UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Sciences
School of Chemistry

Developing and Evaluating Implicit Solvent Models for Catalytic Metallic Surfaces

by

Gabriel Adrian Bramley

ORCiD: 0000-0002-1096-7616

A thesis for the degree of
Doctor of Philosophy

May 2022

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences
School of Chemistry

Doctor of Philosophy

Developing and Evaluating Implicit Solvent Models for Catalytic Metallic Surfaces

by Gabriel Adrian Bramley

Understanding solvent effects at the metallic/liquid interface is critical to improving and analysing heterogeneous catalytic processes. In addition to a growing body of experimental work, computational studies are elucidating how the presence of water affects both the electronic structure and the adsorption thermodynamics of the metallic surface. However, computational methods such as *ab initio* Molecular Dynamics (AIMD) require extensive configurational sampling to obtain equilibrated thermodynamic quantities, precluding their use for wide-ranging studies. In contrast, by considering the dynamic degrees of freedom as an average, implicit solvent methods provide a route to tractable computational simulations of the aqueous environment, while maintaining a quantum mechanical description of metallic/adsorbate interactions. This thesis, in collaboration with the Pacific Northwest National Laboratory (PNNL), describes how implicit solvent approaches can be applied as an inexpensive method of evaluating both the electronic structure of the metal/liquid interface, and the free energy change of adsorption in the aqueous phase for a range of organic adsorbates.

To ensure these calculations are performed accurately and efficiently, developments were made to the linear scaling Density Functional Theory (DFT) code, ONETEP as part of this work. These developments include the implementation of the soft sphere dielectric cavity model, which gives the flexibility to parameterize the solvent model for individual atomic centres. This contrasts with the original electron density based cavity model, which applies a global cavity parameter, leading to poor descriptions of the free energy changes of solvation for systems with mixtures of light organic and heavy metallic species. A surface accessible volume term was also implemented for the non-polar solvation term, which improves the correlation with experimental solvation free energies compared to the surface area non-polar term. Furthermore, a Pulay Hamiltonian mixing routine was implemented in the Ensemble DFT (EDFT) scheme of ONETEP. This approach confers significantly improved convergence behaviour for single point energy calculations performed in this work. This enables more efficient and

accurate simulations of metallic systems, allowing for the evaluation of larger systems studied in later chapters.

Utilising the implemented soft sphere model, this work assesses the ability of the soft sphere model to capture the potential of zero charge and the work function of the metallic/liquid interface. By reparameterizing the implicit solvent model in terms of the work function values calculated from snapshots of an AIMD simulation, we were able to capture the salient electronic structure changes of the solvated metallic surface and electrochemical properties.

Then, combining the accelerated EDFT scheme and the implicit solvent parameterization method used for the potential of zero charge, this thesis concludes with a continuum solvent approach for calculating the aqueous phase adsorption free energy of organic molecules to the Pt(111) surface. In this work, approximations are derived for the entropies of solvation for the metallic surface based on analytical statistical thermodynamic expressions. These approximations allow us to parameterize the implicit solvent mode ΔG_{solv} for the metallic surface, enabling adsorption free energy with reasonable accuracy for a range of coverages and orientations. This opens a route for computationally inexpensive evaluations of adsorption processes at the aqueous Pt(111) interface, which can provide an atomistic understanding of adsorption processes in support of experimental studies.

The work presented in this thesis shows the usefulness of the implicit solvent method in studies of heterogeneous catalytic processes and electrochemical interfaces. The techniques described in this work show that thermodynamic and electrochemical properties can be calculated in a computationally tractable manner with implicit solvent. In future, this could enable high throughput studies for a range of metallic surfaces and adsorbates, aiding the design of catalysts for a range of applications.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:
   (1) Bramley, G.; Nguyen, M.-T.; Glezakou, V.-A.; Rousseau, R.; Skylaris, C.-K. Reconciling Work Functions and Adsorption Enthalpies for Implicit Solvent Models: A Pt (111)/Water Interface Case Study. *J. Chem. Theory Comput.* **2020**, 16 (4), 2703–2715.
   (2) Bramley, G. A.; Nguyen, M.-T.; Glezakou, V.-A.; Rousseau, R.; Skylaris, C.-K. Understanding Adsorption of Organics on Pt(111) in the Aqueous Phase: Insights from DFT Based Implicit Solvent and Statistical Thermodynamics Models. *J. Chem. Theory Comput.* 2022, Article ASAP.

Signed:..... Date:.....

Acknowledgements

At best of times, a PhD. is a strenuous undertaking. Needless to say, the events of the last two years made the process that bit more difficult. As such, it is even more important to extend gratitude to those who ensured I was able to complete this manuscript.

Firstly, thank you to my supervisor, Professor Skylaris for your support through this project. I hope I have become a better scientist through your guidance.

Thank you to my collaborators at PNNL, Vanda Glezakou, Roger Rousseau, and Manh Nguyen. Your insights were invaluable in producing this work and I regret that I could not visit Washington a second time. Special thanks to Manh for the data provided and presented as part of this thesis.

Thank you to the support teams of the various HPC clusters I have used over the last four years. This includes the IRIDIS 4/5 team, the MMM Hub and the ARCHER 2 Support team. Thank you for bearing with my queries.

Thank you to Jacek Dziedzic for keeping my ONETEP implementations in good order, and for your clear explanations on the inner workings of FORTRAN.

Thank you to all the past members of the Skylaris group: Lucas Verga, James Womack, Tom Ellaby, Jolyon Aarons, Gabriele Boschetto, Jamie Whipham, Marjan Famili, and Rebecca Clements. Thank you for your patience and warm company. I hope you're doing well in whatever you've moved on to.

Further thanks to the present members of the group: Dimitris Matthas, Han Chen, Catriona Gibbons, Arihant Bhandari, Julian Holland, Davide Sarpa, Nabeel Anwar, Tom Demeyere and Apostolos Kordatos. I'm sure you will all produce stellar work in the coming years. Special thanks to Tom Demeyere and Davide Sarpa for helping proof read this work.

Thank you to the members of Syma's group we commandeered for our lunchtime, Bastien, Conrado, Dhanushka and Sophie. I hope you all thrive in Oxford or wherever you go.

Thank you to the various housemates over in Shirley. You made that house as much a home as it could be.

Thank you to Rachel, Aria, and Nilam for basically having me as a third housemate when visiting London.

Thank you to the various University of York alumnus who kept me company during the pandemic. A special thanks to Pasky for his frequent life advice on how to get through the PhD process.

Finally, thank you to my parents for making sure I got this far. I hope what I have produced here was worth the time you've dedicated to me.

To my Loving Parents, Brothers and Friends

Chapter 1

Introduction

1.1 Computational Representation Aqueous/Metal Interfaces

Over the last three decades, the increasing sophistication of computing and electronic structure theory has enabled *ab initio* computational modelling to take a more prominent position in research. Even the simplest simulations performed in vacuum can validate experimental observations and provide precise insight into the changes in electronic structure that occur during a reaction. However, omitting the effects of solvation and electrolytes limit the accuracy of computational simulations. The influence of water is especially important when considering processes involving the conversion of biomass, which naturally contain large concentrations of water in the feedstock (see: Table 1.1). Even simple reactions such as the adsorption of H_2 onto the Pt surface, which form a basic step in the hydrodeoxygenation (HDO) reaction, are severely impeded by the introduction of aqueous solvent. Without water, only 0.001 atm partial pressure of hydrogen is required to achieve full coverage. However, water introduces an enthalpic and entropic penalty that increases the pressure to achieve a partial coverage of 0.85 to 1000 bar [1]. As a result, the HDO reaction, which could otherwise be performed at standard pressure in the absence of solvent, requires both high temperatures and high pressures of H_2 to achieve feasible rates under aqueous conditions.

1.1.1 *Ab initio* Molecular Dynamics

Simulating these complex environments at the metallic/solvent interface requires advanced theoretical techniques and greater computational resources compared to vacuum calculations. Molecular representations of water require one to consider the spatial degrees of freedom of the system in order to attain both entropic and enthalpic properties. Methods such as *ab initio* Molecular Dynamics (AIMD) allow the system to propagate over time, while thermostats constrain the thermodynamic conditions of

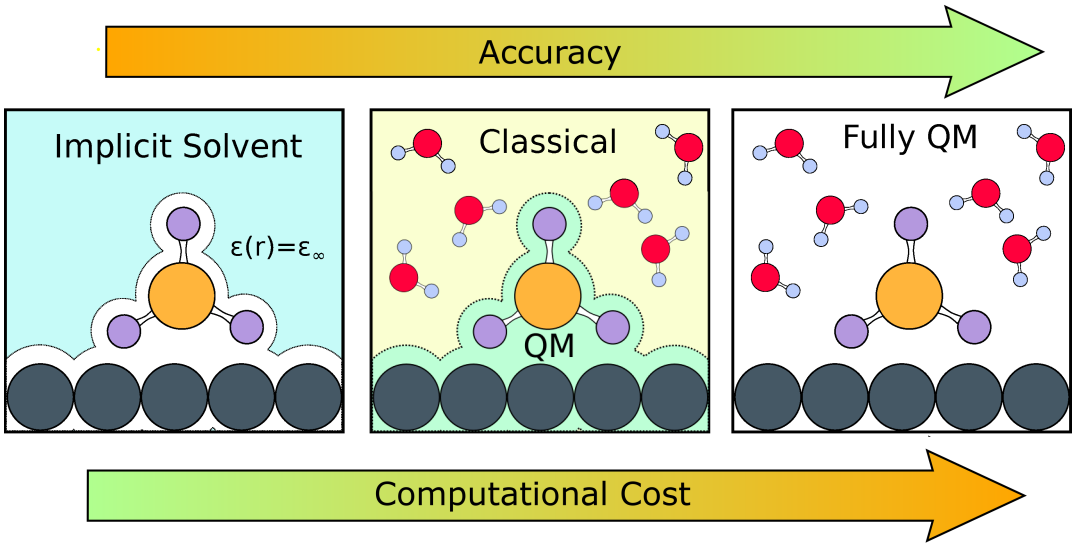

FIGURE 1.1: A simple schematic showing the the major simulation methods used to represent the aqueous phase in *ab initio* simulations. From left to right: implicit solvation, QM/MM and full quantum mechincal calculations.

