conclusion

In this doctoral thesis, we studied electrochemical interfaces, specifically metal/water systems, under the application of an external electric potential using advanced atomistic simulations. We addressed the complexity of these systems by employing machine learning interatomic potentials to accelerate molecular dynamics, achieving scalability comparable to classical force fields together with *ab initio* accuracy. The thesis is motivated by three main current challenges in simulating electrochemical interfaces: 1) exhaustively capturing the dynamical effects and reproducing the full behaviour of water electrolytes via explicit solvent models; 2) accurately describing, at the *ab initio* level, the balance between covalent, hydrogen bonds and dispersion forces that govern the liquid water network; and 3) implementing a rigorous electrification scheme to simulate interfacial processes at constant potential while the surface charge is fluctuating. Machine learning interatomic potentials trained on DFT data can enable such simulations more efficiently than AIMD, but the generation of an *ab initio*-quality training dataset remains the main bottleneck, being computationally demanding, especially when moving beyond common approximations.

In this thesis we presented TRECI (TRansfer learning for ElectroChemical Interfaces), a general, data-efficient workflow to develop ML-FFs for electrochemical interfaces at constant applied potential. The workflow is optimised to minimise the number of DFT-labelled data points needed for training. TRECI achieves data efficiency through 1) an ML architecture based on transfer learning and 2) an effective active learning strategy that guides dataset generation based on current model performance. Instead of training a constant-potential model from scratch, we followed a transfer learning approach, building the descriptors of the new ML-FF on the node features of a pre-existing ML-FF. Data requirements are reduced because the descriptors are not retrained but fully inherited; only the regression block requires new DFT data at constant potential. Following this strategy, we delivered stable

and accurate ML-FFs for representative metal/water interfaces using only about one thousand DFT configurations for each electrode potential value, roughly a 75% reduction in data compared to training from scratch.

This data efficiency allows the use of more advanced, accurate, but computationally demanding DFT simulations to face the discussed modelling challenges. We employed a fully explicit water solvent and the meta-GGA SCAN exchange–correlation functional to correctly describe the liquid water network. We also modelled the applied potential with the complex, computationally expensive Double Reference Method, avoiding common electrification schemes based on implicit solvent approximations.

While the exchange-correlation functional and electrification schemes employed in this work represent a significant improvement over many commonly adopted approaches for electrochemical interfaces, the resulting ML models inevitably inherit the limitations of the underlying DFT reference. The demanding computational cost of the selected DFT set-ups often makes a systematic assessment of its validity limits computationally expensive via standard *ab initio* approaches. Going beyond static tests commonly employed to assess the validity of the DFT functional and electrification scheme, TRECI makes accessible a large number of structural and dynamical properties, also opening future perspectives for benchmarking the predictive power of these DFT setups under more realistic conditions and against experimental observables. In this regard, our data-efficient ML-FF strategy not only extends the spatial and temporal scales accessible to AIMD but also unlocks these demanding setups for production runs. These results can provide valuable feedback into the validity of the DFT description and inform the development of next-generation *ab initio* methodologies for electrochemical interfaces.

In the second part of the thesis, we validated the effectiveness of TRECI by developing ML-FFs for specific case studies: low-index copper surfaces in contact with a water solvent. The main effects of the applied potential were discussed in terms of the modulations of the structure of interfacial water. We clearly identified a threshold potential value over which the interfacial region of chemisorbed water molecules is completely suppressed and the surface has a major hydrophobic behaviour. This threshold potential is surface-specific, resulting from a balance between chemisorption and physisorption. While this picture is obtained in the absence of explicit ions in the solution, the results obtained in this limit of highly dilute electrolyte are still relevant by representing a necessary step to determine the baseline of the intrinsic

properties of the metal/water interface over which the roles of ionic species will be contextualised.

Based on these results of the equilibrium properties of the electrified interfaces, we developed ML-FFs via TRECI to address specific limiting reaction steps for the hydrogen evolution and the electrochemical conversion of CO₂ in multi-carbon products. Performing molecular dynamics simulations coupled with enhanced sampling techniques, we identified the differences between Cu (100) and Cu (111) in terms of reactivity for the CO–CO dimerisation reaction, with the latter surface largely unfavourable for this reaction. Aside from the performance of the electrocatalyst, we highlighted the role of the surface charge and the direct *CO ... H₂O interaction in determining the reactivity of the surface towards the CO dimerisation. While the surface charge does not strongly influence the transition states, the water dipole flip mechanism, induced by the applied potential, changes the number of H₂O molecules interacting with *CO. This indirect action of the applied potential results in a weakening of the C–O bonds in favour of the C–C dimerisation, partially reducing the activation barriers for the dimerisation at negative potentials. On the contrary, the hydrogen evolution reaction, i.e., the main competitive process of the CO₂ electroreduction, shows a more marked dependence on the applied potential. By studying the Volmer step of HER, we identified that, once the electric bias suppresses the region of chemisorbed H₂O molecules and the water dipole flip process is completed, the proton transfer between water and the surface is boosted, lowering the activation barriers. In this way, we clearly proved the applicability of TRECI for ML-FFs able to uncover not only the structural and dynamic properties of electrochemical interfaces under applied potential but also provide insights into the reaction mechanisms, correlating at the atomic scale diverse characteristics of the systems. We stress once again the importance of active-learning protocols to obtain these results when reactive processes are investigated using ML models trained on a limited number of reference configurations. The risk of extrapolation in reactive systems and under large applied potentials remains an important challenge, since these conditions drive the system towards regions of configuration space that can be insufficiently represented in the dataset. In this context, protocols such as DEAL are essential tools to effectively sample and select data points along the complex reaction paths typical of electrochemical processes.

While the effectiveness of TRECI was demonstrated for a specific system considered in this work, one of its main strengths lies in its modular design. The workflow

can be customised both at the level of the underlying DFT reference and of the ML framework, making it adaptable to the requirements of diverse electrochemical systems. We have already shown that, as the complexity of the interfaces increases, it is possible to improve the performance of the transfer learning by employing more advanced general-purpose models, improving in accuracy while not losing in stability. Within this framework, subsequent developments of TRECI will focus on refining computationally inexpensive, universal ML-FFs to establish a general-purpose backbone that is systematically adapted to specific systems. This backbone will then serve as a unique starting point for transfer learning, merging the stability of general-purpose descriptors with the accuracy of domain-specific models. Such developments will simplify active-learning workflows, reduce the need to construct PZC-specific models from scratch, and facilitate the extension of TRECI to increasingly realistic electrochemical systems.

At the same time the modularity of the TRECI scheme will also allow including new ML blocks to face challenges emerging with raising the complexity of the physics of the systems. Future enhancements for TRECI might include ML models to predict atomic charges along the MD trajectory, providing an explicit charge-aware picture to more effectively rationalise the interfacial processes.

Another important improvement to expand the capability of TRECI will be the augmentation of the receptive field of the ML-FFs to treat long-range interactions. The presented picture of electrified metal/water interfaces derived exclusively with short-range models is not significantly affected by the absence of explicit long-range schemes since we have shown that unscreened electrostatic contributions are negligible in the system covered in this thesis. However, long-range interaction effects will arise due to the presence of ionic species in the electrolyte. For these applications, TRECI is not limited in the applicability of long-range approaches, but our framework can actually serve as a fertile ground for the quickly emerging long-range schemes to be used in complex electrochemical interfaces. For example, TRECI can include the recently proposed 'Latent Ewald Summation' scheme (*npj Comput Mater* 11, 80, 2025), which also offers physical insights on the charge. In all the cases, the effectiveness of TRECI is unaffected by these enhancements since its data-efficient protocol always offers a strong basis for further advancements.

In this respect, TRECI is a wide platform ready to start studying not only selected reaction steps but also complex reaction paths, such as the long sequence of reactions of CO₂RR. Moreover, going beyond the simple copper catalyst studied in this thesis,

TRECI models will be useful to better understand the reactivity and selectivity of more advanced electrocatalysts, becoming a useful tool for rationalising and guiding the synthesis of more effective materials for electrocatalysis.

References

- [1] Olle Björneholm, Martin H. Hansen, Andrew Hodgson, Li-Min Liu, David T. Limmer, Angelos Michaelides, Philipp Pedevilla, Jan Rossmeisl, Huaze Shen, Gabriele Tocci, Eric Tyrode, Marie-Madeleine Walz, Josephina Werner, and Hendrik Bluhm. Water at Interfaces. *Chemical Reviews*, 116(13):7698–7726, July 2016.
- [2] Grazia Gonella, Ellen H. G. Backus, Yuki Nagata, Douwe J. Bonthuis, Philip Loche, Alexander Schlaich, Roland R. Netz, Angelika Kühnle, Ian T. McCrum, Marc T. M. Koper, Martin Wolf, Bernd Winter, Gerard Meijer, R. Kramer Campen, and Mischa Bonn. Water at charged interfaces. *Nature Reviews Chemistry*, 5(7):466–485, June 2021.
- [3] Axel Groß and Sung Sakong. Ab Initio Simulations of Water/Metal Interfaces. *Chemical Reviews*, 122(12):10746–10776, June 2022.
- [4] Nestor Perez. Electrochemical Corrosion. In *Materials Science: Theory and Engineering*, pages 835–898. Springer Nature Switzerland, Cham, 2024.
- [5] Koch Gerhardus, Varney Jeff, Thompson Neil, Moghissi Oliver, Gould Melissa, and Payer Joe. International measures of prevention, application, and economics of corrosion technologies study, 2016.
- [6] Zhi Wei Seh, Jakob Kibsgaard, Colin F. Dickens, Ib Chorkendorff, Jens K. Nørskov, and Thomas F. Jaramillo. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science*, 355(6321):eaad4998, 2017.
- [7] Soliu Oladejo Ganiyu and Carlos Alberto Martínez-Huitle. The use of renewable energies driving electrochemical technologies for environmental applications. *Current Opinion in Electrochemistry*, 22:211–220, 2020.
- [8] Javier Rodríguez-Varela, Ivonne L. Alonso-Lemus, Oumarou Savadogo, and Karthikeyan Palaniswamy. Overview: Current trends in green electrochemical energy conversion and storage. *Journal of Materials Research*, 36(20):4071–4083, October 2021.
- [9] Yun Wang, Yiheng Pang, Hui Xu, Andrew Martinez, and Ken S. Chen. PEM Fuel cell and electrolysis cell technologies and hydrogen infrastructure de-

- velopment – a review. *Energy & Environmental Science*, 15(6):2288–2328, 2022.
- [10] Naimat Ullah, Syed Shaheen Shah, Munzir H. Suliman, Fatma Ismail, Samira Kaci, Nor Aishah Saidina Amin, and Muhammad Usman. Electrochemical CO₂ Reduction: A Review toward Sustainable Energy Conversion and Storage. *Energy & Fuels*, 39(41):19614–19646, October 2025. Publisher: American Chemical Society.
- [11] Marie-Cécile Pera, Daniel Hissel, Hamid Gualous, and Christophe Turpin. *Electrochemical components*. John Wiley & Sons, 2013.
- [12] Frank C. Walsh and Derek Pletcher. *Electrochemical Engineering and Cell Design*, chapter 6, pages 95–111. John Wiley & Sons, Ltd, 2014.
- [13] Shripad T Revankar and Pradip Majumdar. *Fuel cells: principles, design, and analysis*. CRC press, 2014.
- [14] Arthur J. Shih, Mariana C. O. Monteiro, Federico Dattila, Davide Pavesi, Matthew Philips, Alisson H. M. Da Silva, Rafaël E. Vos, Kasinath Ojha, Sung-hak Park, Onno Van Der Heijden, Giulia Marcandalli, Akansha Goyal, Matias Villalba, Xiaoting Chen, G. T. Kasun Kalhara Gunasooriya, Ian McCrum, Rik Mom, Núria López, and Marc T. M. Koper. Water electrolysis. *Nature Reviews Methods Primers*, 2(1):84, October 2022.
- [15] Krishnan Veeramani, Gnanaprakasam Janani, Joonyoung Kim, Subramani Surendran, Jaehyoung Lim, Sebastian Cyril Jesudass, Shivraj Mahadik, Hyun-jung lee, Tae-Hoon Kim, Jung Kyu Kim, and Uk Sim. Hydrogen and value-added products yield from hybrid water electrolysis: A critical review on recent developments. *Renewable and Sustainable Energy Reviews*, 177:113227, 2023.
- [16] Mostafa El-Shafie. Hydrogen production by water electrolysis technologies: A review. *Results in Engineering*, 20:101426, 2023.
- [17] Sharad Maheshwari, Yawei Li, Naveen Agrawal, and Michael J. Janik. Density functional theory models for electrocatalytic reactions. In *Advances in Catalysis*, volume 63, pages 117–167. Elsevier, 2018.
- [18] Nicolas Dubouis and Alexis Grimaud. The hydrogen evolution reaction: from material to interfacial descriptors. *Chemical Science*, 10(40):9165–9181, 2019.
- [19] Stephanie Nitopi, Erlend Bertheussen, Soren B. Scott, Xinyan Liu, Albert K. Engstfeld, Sebastian Horch, Brian Seger, Ifan E. L. Stephens, Karen Chan, Christopher Hahn, Jens K. Nørskov, Thomas F. Jaramillo, and Ib Chorkendorff. Progress and Perspectives of Electrochemical CO₂ Reduction on Copper in Aqueous Electrolyte. *Chemical Reviews*, 119(12):7610–7672, June 2019.

- [20] Yong Hyun Moon, Na Yun Kim, Sung Min Kim, and Youn Jeong Jang. Recent advances in electrochemical nitrogen reduction reaction to ammonia from the catalyst to the system. *Catalysts*, 12(9), 2022.
- [21] Yao Zheng, Yan Jiao, Anthony Vasileff, and Shi-Zhang Qiao. The Hydrogen Evolution Reaction in Alkaline Solution: From Theory, Single Crystal Models, to Practical Electrocatalysts. *Angewandte Chemie International Edition*, 57(26):7568–7579, June 2018.
- [22] Huimin Wu, Chuanqi Feng, Lei Zhang, Jiujun Zhang, and David P. Wilkinson. Non-noble Metal Electrocatalysts for the Hydrogen Evolution Reaction in Water Electrolysis. *Electrochemical Energy Reviews*, 4(3):473–507, September 2021.
- [23] Congling Hu, Lei Zhang, and Jinlong Gong. Recent progress made in the mechanism comprehension and design of electrocatalysts for alkaline water splitting. *Energy & Environmental Science*, 12(9):2620–2645, 2019.
- [24] Jing Zhu, Liangsheng Hu, Pengxiang Zhao, Lawrence Yoon Suk Lee, and Kwok-Yin Wong. Recent Advances in Electrocatalytic Hydrogen Evolution Using Nanoparticles. *Chemical Reviews*, 120(2):851–918, January 2020.
- [25] Yueli Quan, Jiexin Zhu, and Gengfeng Zheng. Electrocatalytic reactions for converting CO_2 to value-added products. *Small Science*, 1(10):2100043, 2021.
- [26] Shuyu Liang, Liang Huang, Yanshan Gao, Qiang Wang, and Bin Liu. Electrochemical reduction of CO_2 to CO over transition metal/n-doped carbon catalysts: The active sites and reaction mechanism. *Advanced Science*, 8(24):2102886, 2021.
- [27] Alexander Bagger, Wen Ju, Ana Sofia Varela, Peter Strasser, and Jan Rossmeisl. Electrochemical CO_2 reduction: A classification problem. *ChemPhysChem*, 18(22):3266–3273, 2017.
- [28] Y. Hori. *Electrochemical CO_2 Reduction on Metal Electrodes*, pages 89–189. Springer New York, New York, NY, 2008.
- [29] D Kolb. Reconstruction phenomena at metal-electrolyte interfaces. *Progress in Surface Science*, 51(2):109–173, February 1996.
- [30] Xinfu Xia and Max L. Berkowitz. Electric-Field Induced Restructuring of Water at a Platinum-Water Interface: A Molecular Dynamics Computer Simulation. *Physical Review Letters*, 74(16):3193–3196, April 1995.
- [31] T Iwasita and F.C Nart. In situ infrared spectroscopy at electrochemical interfaces. *Progress in Surface Science*, 55(4):271–340, August 1997.
- [32] G. Cicero, A. Calzolari, S. Corni, and A. Catellani. Anomalous Wetting Layer at the Au(111) Surface. *The Journal of Physical Chemistry Letters*, 2(20):2582–2586, October 2011.

- [33] Andrea Auer, Xing Ding, Aliaksandr S. Bandarenka, and Julia Kunze-Liebhäuser. The Potential of Zero Charge and the Electrochemical Interface Structure of Cu(111) in Alkaline Solutions. *The Journal of Physical Chemistry C*, 125(9):5020–5028, March 2021.
- [34] Michele Giovanni Bianchi, Michele Re Fiorentin, Francesca Risplendi, Candido Fabrizio Pirri, Michele Parrinello, Luigi Bonati, and Giancarlo Cicero. Electrochemical interfaces at constant potential: Data-efficient transfer learning for machine-learning-based molecular dynamics, 2025.
- [35] Yike Huang, Xiaohong Liu, Shu Li, and Tianying Yan. Development of mean-field electrical double layer theory*. *Chinese Physics B*, 25(1):016801, dec 2015.
- [36] Allen J. Bard, Larry R. Faulkner, and Henry S. White. *Electrochemical Methods: Fundamentals and Applications*. John Wiley & Sons, Inc, 2022.
- [37] Matthew A. Brown, Alok Goel, and Zareen Abbas. Effect of electrolyte concentration on the stern layer thickness at a charged interface. *Angewandte Chemie International Edition*, 55(11):3790–3794, 2016.
- [38] Javier Carrasco, Andrew Hodgson, and Angelos Michaelides. A molecular perspective of water at metal interfaces. *Nature Materials*, 11(8):667–674, August 2012.
- [39] Ling-Yun Hu, Xiao-Xuan Huang, Ru-Yu Zhou, Yao-Hui Wang, Zhi-Lin Yang, and Jian-Feng Li. Experimental characterization technique to probe interfacial water. *Journal of Catalysis*, 430:115355, February 2024.
- [40] Andrea Auer, Bernhard Eder, and Franz J. Giessibl. Electrochemical AFM/STM with a qPlus sensor: A versatile tool to study solid-liquid interfaces. *The Journal of Chemical Physics*, 159(17):174201, November 2023.
- [41] Takeshi Fukuma and Ricardo Garcia. Atomic- and Molecular-Resolution Mapping of Solid–Liquid Interfaces by 3D Atomic Force Microscopy. *ACS Nano*, 12(12):11785–11797, December 2018.
- [42] Michael F. Toney, Jason N. Howard, Jocelyn Richer, Gary L. Borges, Joseph G. Gordon, Owen R. Melroy, David G. Wiesler, Dennis Yee, and Larry B. Sorensen. Distribution of water molecules at ag(111)/electrolyte interface as studied with surface x-ray scattering. *Surface Science*, 335:326–332, 1995. Proceedings of the IUVSTA Workshop on Surface Science and Electrochemistry.
- [43] Hoydoo You, C. A. Melendres, Z. Nagy, V. A. Maroni, W. Yun, and R. M. Yonco. X-ray-reflectivity study of the copper-water interface in a transmission geometry under *in situ* electrochemical control. *Physical Review B*, 45(19):11288–11298, May 1992.

- [44] Jonas Lindner, Ulrich Ross, Tobias Meyer, Sung Sakong, Axel Gross, Michael Seibt, and Christian Jooss. Atomic scale ordering of liquid water at a dynamic pt(111) interface under electrochemical conditions imaged by electron holography. *Advanced Energy Materials*, n/a(n/a):e05756, 2026.
- [45] Nawras Abidi, Kang Rui Garrick Lim, Zhi Wei Seh, and Stephan N. Steinmann. Atomistic modeling of electrocatalysis: Are we there yet? *WIREs Computational Molecular Science*, 11(3):e1499, May 2021.
- [46] Michele Re Fiorentin, Michele G. Bianchi, Magnus A. H. Christiansen, Anna Ciotti, Francesca Risplendi, Wei Wang, Elvar Ö. Jónsson, Hannes Jónsson, and Giancarlo Cicero. Methodological frameworks for computational electrocatalysis: From theory to practice. *Small Methods*, 10(5):e01542, 2026.
- [47] Hongyan Zhao, Xinmao Lv, and Yang-Gang Wang. Realistic Modeling of the Electrocatalytic Process at Complex Solid-Liquid Interface. *Advanced Science*, 10(32):2303677, November 2023.
- [48] Chao Zhang, Thomas Sayer, Jürg Hutter, and Michiel Sprik. Modelling electrochemical systems with finite field molecular dynamics. *Journal of Physics: Energy*, 2(3):032005, July 2020.
- [49] Qiang Li, Yixin Ouyang, Shuaihua Lu, Xiaowan Bai, Yehui Zhang, Li Shi, Chongyi Ling, and Jinlan Wang. Perspective on theoretical methods and modeling relating to electro-catalysis processes. *Chemical Communications*, 56(69):9937–9949, 2020.
- [50] Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2 edition, August 2020.
- [51] David S. Sholl and Janice A. Steckel. *Density Functional Theory: A Practical Introduction*. Wiley, 1 edition, March 2009.
- [52] Oliver T. Unke, Stefan Chmiela, Huziel E. Saucedo, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine Learning Force Fields. *Chemical Reviews*, 121(16):10142–10186, August 2021.
- [53] Laura Scalfi, Mathieu Salanne, and Benjamin Rotenberg. Molecular Simulation of Electrode-Solution Interfaces. *Annual Review of Physical Chemistry*, 72(1):189–212, April 2021.
- [54] Kathleen Schwarz and Ravishankar Sundararaman. The electrochemical interface in first-principles calculations. *Surface Science Reports*, 75(2):100492, May 2020.
- [55] Jia-Bo Le, Qi-Yuan Fan, Jie-Qiong Li, and Jun Cheng. Molecular origin of negative component of Helmholtz capacitance at electrified Pt(111)/water interface. *Science Advances*, 6(41):eabb1219, October 2020.

- [56] Alejandra Rendón-Calle, Santiago Builes, and Federico Calle-Vallejo. Substantial improvement of electrocatalytic predictions by systematic assessment of solvent effects on adsorption energies. *Applied Catalysis B: Environmental*, 276:119147, November 2020.
- [57] Jia-Bo Le, Xiao-Hui Yang, Yong-Bin Zhuang, Mei Jia, and Jun Cheng. Recent Progress toward Ab Initio Modeling of Electrocatalysis. *The Journal of Physical Chemistry Letters*, 12(37):8924–8931, September 2021.
- [58] Corinne M. Gray, Karthikeyan Saravanan, Guofeng Wang, and John A. Keith. Quantifying solvation energies at solid/liquid interfaces using continuum solvation methods. *Molecular Simulation*, 43(5-6):420–427, 2017.
- [59] Qiang Zhang and Aravind Asthagiri. Solvation effects on dft predictions of orr activity on metal surfaces. *Catalysis Today*, 323:35–43, 2019. Special Issue Honoring Umit S. Ozkan: 2017 ACS Henry H. Storch Award Winner.
- [60] Duy Le. An Explicit-Implicit Hybrid Solvent Model for Grand Canonical Simulations of the Electrochemical Environment, March 2023.
- [61] Oliviero Andreussi, Nicolas Georg Hörmann, Francesco Nattino, Giuseppe Fisicaro, Stefan Goedecker, and Nicola Marzari. Solvent-aware interfaces in continuum solvation. *Journal of Chemical Theory and Computation*, 15(3):1996–2009, 03 2019.
- [62] S. M. Rezwanul Islam, Foroogh Khezeli, Stefan Ringe, and Craig Plaisance. An implicit electrolyte model for plane wave density functional theory exhibiting nonlinear response and a nonlocal cavity definition. *The Journal of Chemical Physics*, 159(23):234117, 12 2023.
- [63] Luyu Yang, Chengkai Jin, and Xunhua Zhao. Hybrid explicit-droplet/implicit solvation model to accelerate constant-potential molecular dynamics simulations. *The Journal of Chemical Physics*, 163(22):224101, 12 2025.
- [64] Nitish Govindarajan, Georg Kastlunger, Joseph A. Gauthier, Jun Cheng, Ivo Filot, Arthur Hagopian, Heine Anton Hansen, Jun Huang, Piotr M. Kowalski, Jinwen Liu, Juan M. Lombardi, Mikael Maraschin, Andrew Peterson, Hermanth S. Pillai, Hector Prats, Conor J. Price, René van Roij, Jan Rossmeisl, Ranga Rohit Seemakurthi, Seung-Jae Shin, Audrey Smith, Jia-Xin Zhu, and Katharina Doblhoff-Dier. The Intricacies of Computational Electrochemistry. *ACS Energy Letters*, 10(9):4277–4288, September 2025.
- [65] Egill Skúlason, Vladimir Tripkovic, Márten E. Björketun, Sigríður Gudmundsdóttir, Gustav Karlberg, Jan Rossmeisl, Thomas Bligaard, Hannes Jónsson, and Jens K. Nørskov. Modeling the electrochemical hydrogen oxidation and evolution reactions on the basis of density functional theory calculations. *The Journal of Physical Chemistry C*, 114(42):18182–18197, 10 2010.