the system according to the selected statistical thermodynamic ensemble. However, obtaining the thermodynamic average of even simple energetic properties requires adequate sampling of configurational space, which becomes increasingly unfeasible for large systems. For example, simulations performed for a small, 84 atom $Pt - H_2O$ interface require 24 ps of simulations - totalling 24,000 DFT single-point energy calculations - to achieve the distribution of energies characteristic of the Canonical Ensemble [2].

Despite these costs, several studies have been conducted into aspects of the HDO reaction using AIMD methods. Studies conducted by *Yoon et al.* [3] show the impact of the aqueous environment on the conversion of phenol to hexanol. The solvent effects are believed to be twofold: (a) The deposition of water alters the electronic structure of the surface, such that the reductive potential of the surface is increased and (b) the keto-enol tautmerisation equilibrium shifts to favour 3-cyclohexenone over 1,3-cyclohexadienol, leading to a favoured cyclohexanone product in the aqueous phase. Other AIMD studies at the aqueous/Pt interface reveal that under alkali conditions, adsorbed OH^- ions lead to charge transfer to the surface, increasing the activation energy of H_2 adsorption. This corresponds to a decrease in the rate of phenol hydrogenation with increasing pH values.

1.1.2 Quantum Mechanics/Molecular Mechanics (QM/MM)

The cost of AIMD calculations makes the study of extended systems unfeasible, necessitating the use of simpler models. One such approach is the QM/MM model, which

vastly reduces computational costs by describing the metallic surface and adsorbates quantum mechanically, and the bulk water through a classical force field. This significantly reduces the computational expense of representing the liquid environment, as the interactions of the MM atoms are described through a simple integration over point charges. However, these techniques require the total energy function to be re-cast to approximate the interaction between the QM and the MM region,

$$E(x_{QM}, x_{MM}) = E_{QM}(x_{QM}) + E_{MM}(x_{MM}) + E(x_{MM}) \tag{1.1}$$

where x_{QM} and x_{MM} are generic descriptors of the QM and MM atomic structures respectively. Using the approach, the thermal effects of solvation which rely on extensive sampling of configurational space can be calculated with significantly reduced computational cost. This is achieved using a QM system with a fixed structure, and the thermal degrees of freedom for water are captured using Newton's equations of motion in an appropriate statistical ensemble.

This approach has been used to calculate the free energy of solvation (ΔG_{solv}) of adsorbates at the metallic surface [4, 5]. Thermal motions were shown to be important in describing the energetics of solvation for small catalytic intermediates, where the entropy of solvation (ΔS_{solv}) is highly negative. The increased ordering of water accounts for a large component of ΔG_{solv}, where $\Delta S_{solv} = +32$ kJ mol^{-1} and the enthalpy of solvation (ΔH_{solv}) is -61 kJ mol^{-1} for CH_2OH^* adsorbed to Pt(111).

More advanced techniques such as QM/MM Free Energy Path (QM / MM-FEP) [6] allow the calculation of activation barriers and free energy changes of chemical reaction processes, while including thermal and solvent effects. This approach calculates the optimal reaction coordinates on the potential mean force surface (PMF). The PMF approximates the free energy surface in terms of the reaction coordinates in an ensemble average (mean field) of the solvent degrees of freedom. This interpretation allows for the evaluation of the potential energy surface, using a set of frozen QM geometries to represent the reaction coordinates between the initial and final states. Schemes such as eSMS (explicit Solvation for Metallic Surfaces) [7] have been developed using the QM/MM-FEP framework to evaluate bond breaking/formations between adsorbates on metallic facets in the aqueous phase [8]. This represents a promising method for evaluating reaction mechanisms compared to *ab initio* MD, allowing for speed-ups of up to an order of magnitude compared to the fully quantum description.

Recent developments to the QM/MM method include the improved description of the solvent/metal interface through purpose-built force fields. One such example is the GAL class of force fields (GAL17 [9]/GAL19 [10]), which are explicitly parameterized to obtain the chemisorption energies at specific sites on noble metal surfaces. This is achieved by applying a set of attractive Gaussian potentials to each surface site, and capturing the angular dependence of the H_2O binding strength through a truncated

Fourier series. This approach can calculate the binding energy of water to various noble metal surfaces with an error of 4.14 kJ mol^{-1} compared to DFT calculations. Follow-up studies applied the GAL17 force field to the adsorption of phenol to Pt(111) in the aqueous phase. They found this force field supported the experimentally observed qualitative decrease of the adsorption free energy in the aqueous phase compared to the vacuum phase [11]. However, by ignoring the configurational degrees of freedom of the adsorbate, this method does not account for the large negative entropies associated with the adsorption [12].

1.1.3 Implicit Solvent

Even with the computational savings brought about by the QM/MM method, equilibration of the MM system can still take a substantial amount of time and computational effort. This is especially the case if sampling is performed over a range of reaction coordinates along a free-energy path, or if many adsorbates need to be sampled in high-throughput studies. This can be limiting in studies that hope to evaluate a large assay of target molecules. Therefore, it can sometimes be useful to completely eliminate the explicit representation of the solvent, instead using a set of experimentally derived bulk properties to model the solute/solvent interactions.

One such approach is the implicit (continuum) solvent model [13], which treats the solvent system as a dielectric continuum surrounding the system of interest. This polarising medium emulates the effect of water by representing the solute/solvent interactions through the potential changes in a bulk dielectric. This essentially integrates out the spatial degrees of freedom of the solvent analogously to the PMF. As a well established technique, the implicit solvent model has been implemented in a range of forms including the Conductor-like Screening Model (COSMO) [14] and the Polarizable Continuum Model (PCM) [15]. A more complete description of the theory of the Polarizable Continuum Model applied to quantum chemistry codes is given in Section 2.5.

The implicit solvent model has already demonstrated its ability to model transition states for metallic surfaces [16]. *Lu et al.* [17] show the utility of this approach with the mean-field microkinetic modelling of the HDO for guaiacol over Pt(111), where the relative computational cost effectiveness of the solvent model allows the necessary calculation of tens to hundreds of elementary reaction steps. These aid in confirming important mechanistic details, such as the inactivity of the deoxygenation route of guaiacol on Pt(111) surfaces. This provides further evidence of the processes occurring at the cocatalyst, where hydrogenation occurs on the noble metal surface and deoxygenation occurs exclusively on the support. Although this new technique has yet to be used more broadly, these findings demonstrate that it has potential to guide catalysis research by allowing a semi-quantitative account of environmental effects on kinetics.

However, a well-known limitation of the implicit solvent model is the poor description of strong bonding interactions between the solvent and solute. This is especially prevalent for species which form hydrogen bonds with water. Comparisons with QM/MM studies show that for catalytic intermediates such as CH_2OH^* at the Pt(111) surface, ΔG_{solv} is underestimated by 12 kJ mol^{-1} [5].

1.2 Conversion of Biomass to Biofuels

Simulations of metallic surfaces and the solvent environment can guide material design for industrial processes. One such application is the conversion of crude bio oils to biofuels, which show promise as a sustainable and carbon neutral alternative to non-renewable fossil fuels [18]. Although the biofuel sector is expanding 7% year-on-year, continuous growth of 10% is required to meet the sustainability targets set by the International Energy Agency [19]. In order to aid the expansion of this industry, further research must be conducted to overcome the challenges limiting the economic feasibility of biofuel production. Solutions to these problems lie not only in improved logistics [20] and optimised plant design [21], but also in forming a rational understanding of the heterogeneous catalytic processes used at various stages of biomass conversion.

Early biomass sources, categorized as the first generation, are generated from purpose grown food grade crops to produce bioethanol [22] [23]. However, the sustainability of these feedstocks are questioned given their competition with domestic food markets, which is projected to increase the price of commodities such as cereals up to 34% [24]. Further concerns have been raised about the environmental impact of intensive farming related to first-generation biomass production, namely the depletion of aquifers and deforestation [25]. This mitigates some of the environmental benefits that biofuels have over fossil fuels. To tackle these issues, the second generation of biomass feedstocks has been developed. These feedstocks are derived from agricultural, forestry and household waste as well as purpose-grown non-food crops.