- [66] Christopher D. Taylor, Sally A. Wasileski, Jean-Sebastien Filhol, and Matthew Neurock. First principles reaction modeling of the electrochemical interface: Consideration and calculation of a tunable surface potential from atomic and electronic structure. *Physical Review B*, 73(16):165402, 2006.
- [67] Federico Raffone, Rémi Khatib, Marialore Sulpizi, and Clotilde Cucinotta. Revealing the molecular interplay of coverage, wettability, and capacitive response at the Pt(111)-water solution interface under bias. *Communications Chemistry*, 8(1):58, February 2025.
- [68] Zhaoming Xia and Hai Xiao. Grand Canonical Ensemble Modeling of Electrochemical Interfaces Made Simple. *Journal of Chemical Theory and Computation*, 19(15):5168–5175, August 2023.
- [69] Marko M. Melander, Mikael J. Kuisma, Thorbjørn Erik Kjøppen Christensen, and Karoliina Honkala. Grand-canonical approach to density functional theory of electrocatalytic systems: Thermodynamics of solid-liquid interfaces at constant ion and electrode potentials. *The Journal of Chemical Physics*, 150(4):041706, 11 2018.
- [70] Nicolas G. Hörmann, Oliviero Andreussi, and Nicola Marzari. Grand canonical simulations of electrochemical interfaces in implicit solvation models. *The Journal of Chemical Physics*, 150(4):041730, 01 2019.
- [71] Arthur Hagopian, Marie-Liesse Doublet, Jean-Sébastien Filhol, and Tobias Binninger. Advancement of the homogeneous background method for the computational simulation of electrochemical interfaces. *Journal of Chemical Theory and Computation*, 18(3):1883–1893, 2022.
- [72] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.
- [73] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.
- [74] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.
- [75] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566–569, Aug 1980.
- [76] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048–5079, May 1981.
- [77] John P. Perdew and Wang Yue. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B*, 33:8800–8802, Jun 1986.
- [78] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

- [79] Mohan Chen, Hsin-Yu Ko, Richard C. Remsing, Marcos F. Calegari Andrade, Biswajit Santra, Zhaoru Sun, Annabella Selloni, Roberto Car, Michael L. Klein, John P. Perdew, and Xifan Wu. Ab initio theory and modeling of water. *Proceedings of the National Academy of Sciences*, 114(41):10846–10851, October 2017.
- [80] Jianwei Sun, Adrienn Ruzsinszky, and John P. Perdew. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Physical Review Letters*, 115(3):036402, July 2015.
- [81] Jianwei Sun, Richard C. Remsing, Yubo Zhang, Zhaoru Sun, Adrienn Ruzsinszky, Haowei Peng, Zenghui Yang, Arpita Paul, Umesh Waghmare, Xifan Wu, Michael L. Klein, and John P. Perdew. Scan: An efficient density functional yielding accurate structures and energies of diversely-bonded materials, 2015.
- [82] James W. Furness, Aaron D. Kaplan, Jinliang Ning, John P. Perdew, and Jianwei Sun. Accurate and Numerically Efficient r^2 SCAN Meta-Generalized Gradient Approximation. *The Journal of Physical Chemistry Letters*, 11(19):8208–8215, October 2020.
- [83] James W. Furness, Aaron D. Kaplan, Jinliang Ning, John P. Perdew, and Jianwei Sun. Construction of meta-GGA functionals through restoration of exact constraint adherence to regularized SCAN functionals. *The Journal of Chemical Physics*, 156(3):034109, January 2022.
- [84] H Hellmann. *Einführung in die quantumchemie*. 1937.
- [85] R. P. Feynman. Force in molecules. *Phys. Rev.*, 56:340–343, 1939.
- [86] Efrem Braun, Justin Gilmer, Heather B. Mayes, David L. Mobley, Jacob I. Monroe, Samarjeet Prasad, and Daniel M. Zuckerman. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science*, 1(1), 2019.
- [87] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, Jul 1967.
- [88] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, Mar 1985.
- [89] Ville Korpelin, Toni Kiljunen, Marko M. Melander, Miguel A. Caro, Henrik H. Kristoffersen, Nisha Mammen, Vesa Apaja, and Karoliina Honkala. Addressing Dynamics at Catalytic Heterogeneous Interfaces with DFT-MD: Anomalous Temperature Distributions from Commonly Used Thermostats. *The Journal of Physical Chemistry Letters*, 13(11):2644–2652, March 2022. Publisher: American Chemical Society (ACS).
- [90] Glenn J. Martyna, Michael L. Klein, and Mark Tuckerman. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*, 97(4):2635–2643, August 1992.

- [91] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, January 2007.
- [92] Dominik Marx and Michele Parrinello. Ab initio path-integral molecular dynamics. *Zeitschrift für Physik B Condensed Matter*, 95(2):143–144, June 1994.
- [93] Michele Ceriotti, Wei Fang, Peter G. Kusalik, Ross H. McKenzie, Angelos Michaelides, Miguel A. Morales, and Thomas E. Markland. Nuclear Quantum Effects in Water and Aqueous Systems: Experiment, Theory, and Current Challenges. *Chemical Reviews*, 116(13):7529–7550, July 2016. Publisher: American Chemical Society.
- [94] Mark E Tuckerman. Ab initio molecular dynamics: basic concepts, current trends and novel applications. *Journal of Physics: Condensed Matter*, 14(50):R1297–R1355, December 2002.
- [95] Judith A. Harrison, J. David Schall, Sabina Maskey, Paul T. Mikulski, M. Todd Knippenberg, and Brian H. Morrow. Review of force fields and intermolecular potentials used in atomistic computational materials research. *Applied Physics Reviews*, 5(3):031104, August 2018.
- [96] Itai Leven, Hongxia Hao, Songchen Tan, Xingyi Guan, Katheryn A. Penrod, Dooman Akbarian, Benjamin Evangelisti, Md Jamil Hossain, Md Mahbubul Islam, Jason P. Koski, Stan Moore, Hasan Metin Aktulga, Adri C. T. van Duin, and Teresa Head-Gordon. Recent Advances for Improving the Accuracy, Transferability, and Efficiency of Reactive Force Fields. *Journal of Chemical Theory and Computation*, 17(6):3237–3251, June 2021. Publisher: American Chemical Society.
- [97] Sergei Manzhos and Tucker Carrington. Neural Network Potential Energy Surfaces for Small Molecules and Reactions. *Chemical Reviews*, 121(16):10187–10217, August 2021. Publisher: American Chemical Society (ACS).
- [98] Teddy Koker, Keegan Quigley, Eric Taw, Kevin Tibbetts, and Lin Li. Higher-order equivariant neural networks for charge density prediction in materials. *npj Computational Materials*, 10(1):161, July 2024.
- [99] Linfeng Zhang, Mohan Chen, Xifan Wu, Han Wang, Weinan E, and Roberto Car. Deep neural network for the dielectric response of insulators. *Physical Review B*, 102(4), July 2020. Publisher: American Physical Society (APS).
- [100] Tsz Wai Ko, Jonas A. Finkler, Stefan Goedecker, and Jörg Behler. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nature Communications*, 12(1), January 2021. Publisher: Springer Science and Business Media LLC.
- [101] Baron Peters. *Reaction Rate Theory and Rare Events*. Elsevier, Amsterdam ; Cambrige, MA, 2017.

- [102] Stephen J. Klippenstein, Vijay S. Pande, and Donald G. Truhlar. Chemical Kinetics and Mechanisms of Complex Systems: A Perspective on Recent Theoretical Advances. *Journal of the American Chemical Society*, 136(2):528–546, January 2014. Publisher: American Chemical Society.
- [103] Danya Lyu, Jinchang Xu, and Zhenyou Wang. Time-resolved in situ vibrational spectroscopy for electrocatalysis: challenge and opportunity. *Frontiers in Chemistry*, 11:1231886, July 2023.
- [104] Jérôme Hénin, Tony Lelièvre, Michael R Shirts, Omar Valsson, and Lucie Delemotte. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living Journal of Computational Molecular Science*, 4(1), 2022.
- [105] Omar Valsson, Pratyush Tiwary, and Michele Parrinello. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annual Review of Physical Chemistry*, 67(1):159–184, May 2016.
- [106] Enrico Trizio, Andrea Rizzi, Pablo M. Piaggi, Michele Invernizzi, and Luigi Bonati. Advanced simulations with plumed: Opes and machine learning collective variables, 2024.
- [107] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters*, 100(2):020603, January 2008.
- [108] Michele Invernizzi and Michele Parrinello. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *The Journal of Physical Chemistry Letters*, 11(7):2731–2736, April 2020.
- [109] Giovanni Bussi and Alessandro Laio. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics*, 2(4):200–212, April 2020.
- [110] Jörg Behler and Michele Parrinello. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters*, 98(14):146401, April 2007.
- [111] Joe D. Morrow, John L. A. Gardner, and Volker L. Deringer. How to validate machine-learned interatomic potentials. *The Journal of Chemical Physics*, 158(12):121501, March 2023. arXiv:2211.12484 [physics].
- [112] Felix Musil, Andrea Grisafi, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Physics-inspired structural representations for molecules and materials, August 2021. arXiv:2101.04673 [physics].
- [113] Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B*, 99(1):014104, January 2019.

- [114] Dávid Péter Kovács, Cas Van Der Oord, Jiri Kucera, Alice E. A. Allen, Daniel J. Cole, Christoph Ortner, and Gábor Csányi. Linear Atomic Cluster Expansion Force Fields for Organic Molecules: Beyond RMSE. *Journal of Chemical Theory and Computation*, 17(12):7696–7711, December 2021.
- [115] Ilyes Batatia, Simon Batzner, Dávid Péter Kovács, Albert Musaelian, Gregor N. C. Simm, Ralf Drautz, Christoph Ortner, Boris Kozinsky, and Gábor Csányi. The Design Space of E(3)-Equivariant Atom-Centered Interatomic Potentials, November 2022. arXiv:2205.06643 [stat].
- [116] Jonathan Vandermause, Yu Xie, Jin Soo Lim, Cameron J. Owen, and Boris Kozinsky. Active learning of reactive Bayesian force fields applied to heterogeneous catalysis dynamics of H/Pt. *Nature Communications*, 13(1):5183, September 2022.
- [117] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry, 2017.
- [118] Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields, January 2023. arXiv:2206.07697 [stat].
- [119] Junfan Xia, Yaolong Zhang, and Bin Jiang. Accuracy Assessment of Atomistic Neural Network Potentials: The Impact of Cutoff Radius and Message Passing. *The Journal of Physical Chemistry A*, 127(46):9874–9883, November 2023.
- [120] Ammar Daskin. A unifying primary framework for QGNNs from quantum graph states. *The European Physical Journal Special Topics*, October 2024.
- [121] Yutack Park, Jaesun Kim, Seungwoo Hwang, and Seungwu Han. Scalable Parallel Algorithm for Graph Neural Network Interatomic Potentials in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*, 20(11):4857–4868, June 2024. Publisher: American Chemical Society.
- [122] Kevin P. Murphy. *Machine learning: a probabilistic perspective*. Adaptive computation and machine learning series. MIT Press, Cambridge, Mass., 4. print. (fixed many typos) edition, 2013.
- [123] Jonathan Vandermause, Steven B. Torrisi, Simon Batzner, Yu Xie, Lixin Sun, Alexie M. Kolpak, and Boris Kozinsky. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Computational Materials*, 6(1):20, March 2020.
- [124] Gurjot Dhaliwal, Prasanth B. Nair, and Chandra Veer Singh. Machine learned interatomic potentials using random features. *npj Computational Materials*, 8(1):7, January 2022.

- [125] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.
- [126] Pietro Novelli, Giacomo Meanti, Pedro J. Buigues, Lorenzo Rosasco, Michele Parrinello, Massimiliano Pontil, and Luigi Bonati. Fast and Fourier features for transfer learning of interatomic potentials. *npj Computational Materials*, 11(1):293, September 2025.
- [127] Christoph Schran, Krystof Brezina, and Ondrej Marsalek. Committee neural network potentials control generalization errors and enable active learning, September 2020.
- [128] Ryan Jacobs, Dane Morgan, Siamak Attarian, Jun Meng, Chen Shen, Zhenghao Wu, Clare Yijia Xie, Julia H. Yang, Nongnuch Artrith, Ben Blaiszik, Gerbrand Ceder, Kamal Choudhary, Gabor Csanyi, Ekin Dogus Cubuk, Bowen Deng, Ralf Drautz, Xiang Fu, Jonathan Godwin, Vasant Honavar, Olexandr Isayev, Anders Johansson, Boris Kozinsky, Stefano Martiniani, Shyue Ping Ong, Igor Poltavsky, Kj Schmidt, So Takamoto, Aidan P. Thompson, Julia Westermayr, and Brandon M. Wood. A practical guide to machine learning interatomic potentials – Status and future. *Current Opinion in Solid State and Materials Science*, 35:101214, March 2025.
- [129] Simone Perego and Luigi Bonati. Data efficient machine learning potentials for modeling catalytic reactivity via active learning and enhanced sampling. *npj Computational Materials*, 10(1):291, December 2024.
- [130] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning, 2020.
- [131] Michael S. Chen, Joonho Lee, Hong-Zhou Ye, Timothy C. Berkelbach, David R. Reichman, and Thomas E. Markland. Data-Efficient Machine Learning Potentials from Transfer Learning of Periodic Correlated Electronic Structure Methods: Liquid Water at AFQMC, CCSD, and CCSD(T) Accuracy. *Journal of Chemical Theory and Computation*, 19(14):4510–4519, July 2023.
- [132] Viktor Zaverkin, David Holzmüller, Luca Bonferraro, and Johannes Kästner. Transfer learning for chemically accurate interatomic neural network potentials. *Phys. Chem. Chem. Phys.*, 25(7):5383–5396, 2023.
- [133] Jon López-Zorrilla, Xabier M. Aretxabaleta, and Hegoi Manzano. Exploring the Polymorphism of Dicalcium Silicates Using Transfer Learning Enhanced Machine Learning Atomic Potentials. *Journal of Chemical Theory and Computation*, 20(17):7682–7690, September 2024. Publisher: American Chemical Society.

- [134] An Niza El Aisnada, Kajjana Boonpalit, Robin van der Kruit, Koen M. Draijer, Jon López-Zorrilla, Masahiro Miyauchi, Akira Yamaguchi, and Nongnuch Artrith. Cost-Effective Strategy of Enhancing Machine Learning Potentials by Transfer Learning from a Multicomponent Data Set on `ænet-PyTorch`. *The Journal of Physical Chemistry C*, 129(1):658–669, January 2025. Publisher: American Chemical Society.
- [135] John Falk, Luigi Bonati, Pietro Novelli, Michele Parrinello, and Massimiliano Pontil. Transfer learning for atomistic simulations using GNNs and kernel mean embeddings, January 2024. arXiv:2306.01589 [cs].
- [136] Sebastien Röcken and Julija Zavadlav. Enhancing Machine Learning Potentials through Transfer Learning across Chemical Elements. *Journal of Chemical Information and Modeling*, 65(14):7406–7414, July 2025. Publisher: American Chemical Society.
- [137] E. O. Khazieva, N. M. Chtchelkatchev, and R. E. Ryltsev. Transfer learning for accurate description of atomic transport in Al–Cu melts. *The Journal of Chemical Physics*, 161(17):174101, November 2024.
- [138] Xu Huang, Bowen Deng, Peichen Zhong, Aaron D. Kaplan, Kristin A. Persson, and Gerbrand Ceder. Cross-functional transferability in foundation machine learning interatomic potentials. *npj Computational Materials*, 11(1):313, October 2025.
- [139] Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, November 2022.
- [140] Shivam Tripathi, Luigi Bonati, Simone Perego, and Michele Parrinello. How Poisoning Is Avoided in a Step of Relevance to the Haber–Bosch Catalysis. *ACS Catalysis*, 14(7):4944–4950, April 2024. Publisher: American Chemical Society.
- [141] Simone Perego, Maximilian Purcel, Yannick Baum, Shilong Chen, Astrid Sophie Müller, Michele Parrinello, Malte Behrens, Martin Muhler, and Luigi Bonati. No Time for Nitrides: How Cobalt Alloying Promotes Iron Catalysts for Ammonia Decomposition. *ACS Catalysis*, 15(19):16690–16702, October 2025. Publisher: American Chemical Society.
- [142] Shivam Tripathi, Luigi Bonati, Simone Perego, and Michele Parrinello. Nitrogen adsorption and dissociation on flat and stepped $\text{Fe}(110)$ surfaces. *Journal of Catalysis*, 454:116576, 2026.
- [143] Stephan N. Steinmann, Qing Wang, and Zhi Wei Seh. How machine learning can accelerate electrocatalysis discovery and optimization. *Mater. Horiz.*, 10:393–406, 2023.

- [144] Suresh Kondati Natarajan and Jörg Behler. Neural network molecular dynamics simulations of solid–liquid interfaces: water at low-index copper surfaces. *Physical Chemistry Chemical Physics*, 18(41):28704–28725, 2016.
- [145] August E. G. Mikkelsen, Jakob Schiøtz, Tejs Vegge, and Karsten W. Jacobsen. Is the water/Pt(111) interface ordered at room temperature? *The Journal of Chemical Physics*, 155(22), December 2021.
- [146] Peter S. Rice, Zhi-Pan Liu, and P. Hu. Hydrogen Coupling on Platinum Using Artificial Neural Network Potentials and DFT. *The Journal of Physical Chemistry Letters*, 12(43):10637–10645, November 2021.
- [147] Jia-Xin Zhu and Jun Cheng. Machine Learning Potential for Electrochemical Interfaces with Hybrid Representation of Dielectric Response. *Physical Review Letters*, 135(1), July 2025.
- [148] Xi Chen, Muammar El Khatib, Per Lindgren, Adam Willard, Andrew J. Medford, and Andrew A. Peterson. Atomistic learning in the electronically grand-canonical ensemble. *npj Computational Materials*, 9(1):73, May 2023.
- [149] Nicolas Bergmann, Nicéphore Bonnet, Nicola Marzari, Karsten Reuter, and Nicolas G. Hörmann. Machine Learning the Energetics of Electrified Solid-Liquid Interfaces. *Physical Review Letters*, 135(14):146201, September 2025.
- [150] Letian Chen, Yun Tian, Xu Hu, Suyu Chen, Huijuan Wang, Xu Zhang, and Zhen Zhou. A constant potential reactor framework for electrochemical reaction simulations, 2024.
- [151] Masnun Naher, Miguel A. González, Craig M. Williams, and Paul V. Bernhardt. Emerging applications and mechanistic insights of copper mediated electrocatalysts in organic transformations. *Chemical Society Reviews*, 54(15):7304–7338, 2025.
- [152] Thomas E. Gartner, Linfeng Zhang, Pablo M. Piaggi, Roberto Car, Athanasios Z. Panagiotopoulos, and Pablo G. Debenedetti. Signatures of a liquid–liquid transition in an ab initio deep neural network model for water. *Proceedings of the National Academy of Sciences*, 117(42):26040–26046, October 2020.
- [153] William M. Haynes, editor. *CRC Handbook of Chemistry and Physics*, 93rd Edition. CRC Press, April 2016.
- [154] Christopher D Taylor, Robert G Kelly, and Matthew Neurock. First-Principles Prediction of Equilibrium Potentials for Water Activation by a Series of Metals. *Journal of The Electrochemical Society*, 2007.
- [155] Chunyi Zhang, Marcos F. Calegari Andrade, Zachary K. Goldsmith, Abhinav S. Raman, Yifan Li, Pablo M. Piaggi, Xifan Wu, Roberto Car, and Annabella Selloni. Molecular-scale insights into the electrical double layer at oxide-electrolyte interfaces. *Nature Communications*, 15(1):10270, November 2024.

- [156] Chunyi Zhang, Fujie Tang, Mohan Chen, Linfeng Zhang, Diana Y. Qiu, John P. Perdew, Michael L. Klein, and Xifan Wu. Modeling liquid water by climbing up Jacob's ladder in density functional theory facilitated by using deep neural network potentials. *arXiv*, April 2021. arXiv:2104.14410.
- [157] Alenka Luzar and David Chandler. Hydrogen-bond kinetics in liquid water. *Nature*, 379(6560):55–57, January 1996.
- [158] A. Michaelides. Density functional theory simulations of water–metal interfaces: waltzing waters, a novel 2D ice phase, and more. *Applied Physics A*, 85(4):415–425, October 2006.
- [159] Jiabo Le, Qiyuan Fan, Laura Perez-Martinez, Angel Cuesta, and Jun Cheng. Theoretical insight into the vibrational spectra of metal–water interfaces from density functional theory based molecular dynamics. *Physical Chemistry Chemical Physics*, 20(17):11554–11558, 2018.
- [160] Sha Jin, Xue Fan, Caleb Stamper, Richard A. Mole, Yuanxi Yu, Liang Hong, Dehong Yu, and Matteo Baggioli. On the temperature dependence of the density of states of liquids at low energies. *Scientific Reports*, 14(1):18805, August 2024. arXiv:2304.14609.
- [161] Tobias Morawietz, Ondrej Marsalek, Shannon R. Pattenau, Louis M. Streacker, Dor Ben-Amotz, and Thomas E. Markland. The Interplay of Structure and Dynamics in the Raman Spectrum of Liquid Water over the Full Frequency and Temperature Range. *The Journal of Physical Chemistry Letters*, 9(4):851–857, February 2018.
- [162] Deepak Ojha, Kristof Karhan, and Thomas D. Kühne. On the Hydrogen Bond Strength and Vibrational Spectroscopy of Liquid Water. *Scientific Reports*, 8(1):16888, November 2018.
- [163] Yujin Tong, Tobias Kampfrath, and R. Kramer Campen. Experimentally probing the libration of interfacial water: the rotational potential of water is stiffer at the air/water interface than in bulk liquid. *Physical Chemistry Chemical Physics*, 18(27):18424–18430, 2016.
- [164] José A. Martinez-Gonzalez, Niall J. English, and Aoife A. Gowen. Understanding the interface between silicon-based materials and water: Molecular-dynamics exploration of infrared spectra. *AIP Advances*, 7(11):115105, November 2017.
- [165] Masatoshi Osawa, Minoru Tsushima, Hirokazu Mogami, Gabor Samjeské, and Akira Yamakata. Structure of Water at the Electrified Platinum–Water Interface: A Study by Surface-Enhanced Infrared Absorption Spectroscopy. *The Journal of Physical Chemistry C*, 112(11):4248–4256, March 2008.
- [166] Olle Björneholm, Martin H. Hansen, Andrew Hodgson, Li-Min Liu, David T. Limmer, Angelos Michaelides, Philipp Pedevilla, Jan Rossmeisl, Huaze Shen, Gabriele Tocci, Eric Tyrode, Marie-Madeleine Walz, Josephina Werner, and

- Hendrik Bluhm. Water at Interfaces. *Chemical Reviews*, 116(13):7698–7726, July 2016.
- [167] Paula Sebastián-Pascual, Francisco J. Sarabia, Víctor Climent, Juan M. Feliu, and María Escudero-Escribano. Elucidating the structure of the cu-alkaline electrochemical interface with the laser-induced temperature jump method. *The Journal of Physical Chemistry C*, 124(42):23253–23259, 2020.
- [168] J. Lecoœur and J.P. Bellier. Potentiels de charge nulle d’électrodes monocristallines de cuivre d’orientation (111) et (100) au contact de solutions aqueuses de perchlorate de potassium. *Electrochimica Acta*, 30(8):1027–1033, 1985.
- [169] Peng Li, Jun Huang, Youcheng Hu, and Shengli Chen. Establishment of the potential of zero charge of metals in aqueous solutions: Different faces of water revealed by ab initio molecular dynamics simulations. *The Journal of Physical Chemistry C*, 125(7):3972–3979, 2021.
- [170] Thomas A. Manz and Nidia Gabaldon Limas. Introducing ddec6 atomic population analysis: part 1. charge partitioning theory and methodology. *RSC Adv.*, 6:47771–47801, 2016.
- [171] Thomas A. Manz and David S. Sholl. Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Nonperiodic Materials. *Journal of Chemical Theory and Computation*, 6(8):2455–2468, August 2010.
- [172] R. F. W. Bader. Atoms in molecules. *Accounts of Chemical Research*, 18(1):9–15, January 1985. Publisher: American Chemical Society.
- [173] Graeme Henkelman, Andri Arnaldsson, and Hannes Jónsson. A fast and robust algorithm for bader decomposition of charge density. *Computational Materials Science*, 36(3):354–360, 2006.
- [174] Lang Li, Yun-Pei Liu, Jia-Bo Le, and Jun Cheng. Unraveling molecular structures and ion effects of electric double layers at metal water interfaces. *Cell Reports Physical Science*, 3(2):100759, February 2022.
- [175] D. F. Coker, R. E. Miller, and R. O. Watts. The infrared predissociation spectra of water clusters. *The Journal of Chemical Physics*, 82(8):3554–3562, 04 1985.
- [176] Ahmed T. Ayoub, Jack Tuszynski, and Mariusz Klobukowski. Estimating hydrogen bond energies: comparison of methods. *Theoretical Chemistry Accounts*, 133(8):1520, August 2014.
- [177] Luis Rincón, Rafael Almeida, David Garcia-Aldea, and Helena Diez Y Riega. Hydrogen bond cooperativity and electron delocalization in hydrogen fluoride clusters. *The Journal of Chemical Physics*, 114(13):5552–5561, April 2001.

- [178] A. Otero-de-la Roza, Erin R. Johnson, and Víctor Luaña. Critic2: A program for real-space analysis of quantum chemical interactions in solids. *Computer Physics Communications*, 185(3):1007–1018, March 2014.
- [179] Sergio Trasatti. Work function, electronegativity, and electrochemical behaviour of metals: Iii. electrolytic hydrogen evolution in acid solutions. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 39(1):163–184, 1972.
- [180] Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014.
- [181] Kareem M. Gameel, Icell M. Sharafeldin, Amr U. Abourayya, Ahmed H. Biby, and Nageh K. Allam. Unveiling CO adsorption on Cu surfaces: new insights from molecular orbital principles. *Physical Chemistry Chemical Physics*, 20(40):25892–25900, 2018.
- [182] Scott K. Shaw, Antonio Berná, Juan Miguel Feliu, Richard J. Nichols, Timo Jacob, and David J. Schiffrin. Role of axially coordinated surface sites for electrochemically controlled carbon monoxide adsorption on single crystal copper electrodes. *Physical Chemistry Chemical Physics*, 13(12):5242, 2011.
- [183] Joseph H. Montoya, Chuan Shi, Karen Chan, and Jens K. Nørskov. Theoretical insights into a co dimerization mechanism in co₂ electroreduction. *The Journal of Physical Chemistry Letters*, 6(11):2032–2037, 2015. PMID: 26266498.
- [184] Robert B. Sandberg, Joseph H. Montoya, Karen Chan, and Jens K. Nørskov. Co-co coupling on cu facets: Coverage, strain and field effects. *Surface Science*, 654:56–62, 2016.
- [185] Wei Wang, Mattia Salomone, Michele Re Fiorentin, Francesca Risplendi, and Giancarlo Cicero. Activating c–c coupling on copper during co₂rr: Charge-controlled design of alloy catalysts. *ACS Electrochemistry*, 1(11):2512–2520, 2025.
- [186] Georg Kastlunger, Lei Wang, Nitish Govindarajan, Hendrik H. Heenen, Stefan Ringe, Thomas Jaramillo, Christopher Hahn, and Karen Chan. Using ph dependence to understand mechanisms in electrochemical co reduction. *ACS Catalysis*, 12(8):4344–4357, 2022.
- [187] Kumudu Mudiyanse, Yixiong Yang, Friedrich M. Hoffmann, Octavio J. Furlong, Jan Hrbek, Michael G. White, Ping Liu, and Darío J. Stacchiola. Adsorption of hydrogen on the surface and sub-surface of cu(111). *The Journal of Chemical Physics*, 139(4):044712, 07 2013.
- [188] G Lee and E.W Plummer. High-resolution electron energy loss spectroscopy study on chemisorption of hydrogen on cu(111). *Surface Science*, 498(3):229–236, 2002.