The primary components of biomass are lignin, cellulose, and a mixture of short-chain sugars [26], all of which must be refined to short-chain hydrocarbons to be used as fuels. Several promising techniques have been developed to convert raw biomass into feasible fuel sources, such as gasification and pyrolysis/liquification. Fast pyrolysis provides an easily implementable synthetic route for producing crude bio-oil, producing up to 60% yields of potentially usable fuel by mass [27]. This technique involves heating biomass to high temperatures (approximately 500 ^{o} C) in the absence of oxygen with short residence times, followed by rapid condensation of the resulting aerosols. The liquid phase product is furthermore easily stored and transported, especially compared to the products of the gaseous biomass conversion process. However, bio-oil produced

Component	Corn Husks[a]	Pine[b]
Water	25	24
Aldehydes	1	7
Acids	6	4
Carbohydrates	5	34
Phenols	4	15
Ketones	11	4
Furans	2	3
Alcohols	0	2
Unknown	46	57

TABLE 1.1: The bio-oil composition by wt% produced from fast pyrolysis from two feedstocks. Compiled by *Mortenson et al.* [29]. [a] *Mullen et al.* [30]. [b] *Oasmaa et al.* [31]

by fast pyrolysis has a considerable oxygen content (up to 60% wt) compared to non-renewable fossil fuels. The high oxygen content imparts many undesirable properties on the fuel, such as high viscosity, low volatility, and low energy density. Furthermore, the low thermal stability of the oxygen groups leads to the polymerisation of the fuel (coking) at operating temperatures. These properties were found to be disadvantageous in the first tests of diesel engines, where the coking process and acidic impurities in unrefined bio-oil damaged key components of the engine after prolonged operation [28]. These problems preclude the use of crude bio-oils produced by fast pyrolysis as a transportation fuel, limiting its application to standing boilers.

In order to broaden the utility of bio-oils to both automotive and aviation fuels, additional processing must be performed to reduce the oxygen content. Furthermore, since the crude bio-oil product is a mixture of ethers, ketones, and phenolic compounds (the components are shown in Table 1.1), the process must be robust enough to eliminate a range of functional groups containing oxygen. In answer to this, the combination of fast pyrolysis and catalytic upgrading through hydrodeoxygenation (HDO) has emerged as one of the most prevalent routes to convert raw biomass to transportation fuel [33]. The HDO step is typically performed with zeolite/alumina/silicate supported metallic nanoparticles of Pt, Ru, Cu or Ni, [34] or sulfided Co-Mo/Ni-Mo co-catalysts [35]. The latter catalysts represent the most common class in the early development of hydroupgrading, but the said catalysts rapidly deactivate under aqueous conditions [36] and require high pressures of H_2 to perform hydrogenation [21]. The noble metal class of catalysts are a promising alternative, exhibiting higher yields compared to the sulphided co-catalysts at lower hydrogen concentrations (3-5 MPa) [37], and greater resistance to deactivation under aqueous conditions. Figure 1.3 shows how conversion is carried out for a lignin-derived compound, guaiacol, with Pt/AlO_3. Herein, the metallic surface activates H_2 via. dissociative adsorption for hydrogenation of the $C = C$ groups, while the support provides a Brønsted acid site for the subsequent dehydration reaction. The optimisation of the design and composition of these catalysts is an active area of research. The influence of the support has been highlighted, where the

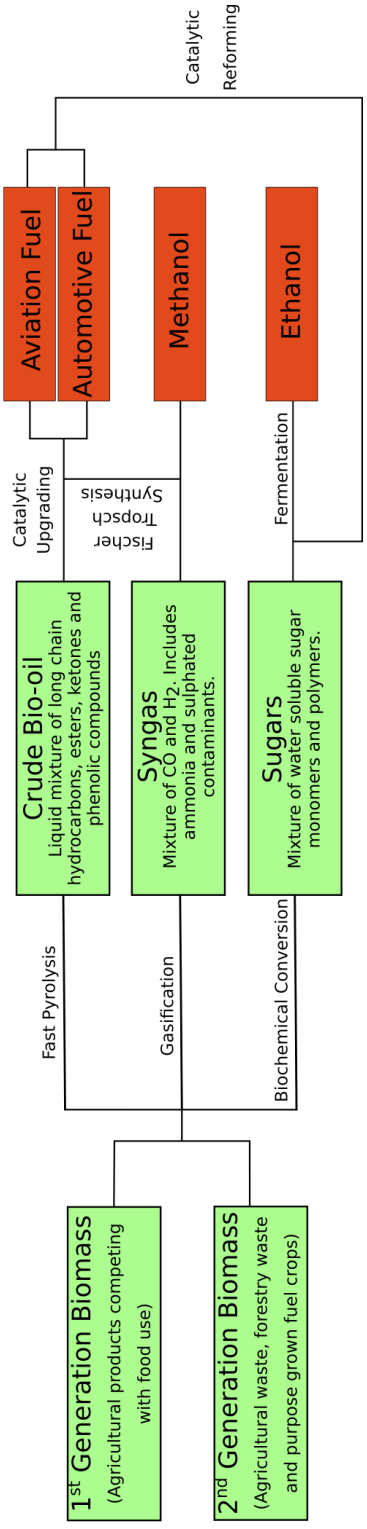

FIGURE 1.2: An overview of the processes which convert biomass into useful biofuels. Adapted from *Yue et al.* [32]

.

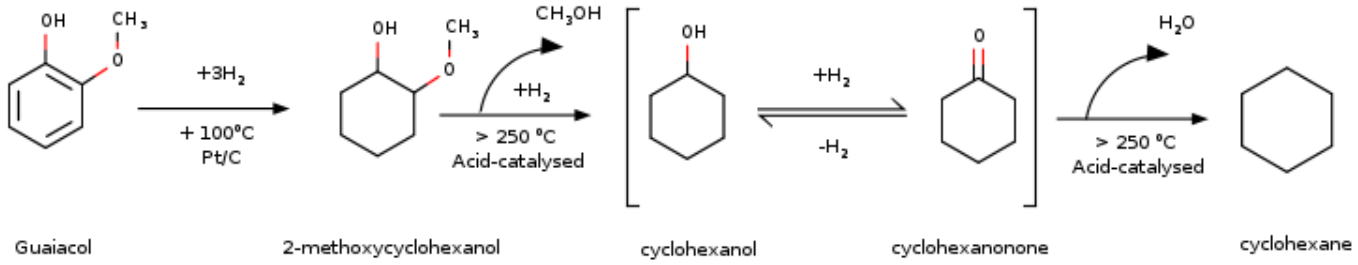

FIGURE 1.3: A simplified scheme of the conversion of guaiacol on a Pt/AlO_3 catalyst. Scheme adapted from *Lee et al.* [38]

zeolite support is known to be susceptible to deactivation from phenolic compounds via. coking [39]. However, alumina supports have recently shown promise as a material robust to poisoning [40]. Furthermore, the yield of these reactions are sensitive to temperature effects, where the favoured product of HDO changes from cycloalkanes at 200°C to phenolic/polymeric products at 300°C [41]. In addition, the choice of Ru nanoparticle species as opposed to Pt can improve the yields of alkane products from 68.6% to 87.2% [42].

The precise understanding of this reaction is complicated by the interplay of support and nanoparticle effects, coupled with reaction conditions such as the use of high temperatures, pressure and acidity. As a result, computational studies can be used to provide insights into the precise mechanisms of the HDO reaction. However, producing models for the HDO reaction is a complex task that requires careful consideration of the environment.

1.3 Experimental Techniques for Measuring Adsorption Free Energies

Accurate measurements of the energetics of adsorption are a vital part of understanding heterogeneous catalytic processes. Especially in complex multicomponent mixtures, the competitive adsorption of different species can impact the turnover frequencies of the target reaction [43], or lead to catalytic poisoning if certain species bind irreversibly to the catalyst compared to the desired reactant [44]. Furthermore, the presence of solvent influences the thermodynamics of adsorption [45–47]. This is demonstrated both enthalpically, by the energetic penalty of displacing the adsorbed H_2O molecules from the catalytic surface [48], and entropically through the favourable transfer of H_2O from dynamically restricted adsorbed configurations to the liquid phase [47, 49].

However, measuring the energetics of adsorption in the aqueous phase is particularly challenging compared to the case in the gas phase. In experimental studies, quantitative measurements such as calorimetry are significantly less sensitive due to heat

dissipation through the solvent medium, meaning a large number of adsorption events must be induced to adequately measure the heat released [50, 51]. Furthermore, the structure and orientation of the adsorbates on the surface are sensitive to a variety of properties within the aqueous environment, including the concentration of the adsorbate [52] and the presence of electrolytes [53]. As such, the measured binding modes can often be ambiguous compared to ultrahigh vacuum experiments, which is further complicated by many traditional microscopy techniques that require ultrahigh vacuums to avoid scattering of the incident beam [54]. However, techniques using ultrathin stable liquid layers have recently been developed for surface studies, which allows the measurement of high-resolution images of the surface/adsorbate interface in the aqueous phase [54].

In computational studies, measurements of thermodynamic quantities at the aqueous interface require extensive configurational sampling of the adsorbed system, typically through molecular dynamics. However, these techniques require the calculation of many thousands of snapshots to obtain equilibrium quantities. This is especially problematic for *ab initio* treatments, where individual energy and force evaluations are computationally expensive. This is further exacerbated by the large amount of H_2O molecules introduced to the simulation cell, required to adequately simulate the solvent environment [55]. Simplified models such as QM/MM (quantum mechanics/molecular mechanics) [10], which separate the system into a quantum and classical region, allowing for more tractable calculations of adsorption processes. Alternatively, one can use methods such as implicit solvation (or continuum solvation), which remove the requirement to perform configurational sampling by averaging out the dynamic degrees of freedom for water [13, 56], drastically reducing the computational expense for calculating properties in the solvent phase.

One of the primary objectives of this work is to provide methods for computing the free energy changes associated with solvation, making use of the reduced computational costs of the implicit solvent model. We will apply our implicit solvation approach to the measurement of the free energy of adsorption under aqueous conditions (ΔG^{solv}_{ads}) of aromatic molecules on the surface of Pt(111). Furthermore, the accuracy of our adsorption and solvation model will be assessed by comparing our ΔG^{solv}_{ads} to experimental literature values.

To better inform our discussion, we will provide a brief account of the evolving techniques used to obtain ΔG^{solv}_{ads} experimentally. Early studies concerning the adsorption of aliphatic and aromatic molecules to metallic electrodes in the aqueous phase were performed largely through FTIR (Fourier Transformed Infrared Spectroscopy) and radiotracer techniques [45, 46]. These experiments obtain ΔG^{solv}_{ads} values by calculating $K^{solv}_{eq,\theta}$ as a function of coverage (θ) over a range of concentrations.

$$K^{solv}_{eq,\theta} = \exp\Big\{(-\Delta G^{0}_{ad,aq,\theta}/RT)\Big\} = \exp\Big\{(-\Delta H^{0}_{ad,aq,\theta}/RT)\Big\}\exp\Big\{(\Delta S^{0}_{ad,aq,\theta}/R)\Big\}, \quad (1.2)$$

where $\Delta G^0_{ad,aq,\theta}$, $\Delta H^0_{ad,aq,\theta}$, and $\Delta S^0_{ad,aq,\theta}$ are the standard state changes in Gibbs free energy, enthalpy, and entropy of adsorption in the aqueous phase respectively. To calculate the coverage for a given concentration, the surface concentration of the adsorbate, Γ (measured in mol cm^{-2}) is measured through the radiation signal of adsorbed C^{13} isotopes. Coverage is then calculated as,

$$\theta = \frac{\Gamma}{\Gamma_{max}}, \tag{1.3}$$

where Γ is normalised against the surface concentration of one monolayer of coverage, Γ_{max}. The models of the isotherm for adsorption are then constructed by plotting c_{ads} vs. θ. Overall, the adsorption process is described as,

$$\mathrm{Mol_{(aq)} + nH_2O_{(ads)} \longrightarrow Mol_{(ads)} + nH_2O_{(l)}}, \tag{1.4}$$

where n is number of water molecules displaced by the adsorption of the water molecule. The concentration of the adsorbate in solution and θ (which is unitless throughout) are related by the Temkin isotherm, which takes into account the lateral repulsions between adsorbates [57]. In the linear form, the Temkin isotherm is defined as,

$$\theta = \frac{RT}{r_1} \ln K^{solv}_{eq,\theta} + \frac{RT}{r_1} \ln c_{ads}, \tag{1.5}$$

where r_1 is the Temkin constant (which describes the variation of ΔG_{ads} with coverage [47]), R is the gas constant, T is the temperature, and $K^{solv}_{eq,\theta}$ is the equilibrium constant.

Empirical studies of organic adsorption to metallic surfaces continued in the gas phase through the SCAC (Single Crystal Adsorption Calorimetry) method [58], but until recently there has been a dearth of reported adsorption free energy values for the aqueous phase. However, measurements of ΔG^{solv}_{ads} are undergoing a revival, brought about by Cyclic Voltammetry (CV) techniques [49, 59, 60].

Similarly to the radiotracer approach, the CV method is used to obtain the coverage of organic adsorbates at a given concentration. This technique exploits the hydrogen underpotential deposition process, where hydrogen reversibly adsorbs to the metallic surface at characteristic potentials in the underpotential region,

$$\mathrm{M\text{-}H \rightleftharpoons H^+_{aq} + e^- + M_s{}^*}, \tag{1.6}$$

where M−H represent the adsorbed hydrogen to the metallic surface, and M* represents an empty site. In equal electrochemical conditions, the potentials where hydrogen is deposited is determined by both the species of the metal and the crystallographic direction of the facet. The number of hydrogens deposited on the metallic surface can be calculated by the charge density (A cm^{-2} over the area of the surface), which is integrated over the potential swept over the underpotential region. As organic adsorbates

block the sites which can be occupied by the hydrogen adatoms, the coverage of the adsorbates can be measured by the relative loss of current transferred during the CV potential sweep. This is measured as,

$$\frac{\theta/(1-\theta)}{\theta^0/(1-\theta^0)} = K^{solv}_{eq,\theta}\frac{c_{ads}}{c_0}, \tag{1.7}$$

where θ^0 is taken as 0.5 to simplify the expression $\theta^0/(1-\theta^0)$ to 1. So far, this technique has been used to measure the adsorption free energy changes of various aromatic compounds to the Pt(111) surface.

Experimental data from electrochemical and radiotracer studies generally agree quantitatively with one another (Table 1.2 and 1.3) - the adsorption free energies for organics in the aqueous phase are significantly less thermodynamically favourable than in the gas phase. Monolayer coverage was reported as unobtainable at high concentrations. This is rationalised as either the inability to determine the true number of surface Pt atoms, the existence of sites inaccessible to the adsorbate or sites where the kinetics of water desorption/adsorbate adsorption are unfavourable. However, these early studies agree overall with the qualitative decrease of ΔG^{solv}_{ads} with respect to ΔG^{vac}_{ads} as found in modern measurements (1.3. Therefore, these older studies will provide a valuable resource in parameterizing our solvent model and testing its robustness to systems beyond cyclic aromatics.

TABLE 1.2: Vacuum phase heats of adsorption for a range of aromatic and common hydrocarbons with oxygen-bearing functional groups.

Molecule	θ	ΔH^{vac}_{ads} / kJmol^{-1}
Phenol (Pt(111)/Ni(111))[a]	1/16	-194, -175
Benzene (Pt(111)/Ni(111))[b]	1/16	-188, -183
Cyclohexene Pt(111)[c]	0	-130
Naphthalene Pt(111)[d]	0	-300
Methanol Pt(111)[e]	0	-60.5
Methoxy Pt(111)[f]	0	-187

[a] Ref. [61], Single Crystal Adsorption Calorimetry (SCAC), 100K. [b] Ref. [62], SCAC, 90K. [c] Ref. [63], SCAC, 300K, $\Delta H_{ads} = (197 - 48\theta - 83\theta^2)$. [d] Ref. [64], SCAC, 300K, $\Delta H_{ads} = (300 - 34\theta - 199\theta^2)$. [e] Ref. [65], SCAC, 100K, between $0 - \frac{1}{3}$ coverage $\Delta H_{ads} = (60.5 - 19.3\theta)$. [f] Ref. [65], derived from Hess cycle.

These studies have both attempted to derive further values such as the enthalpy and entropies of adsorption, the bond strength of phenol and the rate constant. Firstly, the Van't Hoff equation can be rearranged into the Van't Hoff plot,

$$\frac{\Delta G^{solv}_{ads}}{RT} = \frac{\Delta H^{solv}_{ads}}{RT} + \frac{\Delta S^{solv}_{ads}}{R}. \tag{1.8}$$

TABLE 1.3: Aqueous phase change in adsorption free energy and heat of adsorption under standard conditions on the Pt(111) facet. $\Delta G^{0}_{ad,aq,\theta=0}$ refers to the free energy changes of adsorption proceeding from components in their aqueous phase.

Molecule Adsorbed	$\Delta G^{solv}_{ads,\theta=0}$ / kJ mol^{-1}	$\Delta H^{solv}_{ads,\theta=0}$ / kJ mol^{-1}	$\Delta S^{solv}_{ads,\theta=0}$ / J K^{-1} mol^{-1}
Phenol[a]	-9	-71	-
Phenol[b]	-14	$+38 \pm 19$	$+174 \pm 64$
Phenol[c]	-40	$+12$	-
Benzaldehyde[a]	-31	-43	-
Benzaldehyde[d]	-	-44	-
Cyclohexanol[a]	-17	$+2$	-
Benzyl Alcohol[a]	-33	-32	-
Butanol[c]	-33	-17	$+59$
Valeric Acid[c]	-38	-13	$+81$
Benzoic Acid[c]	-42	$+13$	$+191$
Naphthol[c]	-54	$+17$	$+244$
Naphtholic Acid[c]	-63	$+17$	$+254$

[a] Ref. [45], Cyclic Voltammetry, standard conditions. [b] Ref. [49], Cyclic Voltammetry, standard conditions. [c] Ref. [47], Radiotracer/FTIR, standard conditions. [d] Ref. [66], Aqueous phase calorimetry.