- [189] Lang Xu, Joshua Lin, Yunhai Bai, and Manos Mavrikakis. Atomic and Molecular Adsorption on Cu(111). *Topics in Catalysis*, 61(9-11):736–750, June 2018.
- [190] Zhijie Chen, Xiaoguang Duan, Wei Wei, Shaobin Wang, and Bing-Jie Ni. Recent advances in transition metal-based electrocatalysts for alkaline hydrogen evolution. *Journal of Materials Chemistry A*, 7(25):14971–15005, 2019.
- [191] S. Trasatti. The absolute electrode potential: an explanatory note (recommendations 1986). *Pure and Applied Chemistry*, 58(7):955–966, 1986.
- [192] Jean-Sébastien Filhol and Matthew Neurock. Elucidation of the electrochemical activation of water over pd by first principles. *Angewandte Chemie International Edition*, 45(3):402–406, 2006.
- [193] D. Cao, G.-Q. Lu, A. Wieckowski, S. A. Wasileski, and M. Neurock. Mechanisms of Methanol Decomposition on Platinum: A Combined Experimental and ab Initio Approach. *The Journal of Physical Chemistry B*, 109(23):11622–11633, June 2005.
- [194] Per Lindgren, Georg Kastlunger, and Andrew A. Peterson. Electrochemistry from the atomic scale, in the electronically grand-canonical ensemble. *The Journal of Chemical Physics*, 157(18):180902, 2022.
- [195] Ilyes Batatia, Chen Lin, Joseph Hart, Elliott Kassoar, Alin M. Elena, Sam Walton Norwood, Thomas Wolf, and Gábor Csányi. Cross learning between electronic structure theories for unifying molecular, surface, and inorganic crystal foundation force fields, 2025.
- [196] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Physical Review Letters*, 104(13):136403, April 2010.
- [197] Volker L. Deringer and Gábor Csányi. Machine learning based interatomic potential for amorphous carbon. *Physical Review B*, 95(9):094203, March 2017.
- [198] Mianle Xu, Sihang Liu, Sudarshan Vijay, Thomas Bligaard, and Georg Kastlunger. Benchmarking water adsorption on metal surfaces with ab-initio molecular dynamics. *ChemRxiv*, February 2024.
- [199] B. Hammer, L. B. Hansen, and J. K. Nørskov. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Phys. Rev. B*, 59:7413–7421, Mar 1999.
- [200] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge 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(15):154104, 04 2010.

- [201] Haowei Peng, Zeng-Hui Yang, John P. Perdew, and Jianwei Sun. Versatile van der waals density functional based on a meta-generalized gradient approximation. *Phys. Rev. X*, 6:041005, Oct 2016.
- [202] Roger Nadler and Javier Fernandez Sanz. First-principles molecular dynamics simulations of the H₂O / Cu(111) interface. *Journal of Molecular Modeling*, 18(6):2433–2442, June 2012.
- [203] Katrin Tonigold and Axel Groß. Dispersive interactions in water bilayers at metallic surfaces: A comparison of the PBE and RPBE functional including semiempirical dispersion corrections. *Journal of Computational Chemistry*, 33(6):695–701, March 2012.
- [204] Kaisheng Song and Jun Li. The neural network based δ -machine learning approach efficiently brings the dft potential energy surface to the ccSD(t) quality: a case for the OH + CH₃OH reaction. *Phys. Chem. Chem. Phys.*, 25:11192–11204, 2023.

Appendix A

Double Reference Method

A central element of this thesis, differing from other ML-FF works in electrochemistry, is the electrification scheme used to reproduce the effects of a constant applied bias in the dataset. Here we follow the “double reference method” (DRM) [66]. In any constant potential scheme, the charge on the electrode is allowed to vary such that the whole system is at equilibrium with a target applied potential [46]. Two main ingredients are required to achieve a constant potential calculation: 1) a scheme allowing to finely and continuously tune the surface charge of the electrode and 2) a reliable procedure to establish a connection between the surface charge and the applied potential. The DRM implements these ingredients through 1) a homogeneous background method (HBM) and 2) a chain of potential references involving auxiliary systems to rigorously define the electrode potential.

A varying surface charge (i.e., an electronically grand canonical ensemble) can be easily achieved in any electronic structure code by adding/removing a fractional number of electrons to the metal and the corresponding electronic charge to the system [17]. Since DFT codes commonly rely on periodic boundary conditions, there is a need for an approach to avoid divergences in the electrostatic energy and guarantee the charge neutrality when extra charge is added/removed to the electrode. The DRM follows the most straightforward approach of adding a homogeneous background countercharge (a.k.a. jellium) compensating for the extra charge of the electrode. Additional considerations on the HBM are reported in Section A.1.

In addition to the control on the surface charge, a constant potential scheme requires the definition of the potential associated with the electrode. The electrode potential

U_{SHE} (referred to the Standard Hydrogen Electrode SHE) is computed as:

$$U_{SHE} = (\Phi - \Phi_{SHE})/e, \quad (\text{A.1})$$

where Φ is the work function of the system and it depends on the surface charge of the electrode, while Φ_{SHE} is the potential of SHE with respect to a vacuum (commonly assumed equal to 4.44 V) [17, 191]. The DRM defines the work function and establishes its connection with the extra surface charge via a robust series of potential references involving two auxiliary systems such that this definition is not influenced by the counter-charge and any fictitious solvent-vacuum interface. Specifically, the first auxiliary system has no extra charge (i.e., no applied potential) and a vacuum region to define the vacuum level; the second is with no extra charge and no vacuum and represents a reference for the same system but with extra charge (i.e., with applied potential). In this manner, any inconsistency in the energy reference between neutral and charged systems, as well as the spurious contributions to the work function arising from the fictitious solvent–vacuum interfaces, is rigorously accounted for. More details on the series of potential references are reported in Section A.2.

The effectiveness of the DRM in conjunction with the HBM in the description of electrified metal/water interfaces and electrochemical reactions has been extensively proved in several articles [66, 192, 193], justifying its use to reproduce the effects of the applied potential in our dataset for the ML. Moreover, the DRM can be readily implemented in most DFT codes, as these typically allow specifying a fractional charge for the system and automatically compensate it with a homogeneous counter-charge. The potential references required by DRM can also be easily extracted from standard DFT output files. Unlike other constant-potential approaches (e.g., implicit-solvent-based schemes), DRM does not require any specific external plug-ins that might clash with the parallelisation capabilities of the DFT codes on modern HPC architectures. However, managing the chain of potential references and continuously updating the electrode charge in a constant-potential scheme can become tedious and time-consuming, especially in high-throughput workflows such as ML dataset generation. To address this, we developed a *Python* calculator that automates the entire DRM process within an electronically grand-canonical ensemble. Our implementation builds on the fully converged constant-potential (FCP) ASE calculator proposed in Ref. [68], extending it to define the electrode potential according to DRM. The calculator is integrated into a workflow that automatically evaluates

DRM auxiliary systems and performs constant-potential calculations at fixed geometries for multiple target potentials. The DRM calculator is freely available at <https://github.com/michelegiovannibianchi/DoubleReferenceMethod-FCP-calculator>. An excerpt of the core code lines is provided in Section A.3.

A.1 Considerations on the homogeneous background method

The homogeneous background method (HBM) [66, 71] is employed in the DRM to compensate for the extra electron charge added/removed to the system. Despite the non-physical spreading of the "jellium" charge over the whole system, the HBM can be interpreted as 1) a model for the ionic counter-charge in the region of the electrolyte and 2) an entirely delocalised dopant model in the electrode region [71]. Regarding assertion 1), it has been shown that the jellium, in the presence of the screening action of explicit water molecules, is able to reproduce the effects of the electrode polarisation and mimic the presence of ions at the outer Helmholtz plane [66]. Regarding assertion 2), it has been proved that this diluted dopant has negligible effects on the capacitance of a metal electrode. Indeed, a uniform countercharge and a countercharge distributed according to the solution of the linearised Poisson-Boltzmann equation (as implemented in the Polarizable Continuum Model PCM of *VASPsol*), show equivalent results [71].

In the framework of the HBM, the total grand canonical free energy Ω is evaluated as [66]:

$$\Omega = E_{SCF} + \Delta q \mu_e + \int_0^{\Delta q} \langle V_{tot} \rangle dQ, \quad (\text{A.2})$$

where E_{SCF} is the self-consistent energy, as computed in the DFT code, Δq is the extra charge added/removed to the system, μ_e is the electrochemical potential of the electron system and is commonly evaluated as the negative of the work function [194]. $\langle V_{tot} \rangle$ is the volume-averaged electrostatic potential within the simulation cell. As observed in ref. [71], $\langle V_{tot} \rangle$ is automatically set to zero in *VASP*, independently of the amount of extra charge. Hence, the corrections to the grand canonical energy due to the homogeneous countercharge (i.e., the third term in equation A.2) are easily included.

A.2 Considerations on the chain of potential references in the DRM

The electrode potential is often assumed to be proportional to the work function (see Equation A.1), defined as the energy difference between the vacuum level and the Fermi level. However, evaluating the work function becomes problematic when an external potential is applied. First, defining the vacuum level requires a vacuum region far from the electrode surface, introducing an artificial solvent/vacuum interface. Second, adding extra charge and its compensating background countercharge shifts the energy scale of the system, making the Fermi level incomparable to other energy references from a neutral system. The double reference method resolves these issues by introducing two auxiliary systems (Figure A.1):

- a neutral system ($q = 0$) with a vacuum region,
- a neutral system ($q = 0$) without vacuum.

These references allow a consistent definition of the work function for the charged system ($q \neq 0$, applied potential) without vacuum. In the first auxiliary system, the vacuum level $\phi'(v)$ is referenced to the potential inside the metal slab $\phi'(m)$, assumed unaffected by the solvent–vacuum interface. For that reason, this $\phi'(m)$ can be aligned with the corresponding potential $\phi'_0(m)$ in the middle of the slab in the second system (no vacuum). Within the second system, $\phi'_0(m)$ is then referenced to the potential $\phi'_0(w)$ in the middle of the solvent region. Consequently, the potential in the water region, far from the surface, with respect to the vacuum level is:

$$\phi_0(w) = \phi'_0(w) - \phi'_0(m) + \phi'(m) - \phi'(v). \quad (\text{A.3})$$

Assuming the potential in the solvent far from the interface is unaffected by extra charge, the solvent potentials in the neutral and charged systems ($\phi_0(w)$ and $\phi'_q(w)$) are aligned. Finally, the Fermi level $\phi'_q(f)$ of the charged system is referenced to $\phi'_q(w)$ and corrected using $\phi_0(w)$:

$$\phi_q(f) = \phi'_q(f) - \phi'_q(w) + \phi_0(w). \quad (\text{A.4})$$

This chain of references enables a rigorous evaluation of the Fermi level $\phi_q(f)$ for a charged, vacuum-free system relative to the vacuum level of a neutral system. The

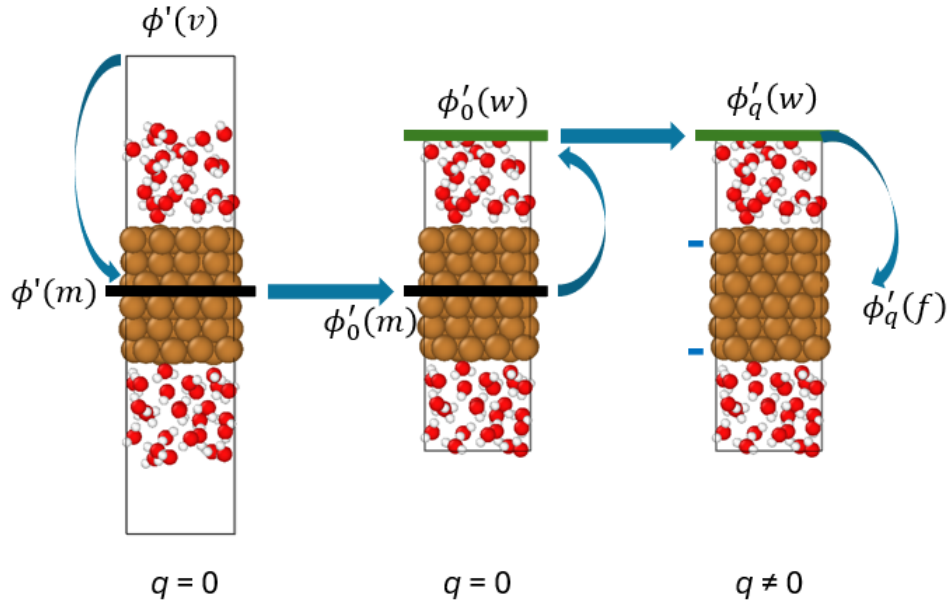


Fig. A.1 Scheme reproducing the systems and the sequence of energy references used to define the work function according to the DRM. From left to right, the system with a vacuum region and no extra charge, without a vacuum and no extra charge, without a vacuum and with extra charge. Systems with symmetric metal/water interfaces are required in the DRM to avoid vacuum regions in the final system.

resulting value is then used in Equation A.1 as the inverse of the work function Φ , properly defining the electrode potential of the system.