Through a linear regression of temperature with respect to ΔG^{solv}_{ads}, $\Delta H^{solv}_{ads,\theta=0}$ and ΔS^{solv}_{ads} can be calculated as the gradient and intercept, respectively, of Equation 1.8. As shown in 1.3, the enthalpies of adsorption in the aqueous phase relative to the gas phase adsorption enthalpies decrease significantly, which is largely attributed to the displacement of water from the Pt(111) surface. However, these processes remain thermodynamically feasible due to the large positive entropies attributed to the configurational freedom gained by water desorbing from the surface and entering the aqueous phase. However, *Akinola et al.* [49] notes that the Van't Hoff approach involves large errors in the measurements of ΔS^{solv}_{ads}, driven by the small variation of K^{solv}_{ads} across the experimental temperature ranges. In the case of phenol adsorbing to the Pt(111) facet, the error of entropy is estimated as ± 64 J K^{-1} mol^{-1} compared to the measured value of $+174$ J K^{-1} mol^{-1} (where the error was obtained by performing repeat measurements of ΔS^{solv}_{ads}

Further techniques have been developed to determine the bond strength (enthalpy of binding for the adsorbate) in the aqueous phase by breaking the adsorption process into its constituent elementary steps. In Bockris' study of the adsorption enthalpy [47], the adsorption process is divided into four enthalpic components, as shown in Figure 1.4: (1) ΔH_1 the enthalpy of desorption for n displaced H_2O molecules from the metal M, (2) ΔH_2, the enthalpy of desolvation for a phenol molecule, (3) ΔH_3, the condensation enthalpy of n H_2O molecules, and (4) ΔH_4, the enthalpy of adsorption of phenol to the metallic surface. In this model, the unknown quantity (ΔH_4) can then be calculated using experimentally derived ΔH_{ads} and the largely available values for the quantities

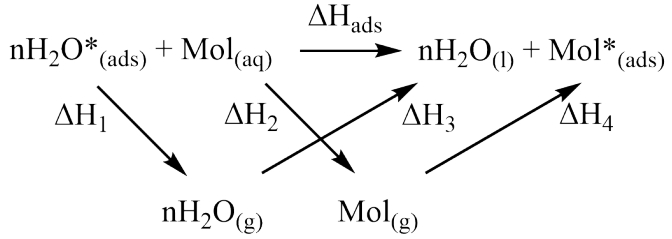

FIGURE 1.4: The individual steps involved in the adsorption of a molecule in the aqueous phase in the metallic surface.

ΔH_1, ΔH_2 and ΔH_3. Once ΔH_4 is derived, the binding strength of ΔH_{bind}^{solv} was derived as $\Delta H_{ads}^{solv} = \Delta H_4 - \Delta H_2$.

However, the ability of these early studies to accurately decompose the total ΔG_{ads}^{solv} into elementary steps were further limited by the lack of data for certain values (e.g., the adsorption enthalpy of water to the Pt surface). In the latter case, ΔH_{ads} was estimated to be -92kJ mol^{-1} compared to the accepted value of -51 kJ mol^{-1} [47], significantly distorting many of their calculated heats of adsorption. Furthermore, these studies largely ignore the energies of desolvating sections of the molecule that are inaccessible to the solvent once adsorbed to the metallic surface.

The bond additivity model developed by *Singh et al.* [48] further expanded this work by including the neglected energetic penalties of breaking water interactions with the adsorbed segment of the molecule. This model also takes the energetically favourable process of reforming the water bonds with the bulk aqueous phase into consideration. Furthermore, the value of ΔH_1 was better represented by experimental measurements of the adhesion energy of water to clean Pt(111) [67]. The bond additivity model predicts that the binding strength for phenol on Pt(111) as $\Delta H_{ads}^{solv} = -136$ kJ mol^{-1}, which is 57 kJ mol^{-1} lower than the corresponding ΔH_{ads}^{vac} at high coverage. However, the calculated ΔH_{ads}^{solv} matches ΔH_{ads}^{vac} at high coverages, where the binding energy decreases by approximately 50 kJ mol^{-1} compared to the limit of $\theta = 0$. Therefore, it has been postulated that phenol forms islands of high coverage in the aqueous phase, implying that the presence of water has little influence on the binding strength of phenol.

1.4 The Entropy of Adsorption

In computational studies, difficulty in calculating the free energy of adsorption arises primarily from the entropic term. As an intrinsically statistical property, the calculation of entropy requires the use of heavily approximated analytical statistical mechanics expressions or configurational sampling across the potential energy surface. As this requires energy evaluations for many thousands of configurations, this approach is intrinsically computationally expensive. As a central quantity in determining ΔG_{ads},

we shall give a brief overview of the various assumptions used to calculate ΔS_{ads} in both the aqueous and vacuum phase. An account of the basic 3D entropy equations for isolated molecules in the gas phase are given in Appendix A to support this discussion.

As the largest entropy component across the surface, many studies are concerned with estimations of the translational entropy of the molecule. In the simplest approximation, one can approximate the entropy of the adsorbate on the surface (ΔS_{ads}) through the 2D ideal gas equation, i.e., the free translator (FT) model [68]. This equation modifies the translational component of entropy to,

$$S_{ads,t}^{vac} = k \ln (q_t^{vac,2D}) + k + kT \left(\frac{\partial \ln (q_t^{vac,2D})}{\partial T} \right)_A, \tag{1.9}$$

where k is Boltzmann's constant, A is the area of the surface and $q_t^{vac,2D}$ is the translational partition function in vacuum across a 2D plane. $q_t^{vac,2D}$ is then defined as,

$$q_t^{vac,2D} = \left(\frac{2\pi mkT}{h^2} \right) A, \tag{1.10}$$

where m is the mass of the adsorbate and h is Planck's constant. The adsorption entropy is calculated relative to the ideal 3D entropy of the molecule, which is measured in a given volume (Appendix A). Although conceptually intuitive, several assumptions are being made of the surface and molecule under study. Firstly, we assume that the diffusion barriers for the adsorbate are very small relative to the thermal energy of the molecule ($kT >> \epsilon_d$). This assumes not only a relatively flat potential energy surface, but also high enough temperatures to overcome ϵ_d. *Jørgensenn et al.* [69] demonstrated that the FT model can act as a lower bound estimate for the loss of entropy, provided that the potential energy surface is relatively flat for the adsorbate in question (e.g., Pt(111)/CO) [69]. However, for species such as oxygen on the Pt(111) surface, they also showed that increased barriers, which hinder translation, lead to the free translator model overestimating the entropy of the adsorbate by a factor of two at room temperature.

At the other extreme, one can assume that at sufficiently low temperatures ($kT << \epsilon_d$), the diffusion barrier is effectively infinite, so the entropy of adsorption is determined by the distribution of the adsorbate across the adsorption sites. Therefore, in this model, the entropy of the adsorbate in the gas phase (S_{ads}^{gas}) is taken in the limit of the ideal 2D lattice gas, that is, the configuration entropy S_{config} [70],

$$S_{config} = S_{ads}^{vac} = R \ln \left(\frac{(1-\theta)}{\theta} \right), \tag{1.11}$$

where the adsorbates are considered indistinguishable and the entropy does not vary significantly with temperature. This term makes a large contribution to adsorbates with large lateral repulsions (such as NH_3 on Mg(100)), where $S_{ads}^{vac} \rightarrow 0$ as $\theta \rightarrow 1$

[12]. Later models such as the Hindered Translator (HT) [71, 72] model bridge the gap between the extremes of the 2D ideal gas and 2D lattice gas assumption by formulating a partition function within a periodic potential and applying a scaling relation to relate the classical mechanical to the quantum mechanical partition function,

$$q_t^{vac,2D} = \frac{M(\frac{\pi r_x}{T_x}) \exp\left[-\frac{r_x}{T_x}\right] \exp\left[-\frac{1}{T_x}\right] I_0^2[\frac{r_x}{T}]}{\left(1 - \exp\left[-\frac{1}{T_x}\right]\right)^2} \exp\left[\frac{2}{(2+16r_x)} T_x\right], \tag{1.12}$$

where M is proportional to the area ($M = A/b^2$ and b is the nearest neighbour distance of a slab with 3-fold symmetry), $r_x = \epsilon/h\nu_x$, which gives the ratio of the diffusion barrier and the energy of the vibrational frequency mode ν_x of the adsorbate on Pt(111) in the x direction, $T_x = kT/h\nu_x$ is the dimensionless temperature normalised to the vibrational frequency, and I_0 is the zero-Bessel function of the first kind, designed to replicate the periodicity of the potential. This approach is significantly more generalisable than the 2D lattice and 2D ideal gas models alone, and allows for entropy approximations for a range of species by explicitly including the ambient temperature and the relative size of the vibrational energy compared to the diffusion barriers. The (HT) model compares favourably with the CPES (complete potential energy sampling) technique [69], which better describes the potential used in the partition function by integrating the potential energy of the adsorbate across different positions within the metallic unit cell. Across a range of temperatures for CO on the Pt(111) surface, the error of HT relative to CPES is on an order of $< 10\%$ - however, as with the FT model, these errors are relatively larger for Pt/O, where the potential energy surface is less homogeneous.