A.3 Code for the fully converged constant potential calculator via DRM

We report the main extracts of the developed *Python* code for the fully converged constant potential (FCP) that relies on the DRM definition for the electrode potential. The variable names reflect the convention used in the equations above. The full code and more info can be found at <https://github.com/michelegiovannibianchi/DoubleReferenceMethod-FCP-calculator>.

The calculator is initialized and called via:

```
## Start calculation of the "Double Reference Method" workflow ##
# Arguments for the Double Reference Workflow
```

```
# - snap: ase atoms, atomic geometry
# - external_bias_vector: list, values of applied potential for which
    the DRM will be applied
# - calc_neutral_no_vacuum: ase calculator, calculator for system
    without a vacuum region and no extra charge
# - calc_neutral_vacuum_no_dipole: ase calculator, calculator for
    system with a vacuum region, but not dipole corrections (
    recommended to get convergence when dipole corrections will be
    switched on)
# - calc_neutral_vacuum_dipole: ase calculator, calculator for system
    with a vacuum region, and with dipole corrections
# - calc_charge: ase calculator, calculator for system without a
    vacuum region and extra charge
# - guess_extra_electrons: int, initial guess of extra electrons to
    add to the system for the first entry value of the
    external_bias_vector
# - C_guess: float, initial guess for the capacitance of the interface.

# - restart: bool, if True the calculation will start from the
    calculation of the system with a vacuum region, and with dipole
    corrections
# - V_SHE: float, value of the standard hydrogen electrode potential

snap=read('POSCAR',format='vasp')

# Prepare arguments for your workflow
workflow_args = dict( external_bias_vector=[-0.5, -0.75, -1.0],
                      calc_neutral_no_vacuum=calc_neutral_no_vacuum,
                      calc_neutral_vacuum_no_dipole=
                          calc_neutral_vacuum_no_dipole,
                      calc_neutral_vacuum_dipole=
                          calc_neutral_vacuum_dipole,
                      calc_charge=calc_charge,
                      guess_extra_electrons=1,
                      C_guess=0.008856,
                      restart=False )
```

```
# Attach evaluator
snap.evaluator = DoubleReferenceEvaluator(DoubleReferenceWorkflow, **
    workflow_args)

# Trigger calculation
result = snap.evaluator.evaluate(snap)
```

In the last lines above, a wrapper for the DoubleReferenceWorkflow function is called. The core lines of this function are:

```
def DoubleReferenceWorkflow(snap,external_bias_vector,
    calc_neutral_no_vacuum,calc_neutral_vacuum_no_dipole,
    calc_neutral_vacuum_dipole,calc_charge,guess_extra_electrons,
    C_guess,restart,V_SHE=None):

    ### Start calculation of the "Double Reference Method" workflow ####
    ...
    element=snap.get_chemical_symbols()
    n_O=element.count('O')
    case_dir=os.getcwd()

    #### 1) System no extra charge + no vacuum
    snap_neutral=snap

    #Call VASP calculator
    snap_neutral.calc=calc_neutral_no_vacuum
    if restart==False:
        print('Neutral: Etot=',snap_neutral.get_total_energy())

    #Average potential with MacroDensity
    input_file = 'neutral/LOCPOT'
    output_file = 'neutral/Neutral_no_vac_planar.dat'
    pot=average_potential(input_file,output_file)

    # Potential in the middle of the slab
    phi_prime_0_m=min(pot)
    # Potential in the middle of the water region
    phi_prime_0_w=(pot[0]+pot[-1])/2
```

```

#Read Fermi level and electron number from OUTCAR
path_OUTCAR='neutral/OUTCAR'
phi_prime_0_f,nelect0=read_fermi_nelect(path_OUTCAR)

#### 2) System with no extra charge + vacuum

snap_vacuum=add_vacuum(snap,n_0)

##### 2.1) No extra charge + vacuum + no dipole corrections
# Call VASP calculator
snap_vacuum.calc=calc_neutral_vacuum_no_dipole
if restart==False:
    print('Neutral vacuum no dipole: Etot=',snap_vacuum.
          get_total_energy())

##### 2.1) No extra charge + vacuum + dipole corrections
# Call VASP calculator
snap_vacuum.calc=calc_neutral_vacuum_dipole
print('Neutral vacuum dipole: Etot=',snap_vacuum.get_total_energy
      ())

#Average potential with MacroDensity
input_file = 'neutral_vacuum/LOCPOT'
output_file = 'neutral_vacuum/Neutral_vac_dipole_planar.dat'
pot=average_potential(input_file,output_file)

# Potential in the middle of the slab
phi_prime_m_r=min(pot)# the minimum of the potential is in the
    fixed layers of the slab

#Read vacuum level from OUTCAR
path_OUTCAR='neutral_vacuum/OUTCAR'
phi_prime_v_r=read_vacuum_level(path_OUTCAR)

##### Compute PZC value via double reference method
#potential in the neutral cell in the middle of water, without a
    vacuum, referred to the system with a vacuum (Equation A.3)

```

```

phi_0_w=phi_prime_0_w-phi_prime_0_m+phi_prime_m_r-phi_prime_v_r

#Fermi level in the neutral cell without a vacuum, referred to the
    vacuum level (Equation A.4)
phi_0_f=phi_prime_0_f-phi_prime_0_w+phi_0_w

#PZC value referred to the SHE (Equation A.1)
U_0_SHE=-V_SHE-phi_0_f

#### 3) System with extra charge + no vacuum

#initialize vectors
nelect_vector=[]
U_vector=[]

#Store the point corresponding to no extra charge
nelect_vector.append(nelect0)
U_vector.append(U_0_SHE)

snap_charge=read('POSCAR',format='vasp')

for V,index_V in zip(external_bias_vector,range(0,len(
    external_bias_vector))):
    V_label=convert_V_to_label(V)

    if index_V==0: #For the first point start from the PZC
        informations

    # FCP calculator
    cal_FCP=FCP2rm(innercalc=calc_charge,fcptxt=f'log-fcp_U_{
        V_label}.txt',U=V,NELECT =nelect0+guess_extra_electrons,
        NELECT0=nelect0,work_ref=V_SHE,C=C_guess, phi_0_w=
        phi_0_w,max_FCP_iter=20)
    snap_charge.calc=cal_FCP
    print(snap_charge.get_total_energy())

    else: #For the other points start from the info of the previous
        points

```



```

#read the last point to calculate a new guess of the
#capacitance and the number of electrons
nelect_last, U_last, C_cal =read_log_fcp(f'log-fcp_U_{
    convert_V_to_label(external_bias_vector[index_V-1])}.txt
    ')
nelect_vector.append(nelect_last)
U_vector.append(U_last)

#Linear fit to estimate the number of electrons to add
coef=np.polyfit(np.array(U_vector),np.array(nelect_vector),
    1)

#start the FCP calculation
cal_FCP=FCP2rm(innercalc=calc_charge,fcptxt=f'log-fcp_U_{
    V_label}.txt',U=V,NELECT =np.polyval(coef,V), C=C_cal,
    NELECT0=nelect0,work_ref=V_SHE,phi_0_w=phi_0_w,
    max_FCP_iter=20)
snap_charge.calc=cal_FCP
print(snap_charge.get_total_energy())

print("Calculation terminated")
return

```

The DoubleReferenceWorkflow function calls the FCP calculator FCP2rm, adapted from ref. [68], whose core lines defining the electrode potential following the DRM are:

```

class FCP2rm(FileIOCalculator):
    """
        Fully constant potential calculator for electrochemistry
        implementing the "Double Reference Method" by Taylor et al.,

        Phys. Rev. B 73, 165402 (DOI: https://doi.org/10.1103/PhysRevB.73.165402)

        This FCP calculator is adapted from https://github.com/hellozhaoming/FCP-vasp-ase
    """

```

```

...
def calculate(...):
    ...
    self.fermi=atomstmp.calc.get_fermi_level()
    # Equation A.1 + A.4
    Ucal=-(self.fermi + self.fermishift)-self.workSHE
    # Equation A.2
    grand_energy_free=energy_free+(self.wf+self.fermishift)*(self.
        Nselect-self.Nselect0)
    ...
    def read_fermishift_2rm(self,outhpath, lines=None,phi_0_w=None)
        :
        """Method that evaluate Fermi shift within the "Double
            Reference Method" """

        ## Average over potential. Use MacroDensity code
        input_file = outhpath+'LOCPOT'
        output_file = outhpath+'Charge_planar.dat'
        pot=average_potential(input_file,output_file)

        phi_prime_q_w=(pot[0]+pot[-1])/2# average potentential in
            the middle of the water region
        # Equation A.4
        E_fs = -phi_prime_q_w+phi_0_w #Fermi shift =-potential in
            vacuum
        return E_fs
    ...

```

Appendix B

Franken descriptors: characteristics, accuracy and usage

In this appendix, we report additional considerations on the importance of identifying the most suitable descriptors (i.e., Franken backbones) for an effective transfer learning approach in TRECI workflow and the resulting accuracy achievable with the different backbones. We first recap the characteristics of the different Franken models and of their associated backbones, reporting a table that can help readers to understand the use of these ML-FFs within the TRECI framework. We then report the results associated with the ML-FFs developed in Chapter 5 to describe the Cu interfaces without reaction (Section B.2), followed by the discussion of models of Chapter 6 where the reactivity is included in the picture (Section B.3).

B.1 Comparison of the different Franken models

Table B.1 provides a comparative overview of the Franken models employed in this thesis, together with their principal characteristics. Since model performance in terms of computational cost, stability, and accuracy depends on a broad range of factors, such as dataset size, number of random features, and is system-dependent, we report only a qualitative ranking (High–Middle–Low). This classification is not intended to be interpreted in absolute terms, but rather as a guideline to support the selection of the most appropriate model for each specific circumstance. Domain-specific backbone models in Chapter 5 are trained on a dataset describing the various

Table B.1 Comparative table of the different Franken models employed in this thesis with information on the pre-trained model over which the backbones are built, their performance and their typical use in the different stages of active learning (AL) or final MD production runs.

Franken	Type	Backbone	Cost	Stability	Accuracy	Use
MP0	General purpose	MACE-MP-0	Low	High	Low	Early stage AL
MH0	General purpose	MACE-MH-0	High	High	Middle	All cycles of AL
PZC-L1	Domain specific	Equivariant model for Cu/H ₂ O at PZC	High	Middle	High	Early / middle stage AL
PZC+Charged-L0	Domain specific	Invariant model for Cu/H ₂ O trained on geometries at PZC and at applied V	Low	Middle/low	High	Middle / late stage AL and production run
PZC-L0	Domain specific	Invariant model for Cu/H ₂ O at PZC	Low	Low	High	Production run

Cu/water interfaces without reactive configurations, whereas the PZC dataset for models in Chapter 6 is further augmented with configurations relevant to the reaction processes.

B.2 Models for the interface without reactive events

B.2.1 Comparative test on the descriptors and the number of random features

Figure B.1 shows a comparison of the accuracy of the Franken models based on MACE-MP0 (i.e., Franken-MP0) and on the domain-specific PZC descriptors (i.e., Franken-PZC) developed to study the Cu (111) surface without reactive events. Accu-

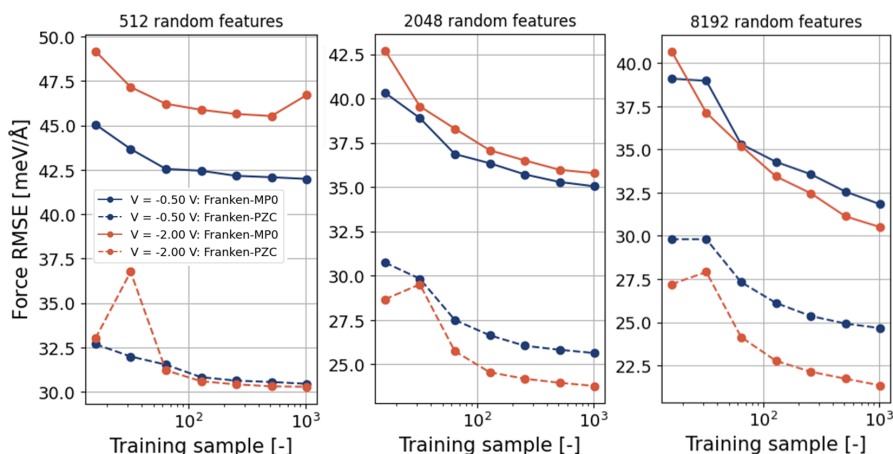


Fig. B.1 Comparison of the RMSE error on force prediction at $U = -0.50$ V and -2.00 vs SHE, using different descriptors (Franken-MP0 vs Franken-PZC) for different training set sizes and numbers of random features.

racy is evaluated on the force RMSE on a validation set of about 150 configurations. This analysis is performed by varying the size of the training set (from 16 up to 1024 configurations) and the number of the random features used to approximate the kernel function. We evaluate the cases where U is -0.50 and -2.00 V vs SHE, as these represent the nearest and most distant conditions for the descriptors trained on the PZC geometries. It is evident that the domain-specific PZC descriptors are superior to the general-purpose MACE-MP0 ones at -0.50 V as well as at -2.00 V, regardless

of the training set size and the number of random features. This test further proves the importance of the domain-specific descriptors in the TRECI strategy.