Experimentally, it has been determined that for adsorbates that exhibit attractive interaction, the 2D ideal gas approximation can be used to describe the entropy of an adsorbate (S_{ads}^{gas}) through a linear relationship with the standard gas phase entropy ΔS_{gas}^{o} [12],

$$S_{ads,gas}^{0}(T) = 0.7 S_{gas}^{0}(T) - 3.3R = 0.68(S_{gas}^{3D,o} - S_{gas,t_{(z)}}^{3D,o}), \tag{1.13}$$

where $S_{gas,t_{(z)}}^{3D,0}$ is the translational entropy associated with the z-direction, perpendicular to the adsorption interface. From these experiments, we obtain a simple but powerful validation of the simple 2D ideal gas model of adsorption in certain scenarios. However, as described in the case of NH_3/Mg(100), this model does not account for adsorbates that strongly repel each other, which means that this equation does not apply to all adsorbates. Furthermore, this model does not take into account the local environments of the adsorbate, which may further restrict the rotational entropy.

Dauenhauser et al. approximates the impact of the local environment by including the entropy change induced by confinement within a zeolite cavity [73]. Applied to zeolites with pores of varying volume, they found that alkanes placed within smaller volumes lose a greater number of rotational degrees of freedom, while the translational entropy

remains constant. As such, they propose a switching function, which modifies the total entropy change as a function of the volume cavity occupied by the adsorbate,

$$\Delta S^0_{ads} = S^0_{1D,t_z} + (F_{rot,slab} + \frac{1}{7}[(1 - \frac{V_{critical}}{2V_{occ}} - 1)])S^0_{rot}, \tag{1.14}$$

where $F_{rot,slab}$ is the number of rotational degrees of freedom lost upon adsorption, and $V_{critical}$ describes the volume in which all rotational degrees of freedom are lost. Comparisons of the entropy between the smallest and largest cavity volumes for C3-C6 alkanes showed that the overall entropy of adsorption from the gaseous phase can decrease by a factor of two due to the confinement. Therefore, including the effect of the local environment is a required to accurately reflect the overall change in entropy [73].

Although a central quantity in the calculation of adsorption thermodynamics in the aqueous phase, there is no definitive theoretical approximation for the desorption entropy of a water layer, much less an experimental value. However, the body of literature allows one to intuit reasonable upper and lower bounds using both experimental quantities, dynamics at the metallic surface, and 2D lattice statistics. Using the ice-like bilayer model at low temperatures, the upper bound of $\Delta S^{H_2O}_{ads}$ can be calculated as the entropy of fusion (ΔS_{fus}) for water, which is derived from readily available thermodynamic data ($S^{(l)}_{H_2O} - S^{(s)}_{H_2O} = 25.1\ \text{J K}^{-1}\ \text{mol}^{-1}$) [49].

In contrast to the upper bound, in the lower bound, it is assumed that water retains its liquid-like dynamics but is constrained to the xy-plane. The simplest representation of adsorption from *Campbell et al.* [12] (Equation 1.13) yields an entropy value of $S^0_{ad}(T) = 104.16\ \text{J K}^{-1}\ \text{mol}^{-1}$, which leads to an nonphysical positive entropy compared to liquid water ($S^o_{(l)}$) $= 69.95\ \text{J K}^{-1}\ \text{mol}^{-1}$ [74]. A simple physical explanation could be that this model does not take into account the relatively highly ordered nature of water, that cannot be simply approximated by the loss of transitional freedom in the z-axis. Theories which take into account further reductions in rotational entropy (such as Dauenhauer's Universal Descriptor of Adsorption Entropy for Confined molecules) [73] similarly yield a positive entropy with respect to liquid water, even if all degrees rotational degrees of freedom are constrained when transitioning from the gas to the adsorbed aqueous phase. In the absence of a reasonable lower bound entropy of adsorption for water, we derive an approximation based on the cavitation entropy of H_2O in Chapter 6.

1.5 Conclusions and Outlook

The study of reaction processes at the metallic surfaces is a rapidly evolving field of enquiry, both from the perspective of experimentalists and theoreticians. Many of the

techniques described in this chapter promise to provide a great deal of insight into the design of heterogeneous catalysts. However, the body of experimental work is relatively small because of the difficulty of measuring thermodynamic and structural properties in the aqueous phase. There is therefore a large scope for computational studies to support experimental work by providing atomistic information unobtainable with current lab techniques. For example, DFT methods in vacuum can indicate the effect of nanoparticle size on adsorbate binding, as well as the change in electronic structure induced by adsorption [75].

This field will likely grow further in the coming years, and a greater amount of experimental data for adsorption processes will aid the development of computational techniques and vice versa. In this work, we construct a framework in which solvent calculations for surface processes can be performed at low computational cost. By doing so, we can validate the conclusions of experimental work by providing structural and electronic information of the adsorbates at the metallic surface, and guide experimental work to metallic surfaces/nanoparticles with favourable catalytic behaviour.

Using the implicit solvent model, this thesis presents a set of methods to simulate effect of the water environment on metallic surfaces. This enables the computation of both electronic properties and adsorption properties at low computational costs. Chapter 2 gives a theoretical overview of DFT and the physical principles of periodic systems employed in the ONETEP electronic structure code. Furthermore, the implicit solvent method is discussed in more detail, including details unique to the ONETEP implementation. Chapter 3 then presents improvements made to the implicit solvent model by reparameterizing non-polar cavity terms, and a distance based dielectric cavity model (the soft sphere model). Further implementations to ONETEP are shown in Chapter 4, where the underlying total energy algorithm used in Ensemble DFT is accelerated using a Pulay mixing algorithm. Using these developments, Chapter 5 explores the use of a reparameterized implicit solvent model to calculate the change of the work function of Pt(111) in the aqueous phase, which is compared to a set of AIMD simulations of an explicit metal/solvent interface. Finally, continuing the calculations on the Pt(111) surface, Chapter 6 uses the implicit solvent model to calculate the free energies of adsorption in the water solvent environment for a set of organic molecules.

Chapter 2

Theory

2.1 Wave Functions and Wave Mechanics

The task of quantum chemistry is to formulate a theory that is both physically reflective of the system and practicable in terms of computational resources. In order to achieve this, vast amounts of thought have been dedicated to evaluating approximations and creating a framework which can be implemented in computational chemistry software. Among the most common contemporary techniques is Density Functional Theory (DFT), which is used extensively throughout this work. In this chapter, we will describe the foundations of this theory, as well as practical developments made over the decades to ensure that DFT can be used accurately and efficiently.

2.1.1 Schrödinger Equation

Schrödinger revolutionised the modern chemist's approach to quantum mechanics by applying the De Broglie relationship to the classical Jacobi-Hamiltonian operator, thereby re-casting a previously intractable problem into a feasible eigenvalue expression [76],

$$\hat{H}|\Psi\rangle = E|\Psi\rangle, \tag{2.1}$$

where the wavefunction (Ψ) is the many-body wavefunction containing all the information for the system, $\hat{H}$ is the Hamiltonian operator and E is the total energy of the system. In effect, this reduces the calculation of the total energy to an eigenvalue problem, provided that a precise form of the Hamiltonian is known. Although the concept of the wave function was introduced with no formal definition itself, Born later postulated that by taking a squared modulus of a normalised wavefunction, one obtains the

probability density ($P(\mathbf{r})$) of the particle represented by Ψ,

$$P(\mathbf{r}) = \int |\Psi(\mathbf{r})|^2 dxdydz, \tag{2.2}$$

where $\mathbf{r}$ is a position vector defining a point in cartesian space (x, y, z). Therefore, the Born interpretation postulates that the probability of finding a particle in a given volume is proportional to $|\Psi(\mathbf{r})|^2$. As $N\Psi(\mathbf{r})$ is also a solution to Equation 2.1, we can normalise Equation 2.2 so that the probability of finding a particle in a given volume is unity,

$$1 = N^2 \int |\Psi(\mathbf{r})|^2 dxdydz, \tag{2.3}$$

For the purposes of quantum chemistry, the Hamiltonian operator [77] for an atomic system of **M** nuclei and **N** electrons is formulated in terms of atomic units as,

$$\hat{H}_{tot} = -\sum_{i=1}^{N} \frac{1}{2} \bigtriangledown_i^2 - \sum_{A=1}^{M} \frac{1}{2M_A} \bigtriangledown_A^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{i,A}} + \sum_{i=1}^{N} \sum_{i>j}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{M} \sum_{B>A}^{N} \frac{Z_A Z_B}{R_{AB}}, \tag{2.4}$$

where each term corresponds to the following properties: (1) the kinetic energy of the **i**th electron, (2) the kinetic energy of the **A**th nucleus with mass M_A, (3) the Coulombic attraction between the electron i and the nucleus A with charge Z_A, with the distance between each nucleus and electron pair defined as $r_{i,A} = |r_i - R_A|$, (4) the repulsive term between each of the electrons, and (5) the repulsive terms between nuclei A and B.

However, even for molecular hydrogen, this form of the Schrödinger equation becomes a many-body problem, for which there is no exact analytical solution. This arises from the coupled motion of nuclei with electrons and, in the case of molecular systems, the nuclei with other nuclei.