B.2.2 Comparative test of PZC and constant potential models

This section reports a comparison of the accuracy of the PZC model trained from scratch with respect to the transfer learning model for the constant potential cases for the Cu surfaces without reactive events. The results for the PZC are obtained with a MACE model with invariant features and 256 channels, while the applied bias cases are obtained via Franken transfer learning using the Franken-PZC-L0 backbone (invariant features and 256 channels) and 8192 random features. Figure B.2 reports the distributions of the error in the forces over the test sets for the different cases (PZC and constant potentials). The error is reported together with the modulus of the forces. The contributions for the different species are reported separately in the different panels. It is evident that the accuracy is similar for all the constant potential cases, and it is comparable with the accuracy of the PZC MACE model trained from scratch but with a larger amount of data, proving the effectiveness of the Franken transfer learning strategy as implemented in the TRECI workflow.

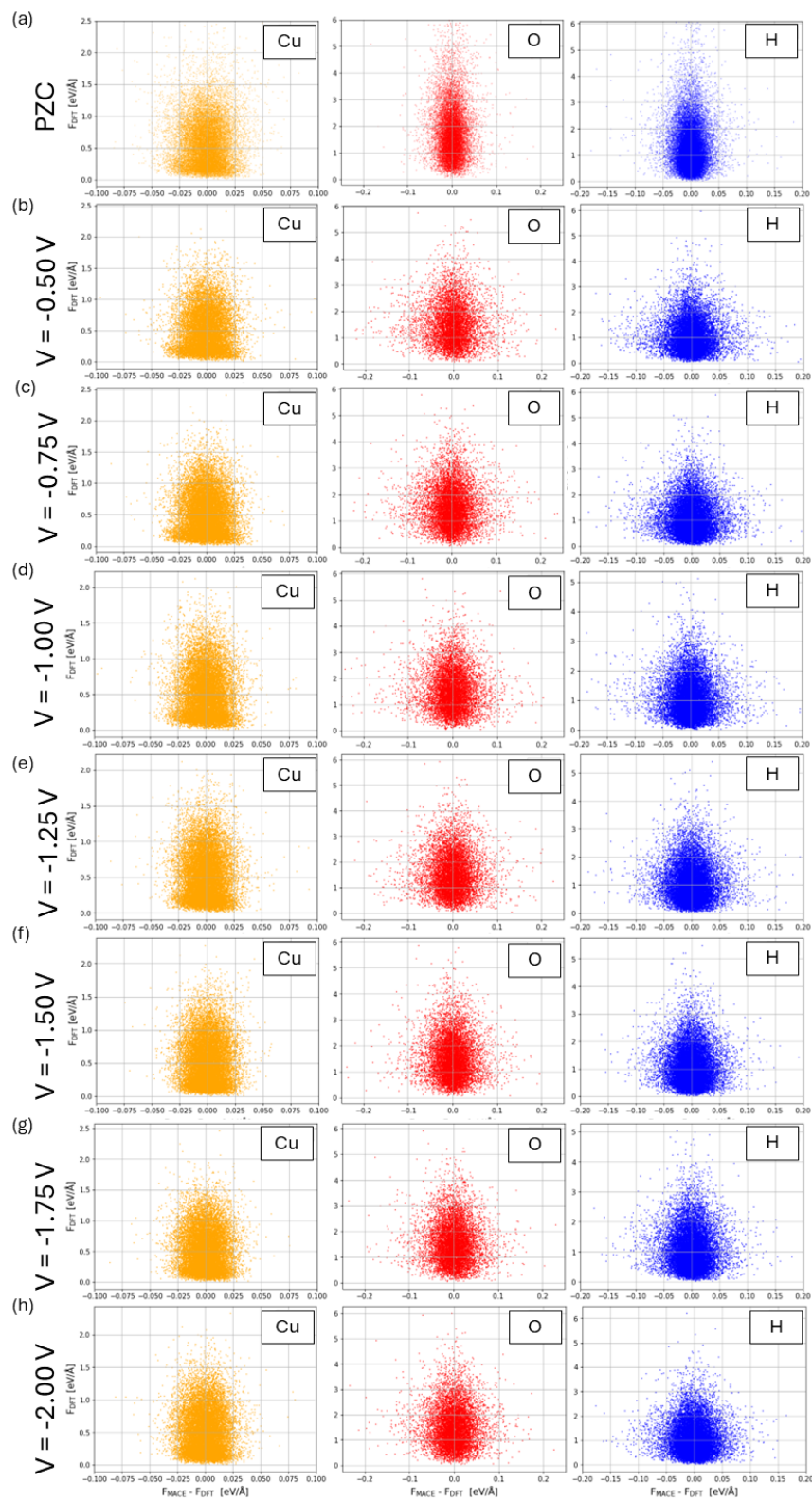


Fig. B.2 Scatter plots for the error in the predictions of the forces vs the force modulus at (a) PZC, (b) -0.50, (c) -0.75, (d) -1.00, (e) -1.25, (f) -1.50, (g) -1.75 and (h) -2.00 V vs SHE. The contributions of the different chemical species are reported separately.

B.3 Models for the reactivity of the interfaces

In this section we discuss the importance of considering additional backbones to achieve a more effective, accurate and data-efficient transfer learnings for models targeting reactions. While the general-purpose backbone based on MACE-MP0 (Franken-MP0) still grants sufficient stability to the MD trajectories also in the presence of reactive events, these descriptors are not the most suitable to describe molecular processes. Indeed, MACE-MP0 is trained on a large Materials Project database mainly composed of small periodic unit cells describing inorganic crystals with few molecular components. Consequently, processes involving proton transfer (e.g., the Volmer step in the HER) or the formation of carbon chains (e.g., the CO dimerisation in the CO₂RR) might be far from the domain on which MACE-MP0 models are trained, resulting in less effective transfer learning for the reactions considered here. On the contrary, the domain-specific descriptors based on the PZC dataset (Franken-PZC), expanded with the reactive configurations collected at the PZC for our MACE model trained from scratch, still stand as the most accurate backbones. Nonetheless, the stability issues for the domain-specific route are more severe when Franken models are employed to drive MD simulations coupled with enhanced sampling techniques required for the study of the reactions. The application of the bias potential associated with the enhanced sampling techniques might push the ML models in different basins of the PES and/or through different transition states (TSs) with respect to the PZC case, challenging the generalisation capability of the domain-specific descriptors.

To better fill the gap between the stable Franken-MP0 and the accurate Franken-PZC, we evaluate other universal MACE-based models better suited for molecular systems to use as general-purpose backbones. Specifically, we draw attention to the MACE Multi-Head universal model family (MACE-MH), whose different heads (i.e., read-out blocks) have been trained on diverse databases, each one tailored to address some specific classes of materials, such as inorganic crystals, organic molecules, metal surface slabs with adsorbate complexes and organic reaction intermediates and TSs [195]. The resulting descriptors of MACE-MH, embedding information from these diverse datasets, have superior cross-domain capability spanning inorganic crystals, molecular systems, surface chemistry, and reactive organic chemistry. Consequently, Franken models based on MACE-MH (Franken-MH) have a more effective general-purpose backbone to target the reaction steps selected in this chapter.

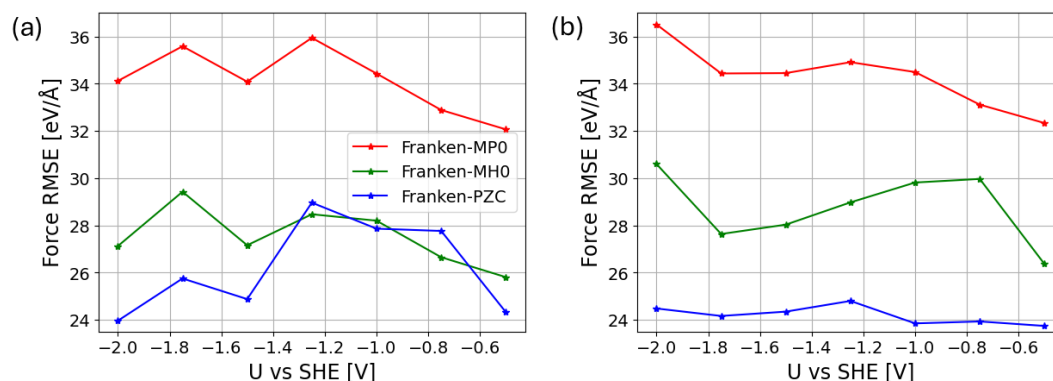


Fig. B.3 RMSE on the force prediction over the validation set vs the applied potential for Franken ML-FFs using different backbones for the model for (a) the Volmer step of HER, (b) the CO dimerisation step for the CO₂RR.

Figure B.3 compares the RMSE of the force prediction of the constant-potential ML models using the MACE-MH-0 descriptors (Franken-MH0) with respect to the Franken-MP0 and Franken-PZC for the selected reaction steps (Volmer for the HER in (a) and CO dimerisation in the CO₂RR in (b)). Franken-MH0 effectively bridges the accuracy gap between Franken-MP0 and Franken-PZC, yielding errors that are close to those of Franken-PZC. At the same time, it retains the robust MD stability characteristic of general-purpose models, producing trajectories comparable in stability to Franken-MP0 and markedly more stable than Franken-PZC. Nonetheless, this solution comes at the expense of a higher computational cost in training and inference due to the more complex architecture employed in MACE-MH-0 (roughly Franken-MH0 is 40÷50% slower in training and inference with respect to Franken-PZC). For these reasons, ongoing tests are finalised to optimise the use of these general-purpose and accurate backbone models during the active learning phase and/or in the final MD production.

Appendix C

Assessments of the modelling framework

In this ancillary section, we present a series of tests assessing both the reliability of the underlying DFT reference data and the performance of the ML models in capturing the key interactions that govern the physics of electrified metal/water interfaces. In particular, we examine the role of specific long-range force contributions that may significantly affect the solvent response to an applied potential (Section C.1), and we assess whether explicit van der Waals corrections to the SCAN exchange–correlation functional are required to accurately reproduce the balance between chemisorption and physisorption processes at the interface (Section C.2). Providing a solid justification for the assumptions made in Chapter 5, i.e., the exclusion of explicit long-range terms in the construction of the ML-FFs and the adoption of SCAN-based DFT results as ground truth, is a crucial step toward validating the reliability of the picture of the electrified metal/water interfaces reported in this thesis.

C.1 Locality of the ML models and long-range interactions

Electrochemical systems are a typical case in which electrostatic long-range interactions might have a significant role [46, 147, 155]. Since all the MPNN schemes employed in this thesis have two interaction layers, with a 6 Å cut-off radius each,

the nominal receptive field of the developed models extends up to 12 Å. However, considering the possible presence of non-local contributions outside this receptive field, we test models with cut-off radii up to 9 Å, as well as a model with a cut-off of 5 Å but with three interaction layers. We benchmark these models on accuracy and prediction of physical observables. Specifically, we perform these tests on the ML-FF for the PZC and for $U = -2.00$ V. As shown in Figure C.1a, there is a minimal

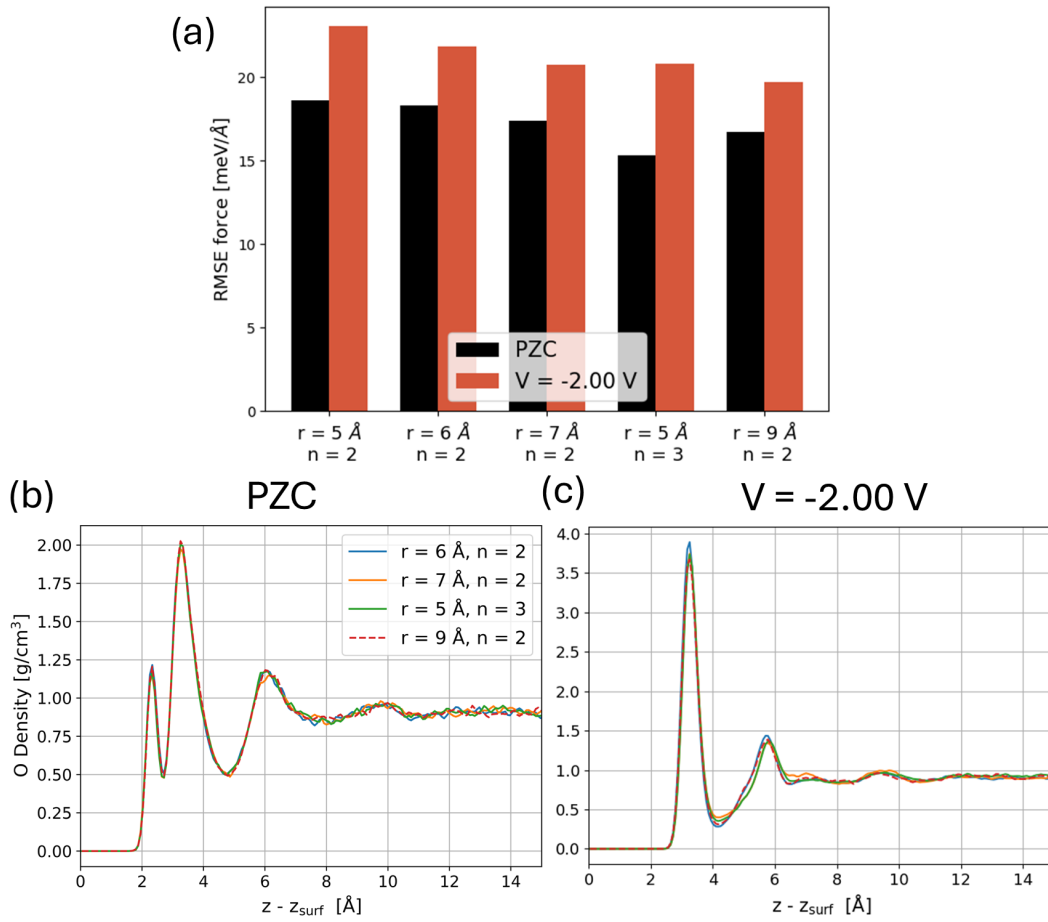


Fig. C.1 Comparison of the effects of different cut-off radii r and/or numbers n of interaction blocks on the accuracy of the models (a) and on the oxygen density profile for the Cu (111) surface at the PZC (b) and at $U = -2.00$ V (c).

improvement in the RMSE on the force prediction as the receptive field is expanded for the PZC as well as for $U = -2.00$ V. We also evaluate these effects on physical observables, such as the oxygen density profile (Figure C.1b-c). It is evident that the solvent density profile is independent of the receptive field of the ML model since the small variations between the models are within the statistical uncertainty, and a

constant density is recovered within 12 Å in all cases. This proves that the recovery of the water bulk properties far from the surface, also in the presence of additional surface charge, is associated with the screening of the water molecules rather than the local nature of the ML-FFs.

We further quantify the non-local contributions outside the receptive field of our models following the procedure proposed by Bartók and Csányi [196, 197]. Given a configuration and selecting an atom, we define a sphere centred on this atom with radius r_{fix} . Then we create multiple copies (i.e., > 20) of this system, keeping the atoms within the sphere frozen while introducing random displacements for all other atoms outside of the sphere. Specifically, the modulus of the displacement is sampled from a normal distribution with a standard deviation equal to 0.1 Å. For all these copies, we compute the forces acting on the central atoms at the DFT level. The locality of the interactions of the central atom is then characterised by plotting the standard deviation of these forces as a function of r_{fix} (see Figure C.2). Any ML model with a receptive field smaller than r_{fix} is not aware of the distortions and is not able to reproduce the contributions of atoms outside the sphere, i.e., these interactions result in long-range terms for the ML-FF. As visible in Figure C.2a, the

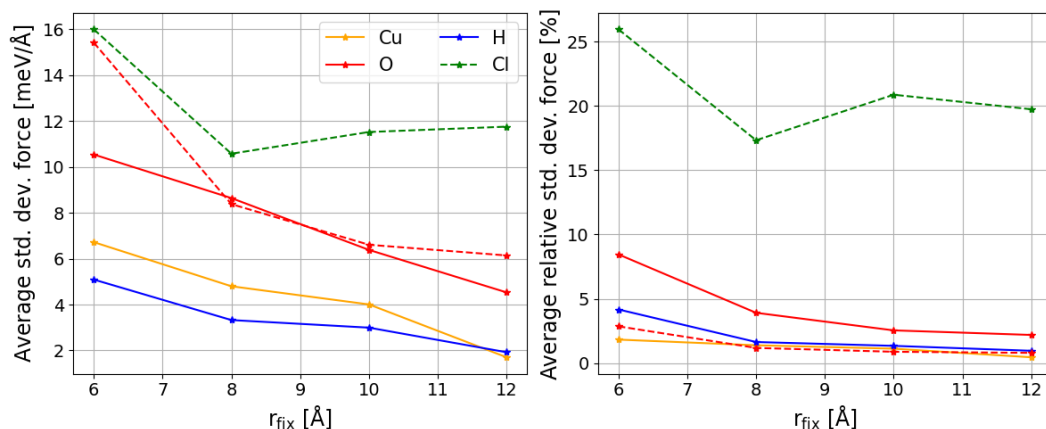


Fig. C.2 Standard deviation (a) and relative standard deviation (b) of forces computed at increasing r_{fix} . Results are reported for the different chemical species, averaging among the force components and among atoms at different distances from the surface. Solid lines refer to a system at -2.0 V and without ions, while dashed curves refer to a system with a Cl^- ion.

standard deviation of the forces decreases at increasing r_{fix} , but even when there are applied high negative potentials (solid lines), the variations rapidly decrease at small values between 8 and 12 Å. The comparison between these standard deviations with respect to the corresponding mean value of the forces further proves that the relative

contribution of the atoms outside the sphere is limited for $r_{fix} > 8 \text{ \AA}$ (Figure C.2b). Notice that the range above which the relative contribution becomes negligible is comparable with the receptive field of our ML models. For comparison, we also consider a Cu/H₂O system in which a Cl atom is added in the solvent region of the simulation cell. The Cl species accepts one electron from the metal slab, getting ionised as Cl⁻. This case is interesting since ions in solution are the typical example in which electrostatic long-range interactions are relevant [155]. In this case, the standard deviation is markedly higher than before, above all for the Cl species. Differently from the case without ions, here both the O and Cl contributions do not markedly decrease at higher r_{fix} , but they tend to reach a plateau value that can be still high, showing that the simple increase of r_{fix} is insufficient to decrease the variability in the forces. This can be considered the typical signature of long-range interactions. On the contrary, in our system, with extra surface charge due to the applied potential but without ions, the non-local contributions above $8 \div 10 \text{ \AA}$ are small, and hence, the employed MPNN architecture with $r = 6 \text{ \AA}$ and $n = 2$ is sufficient to capture all the relevant interactions. For this reason, all the analyses of the properties of the Cu/H₂O interfaces reported in Chapter 5 are obtained with ML-FFs without any specific long-range scheme.

C.2 Effects of the DFT functional on the density profile

The modulation of the solvent density discussed in Chapter 5 is the main characteristic that distinguishes any solid/liquid interface and is the result of a subtle balance between two types of interactions: the interfacial molecules interacting with surface atoms and the interfacial molecules interacting with other water molecules far from the electrode. Hence, it is expected that different DFT exchange-correlation functional approximations describe in different ways the balance between these interactions, resulting in variations in the solvent density [198]. Considering that AIMD simulations of metal/water interfaces at the SCAN level are extremely computationally demanding, the developed ML-FF also facilitates a benchmark of SCAN with respect to more commonly adopted functionals in the field of electrochemical interfaces. For example, it is possible to compare the results here reported with the findings reported in Ref. [144]. In that work, the authors developed an ML potential

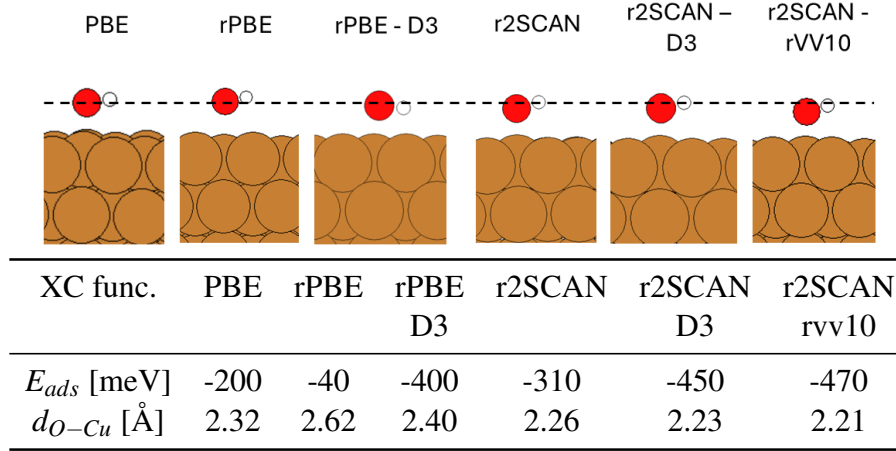


Fig. C.3 Comparison of geometries, adsorption energies E_{ads} and distances d_{O-Cu} between O and the nearest Cu atom for the adsorption of a single water molecule on Cu (111), computed with different exchange-correlation (XC) functionals. Dark orange, red and white balls are Cu, O and H atoms, respectively. The dashed black line is useful to guide the eye.

for Cu/water interfaces, using the rPBE-D3 approximation as the DFT reference for the dataset. The most relevant difference between SCAN ML-FF and rPBE-D3 ML-FF is associated with region II. Indeed, the latter forecasts a significantly higher O density (up to 4 times with respect to the water bulk density) and a narrower peak with respect to the findings reported in our work. Moreover, the H peaks are more shifted towards the surface in the picture predicted by the rPBE-D3 ML-FF. These elements suggest a stronger metal/water interaction predicted by the rPBE-D3 ML-FF

To better identify the difference in the metal/water interaction predicted by the two functionals, we perform a comparative analysis of the adsorption of a single water molecule on the bare Cu (111) surface among different exchange-correlation functionals. We consider also the inclusion of semi-empirical dispersion corrections and van der Waals functionals. Specifically, we focus on three common GGAs (PBE [78], rPBE [199], rPBE with the semi-empirical Grimme-D3 corrections [200]) and three meta-GGAs (r2SCAN, r2SCAN-D3 and the van der Waals functional r2SCAN-rVV10 [201]). Figure C.3 shows a comparative sketch for the adsorption geometries of a single water molecule, as predicted by the different functionals. Table C.3 reports the corresponding Cu-O distances and the adsorption energies. The water adsorption on the Cu top site is considered for all the cases, since this site is well established to be the most stable [202]. Common GGA functionals, such as PBE and rPBE, without any non-local contributions, predict low adsorption energies [203].

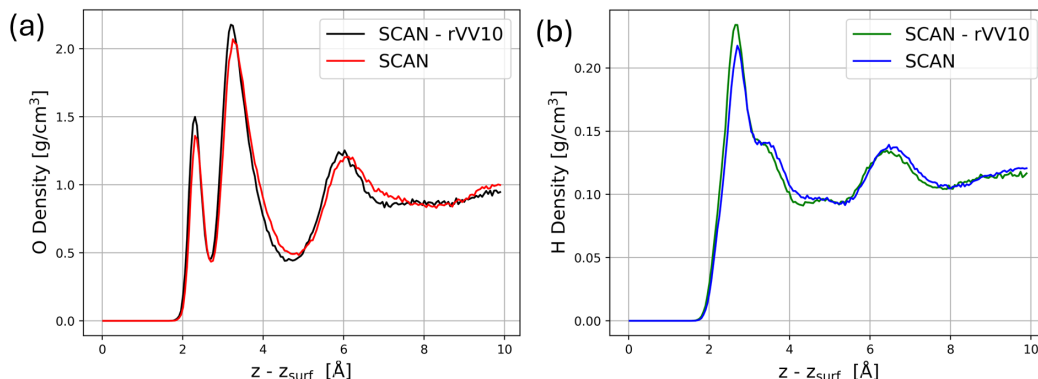


Fig. C.4 Comparison of the water density profiles computed with r2SCAN-rVV10 vs r2SCAN functional: (a) oxygen and (b) hydrogen contributions.

On the contrary, the addition of dispersion contributions to GGA functionals (e.g., rPBE-D3) is important to increase the adsorption energies of water. The SCAN functional family is already able to include some middle-range interactions without any additional van der Waals contributions [81]. For this reason, the adsorption energy of r2SCAN is already similar to the GGA result with dispersion contributions. The addition of D3 corrections or the rVV10 contribution to SCAN further enhances the water-metal interaction. However, the inclusion of van der Waals terms in SCAN does not significantly alter the adsorption geometry, as visible in Figure C.3. On the contrary, the D3 corrections modify the orientation of the adsorbed molecule with respect to the bare rPBE. Indeed, H atoms tend to be more exposed towards the metal surface for rPBE-D3. This proves that the enhancement of the adsorption due to D3 corrections for rPBE is associated with a stronger interaction between H and Cu atoms. Notice that this enhancement is only visible for rPBE-D3 and not for SCAN-D3 or for a van der Waals functional such as SCAN-rVV10. Hence, these different behaviours due to the D3 corrections can justify the previously identified differences in the solvent profile predicted by the two ML-FFs.

Considering the effects of adding van der Waals terms to common GGA functionals, we further investigate these contributions in the Cu/water interaction by developing an ML model trained at the SCAN plus non-local correlation functional rVV10 level. In view of the higher computational cost of rVV10, we do not develop a new ML-FF from scratch, but we follow a "delta-learning" approach [204]. Starting from the dataset computed at the SCAN level, we use the Deal scheme (cf. Section 3.2.2) to select about 410 configurations whose energies and forces are re-labelled at the r2SCAN-rVV10 level. This small dataset is used to train a second ML-FF able to

predict the energy and force differences between r2SCAN-rVV10 and r2SCAN DFT references. Then, MD simulations are performed summing the output of two ML-FFs (the first for the r2SCAN baseline and the second for the difference between these two DFT functionals). The solvent density profiles predicted by the r2SCAN-rVV10 ML model for the Cu (111) are reported in Figure C.4. The effects of van der Waals contributions due to rVV10 are mainly visible in the closest peaks to the surface. Here a higher density is associated with a more ordered structure of water. These results are consistent with the increase of adsorption energy of r2SCAN-rVV10 vs r2SCAN. Nevertheless, the variations in the peak densities are minimal and do not alter the description of the Cu/water interaction. For this reason, we conclude that SCAN alone is already able to reasonably capture the interactions between Cu and water, and the van der Waals interactions due to rVV10 are not necessary. Hence, all the results of this thesis are discussed considering the "plain" SCAN as the DFT reference.