The Hamiltonian can be significantly simplified by introducing the Born-Oppenheimer approximation, which postulates that due to the significantly higher mass of the nuclei compared to the electrons, their velocities are correspondingly far smaller than those of the electrons. Therefore, the nuclei can be considered as approximately stationary in the co-ordinate system. With this approach, the electrons interact with a mean field of positive charge, as opposed to a positively charged body with correlated motions. This treatment simplifies the Hamiltonian in two ways. Firstly, the nuclear kinetic energy term is eliminated as the nuclear velocity is set to zero. Secondly, electron-electron interactions are decoupled from the nuclear interactions, allowing the formulation of a Hamiltonian that depends only on the motion of the electrons in a field of point charges. Furthermore, the nuclear-nuclear interactions can simply be treated as a fixed constant

to the total energy. The results of this can be summarised by the decoupling of the electronic terms from the Hamiltonian, where equation (2.5) shows the electronic Hamiltonian of i^{th} electrons in a field of M positive point charges.

$$\hat{H}_{elec} = -\sum_{i=1}^{N} \frac{1}{2} \bigtriangledown_i^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{i,A}} + \sum_{i=1}^{N} \sum_{i>j}^{N} \frac{1}{r_{ij}} \tag{2.5}$$

$$\hat{H}_{elec}|\Psi\rangle = E_{elec}|\Psi\rangle, \tag{2.6}$$

where the total energy term can be obtained by adding the now constant nuclear-nuclear attraction term.

$$E_{tot} = E_{elec} + \sum_{A=1}^{M} \sum_{B>A}^{N} \frac{Z_A Z_B}{R_{AB}}, \tag{2.7}$$

2.1.2 The Variational Principle

Even with the Born-Oppenheimer approximation, systems as simple as H_2 prove impossible to solve analytically. This is due to the correlated motion between each of the electrons in the electron-electron repulsion term, from which deriving an analytical expression proves impossible because of the interdependency of their positions. However, one can use numerical procedures starting from an approximate first guess of Ψ, and iteratively solving the eigenvalue problem to obtain improved values of Ψ as close as possible to the exact wavefunction. However, this relies on the variational principle, which posits that the energy of the orthonormalised trial wavefunction gives an upper bound solution to the exact ground state energy,

$$\langle \Psi_{trial}|\hat{H}|\Psi_{trial}\rangle \geq \langle \Psi_0|\hat{H}|\Psi_0\rangle = E_{exact}, \tag{2.8}$$

where,

$$\langle \Psi_{trial}|\Psi_{trial}\rangle = 1. \tag{2.9}$$

Proof of this can be derived by representing the trial wave function Ψ_{trial} as a linear combination of fixed eigenstates with expansion coefficients c_i. This defines a set of orthonormal basis functions, which reduces the task of finding an optimal solution to the electronic eigenvalue problem (Φ_{exact} and $(E)_{exact}$) to finding an accurate set of values for c_i.

$$|\Psi_{trial}\rangle = \sum_{i=1}^{N} c_i|\psi_i\rangle, \tag{2.10}$$

where for an orthogonal set of single particle wavefunctions,

$$\langle \psi_i|\psi_j\rangle = \delta_{ij}, \tag{2.11}$$

where $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ for $i \neq j$. These expressions can in turn be substituted into the expression yielding the expectation energy value,

$$E[\psi] = \frac{\langle \Psi_{trial} | \hat{H} | \Psi_{trial} \rangle}{\langle \Psi_{trial} | \Psi_{trial} \rangle}, \tag{2.12}$$

$$= \frac{\sum_{i=1}^{N} |c_i|^2 \int d\mathbf{r} \psi_i(\mathbf{r}) \hat{H} \psi_i(\mathbf{r})}{\sum_{i=1}^{N} |c_i|^2 \int d\mathbf{r} \psi_i(\mathbf{r}) \psi_i(\mathbf{r})} \tag{2.13}$$

given the orthonormality condition for the basis functions of ψ_{trial},

$$E[\psi] = \frac{\sum_{i=1}^{N} |c_i|^2 E_i \int \psi_i(x) \psi_i(x) dx}{\sum_{i=1}^{N} |c_i|^2 \int \psi_i(x) \psi_i(x) dx} = \frac{\sum_{i=1}^{N} |c_i|^2 E_i}{\sum_{i=1}^{N} |c_i|^2}. \tag{2.14}$$

By subtracting the exact energy (E_0) from both sides, one finally obtains the relationship necessary to prove the variational principle.

$$E[\psi] - E_0 = \frac{\sum_{i=1}^{N} |c_i|^2 (E_i - E_0)}{\sum_{i=1}^{N} |c_i|^2} \tag{2.15}$$

As the right-hand term is always greater than or equal to 0, it follows that the trial energy will always be greater or equivalent to the energy obtained by the exact solution, therefore proving the equation (2.8).

This contains a simple but powerful idea: for different wavefunctions calculated through solutions to the Schrödinger equation, there exists a hierarchy of solutions dictated by their energies. We can therefore interpret wavefunctions which give a lower value of E as being of intrinsically higher quality, and those with higher energies being further from E_{exact}. Following from this, one can create iterative procedures from which the wavefunction can be successively improved, until one obtains an upper bound to the exact energy. Therefore, a relatively accurate solution to the eigenvalue problem can be obtained without requiring either a full analytical expression or precise *a priori* knowledge of the wavefunction.

2.2 Density Functional Theory

Early quantum chemistry made extensive use of the Hartree-Fock Theory. The key assumption is the description of electron-electron electrostatic repulsions through a mean-field, as defined in the Coulomb exchange terms ($\mathcal{J}_i(\mathbf{r})$ and $\mathcal{K}_i(\mathbf{r})$) of the closed-shell Fock operator:

$$\hat{f}(\mathbf{r}) = \hat{h} + \sum_{i}^{N/2} 2\hat{\mathcal{J}}_i(\mathbf{r}) - \hat{\mathcal{K}}_i(\mathbf{r}), \tag{2.16}$$

where $\hat{h}$ is the core Hamiltonian operator,

$$\hat{h} = -\frac{1}{2}\nabla^2 + \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{i,A}}. \tag{2.17}$$

However, this assumption ignores the correlated motions between electrons, which introduces an error compared to the exact energy, known as the correlation energy.

$$E_c = E_{\text{exact}} - E_{\text{HF}} \tag{2.18}$$

Post-Hartree Fock methods such as Configuration Interaction (CI) and Moller-Plesset Theory (MP) partially account for the correlation energy by accounting for higher order excitations of the ground state wavefunction by using multicomponent determinants. However, these methods are incredibly computationally costly, and alternative formulations of the eigenvalue problem become attractive. By far, the most popular of these is the Density Functional Theory (DFT), which is founded on reforming the ground state energy equation (Equation (2.4)) into a functional dependent on the electron density as opposed to a many electron wavefunction (Equation (2.19)).

$$E_{elec}[n] = F[n(\mathbf{r})] + \int n(\mathbf{r})[v(\mathbf{r})]d\mathbf{r}, \tag{2.19}$$

where the second term represents the interaction of the electron density ($n(\mathbf{r})$) with an external potential ($v(\mathbf{r})$), and $F[(n(\mathbf{r}))]$ is an unknown functional which is a composite of the kinetic and exchange-correlation energies. In order to make this functional physically useful, two key conditions are required: 1) Equation 2.19 needs to be made applicable to all systems (universal), such that the Hamiltonian is fully known once the number of electrons and external potential are known, and 2) a variational theorem must be derived such that a minimum energy can be found.

2.2.1 Hohenberg-Kohn Theory

Both of the previously mentioned conditions were satisfied by two theorems postulated by Kohn and Hohenberg [78]. The original formulations posed the number of electrons, N and the external potential $v_{ext}(\mathbf{r})$ as the central properties of DFT, but their work showed that there exists an exact energy functional depending solely on $n(\mathbf{r})$. In addition, this functional follows variational principle, meaning the quality of the $n(\mathbf{r})$ can be interpreted in terms of its calculated energy. The derivation of these principles follows from two of their posited theorems.

THEOREM 1 : *The external potential $v_{ext}(\mathbf{r})$ is determined, within a trivial constant, by the electron density $n(\mathbf{r})$.*

This theorem posits that the many particle ground state can be uniquely derived from $n(\mathbf{r})$. This does so by linking each of the individual variables to the electron density, thereby making it the basic variable. The multibody wave function is used to represent the ground state electron density for a closed-shell system as follows,

$$n(\mathbf{r}) = N \int d\mathbf{r}_1 \cdots \int \mathbf{r}_N \Psi(\mathbf{r_1} \ldots \mathbf{r_N}) \Psi^*(\mathbf{r_1} \ldots \mathbf{r_N}), \tag{2.20}$$

where we condense $\Psi(\mathbf{r_1} \ldots \mathbf{r_N})$ as $\Psi(\mathbf{r})$ for ease of discussion.

The proof posed by Hohenberg-Kohn proceeds as follows. Two different multielectron systems are hypothesised: (1) $v_{ext}(\mathbf{r})$ corresponding to Ψ and (2) $v'_{ext}(\mathbf{r})$ corresponding to $\Psi'(\mathbf{r})$. Each of these wave functions are said to map to the same electron density ($\Psi'(\mathbf{r}) \rightarrow n(\mathbf{r})$ and $\Psi(\mathbf{r}) \rightarrow n(\mathbf{r})$), but by necessity $\Psi'(\mathbf{r}) \neq \Psi(\mathbf{r})$ as they give rise to two different Hamiltonians, $\hat{H}$ and $\hat{H}'$. This means each of these wave functions corresponds to solutions of different Schrödinger equations.

Using the minimal energy condition expressed in Section 2.1.2, $\Psi'(\mathbf{r})$ can be taken to be a trial wavefunction of $\hat{H}$, giving rise to a non-minimum energy. As the difference between the Hamiltonians arises through the different external potentials, the following inequality can be derived.

$$E = \langle \Psi(\mathbf{r}) | \hat{H} | \Psi(\mathbf{r}) \rangle < \langle \Psi'(\mathbf{r}) | \hat{H} | \Psi'(\mathbf{r}) \rangle \tag{2.21}$$

$$= \langle \Psi'(\mathbf{r}) | \hat{H}' | \Psi'(\mathbf{r}) \rangle + \langle \Psi'(\mathbf{r}) | \hat{H} - \hat{H}' | \Psi'(\mathbf{r}) \rangle \tag{2.22}$$

$$= E' + \int n(\mathbf{r})[v(\mathbf{r}) - v'(\mathbf{r})] d\mathbf{r} \tag{2.23}$$

This logic can also be applied using $\Psi(\mathbf{r})$ as a trial wave function.

$$E' = \langle \Psi(\mathbf{r}) | \hat{H} | \Psi(\mathbf{r}) \rangle + \langle \Psi(\mathbf{r}) | \hat{H}' - \hat{H} | \Psi(\mathbf{r}) \rangle \tag{2.24}$$

$$= E_0 + \int n(\mathbf{r})[v'(\mathbf{r}) - v(\mathbf{r})] d\mathbf{r} \tag{2.25}$$

Adding these two inequalities yields $E + E' < E' + E$, which is clearly contradictory. Therefore, this theorem proves that $n(\mathbf{r})$ produces a unique $v_{\text{ext}}(\mathbf{r})$, which in turn fixes the value of Hamiltonian for each value of the electron density. In addition, the number of electrons N can be related to $n(\mathbf{r})$ by a simple integration over the whole volume of interest.

$$N = \int_{-\infty}^{\infty} n(\mathbf{r}) d\mathbf{r} \tag{2.26}$$

Therefore, we satisfy the requirement for a universally applicable energy functional dependant only on $n(\mathbf{r})$.

THEOREM 2: *For a trial density,* $\widetilde{n}(\mathbf{r}) \geq 0$ *and,* $\int \widetilde{n}(\mathbf{r}) d\mathbf{r} = NE_0 \leq E_v[\widetilde{n}(\mathbf{r})]$.

This theorem is a simple extension of the principles stated in the first Kohn-Sham postulate and the variational theorem of a wave function-based multielectron minimal energy functional. As we know that the density yielding the exact minimum energy $n_0(\mathbf{r})$ maps to a unique $v_0(\mathbf{r})$ and Ψ_0, it stands to reason that trial density $\widetilde{n}(\mathbf{r})$ will map to a unique trial wavefunction $\widetilde{\Psi}$ and external potential $\widetilde{v}(\mathbf{r})$,

$$\langle\widetilde{\Psi}|\hat{H}|\widetilde{\Psi}\rangle = \int \widetilde{n}(\mathbf{r})v(\mathbf{r})d\mathbf{r} + F_{HK}[\widetilde{n}(\mathbf{r})] = E_v[\widetilde{n}(\mathbf{r})] \geq E_0[n_0(\mathbf{r})]. \tag{2.27}$$

The result of these equations is a powerful *ab initio* technique which, in its formulation, is exact and universal from the outset. However, this implies a full knowledge of the functional $F_{HK}(n)$, which is unfortunately unknown. Therefore, one must make assumptions to obtain an approximate form of the functional in order to usefully apply these results.

2.2.2 Kohn-Sham Theory

Early DFT formulations struggled to reproduce even basic electronic features otherwise available through Hartree-Fock [79]. The central difficulty lay in the inability to exactly formulate the kinetic energy through a functional, taking the electron density as a variable. However, Kohn and Sham [80] re-cast density functional theory in a way which is both easily implemented and significantly more accurate than its predecessors. Kohn and Sham introduced the idea that the electron density can be constructed through a set of N, single particle orbitals:

$$n(\mathbf{r}) = \sum_{i}^{N} |\psi_i(\mathbf{r})|^2, \tag{2.28}$$

where the kinetic energy can be defined as,

$$T_s[n(\mathbf{r})] = \sum_{i}^{N} \langle\psi_i| - \frac{1}{2}\nabla^2|\psi_i\rangle. \tag{2.29}$$

This formulation of the kinetic energy describes a system of N non-interacting electrons. However, $T_s[n]$ does not directly correspond to the functional as defined in $F[n]$. However, by redefining $F[n]$, Kohn and Sham showed that the orbital-dependent kinetic energy term can be separated out, and the remaining errors are grouped into another functional. Therefore, making $T_s[n]$ an exact solution to the kinetic energy portion of the ground state DFT functional. This functional is defined as,

$$F[n(\mathbf{r})] = T_s[n(\mathbf{r})] + V_H[n(\mathbf{r})] + E_{\text{xc}}[n(\mathbf{r})], \tag{2.30}$$

where T_s is the kinetic energy term, V_H is the Hartree potential, and F_{xc} is the exchange-correlation term. The exchange-correlation term accounts for the error between the

exact kinetic energy functional and the system of N non-interacting electrons, and the exact electron-electron interaction energies. This expression is then used to derive the Euler-Lagrange equation for the non-interacting case (a more in depth correlation and discussion of this term can be found in [79]:

$$\mu = v_{\text{eff}}(\mathbf{r}) + \frac{\delta T_s[(\mathbf{r})]}{\delta n(\mathbf{r})}, \tag{2.31}$$

where $v_{\text{eff}}(\mathbf{r})$ is the effective potential defined as,

$$v_{\text{eff}}(\mathbf{r}) = v(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{xc}}(\mathbf{r}), \tag{2.32}$$

and the exchange correlation potential, $v_{\text{xc}}(\mathbf{r})$ is defined as the derivative of the exchange-correlation functional.

$$v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[n]}{\delta n(\mathbf{r})}. \tag{2.33}$$

The central result of the Kohn-Sham formulation lies in the fact that equation (2.32) is a simple rearrangement of the Lagrangian associated with the exact ground state kinetic energy for an interacting electron system. Therefore, using a set of one-electron equations, the electron density, $n(\mathbf{r})$ can be derived through an eigenvalue problem associated with a given v_{eff}.

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{eff}}(\mathbf{r}) \right] \psi_i = \epsilon_i \psi_i \tag{2.34}$$

Together, Equations (2.28), (2.34) and (2.33) form what are known as the Kohn-Sham equations. From this basis, one can construct an iterative procedure from which one can obtain successively more accurate electron densities. This can be achieved by using a trial density, which calculates v_{eff}, which in turn can be applied to the single-electron eigenproblem to obtain a new $n(\mathbf{r})$.

Through this approach, Kohn-Sham secluded all the unknown functional forms into the exchange-correlation term, making the rest of the equation universally applicable to all systems. Many approaches exist to accurately describe this term, such as the local density approximation (LDA) and the generalised gradient approximation (GGA), which form the basis of the most functionals used in modern day quantum chemistry. However, they all use the Kohn-Sham approach as their theoretical basis, which shall be the assumed formulation of DFT throughout the rest of this text.

2.2.3 Finite-Temperature DFT

This approach works demonstrably well for insulators, which by definition possess a large band gap. This formulation begins to break down for conductive systems such as metals due to the constraint of occupancies near Fermi level at 0 K, which are assumed to be either occupied or unoccupied. This assumption leads to numerical instabilities in conductive systems, where states near the Fermi level have near-degenerate eigenvalues, but significant differences in their spatial representations. This results in a phenomenon called 'level-crossing', where the occupancy of states close to the Fermi level repeatedly swaps, leading to large changes in the total energy [81]. With increasing system sizes and decreasing HOMO-LUMO gaps, the number of iterations required to obtain convergence which scales on an order of N or worse, which results from the difficulty in determining the correct occupied and unoccupied orbitals [82].

However, one can improve the scaling of these calculations by relaxing the constraint of binary occupancies. As such, methods such as the finite temperature Kohn-Sham DFT [83, 84] determine occupancies near the Fermi level based on a probability distribution at a constant electronic temperature, which in effect smears the occupancies of the ground state electronic structure near the Fermi level. However, to achieve this the total energy minimisation problem (Equation 2.19) must be re-cast to a grand potential or a Helmholtz free energy,

$$A[T;\{\epsilon_i\},\{|\psi_i\rangle\}] = \sum_i f_i \langle\psi_i|\hat{T} + \hat{V}_{\text{ext}}|\psi_i\rangle + E_H[n] + E_{\text{XC}}[n] - TS[\{f_i\}], \tag{2.35}$$

with k as the Boltzmann constant, μ as the Fermi level, f_i as the fractional occupancies of the corresponding eigenstates and T as the finite electronic temperature. In addition to this occupancy function, one must account for the fact that energies are no longer being calulated at 0 K, which introduces an electronic entropy term that must also be considered in the minimisation procedure. This is done so by considering the Helmholtz free energy as opposed to the pure Kohn Sham energy, which adds an entropy term to the standard energy functional, taking the occupancy of the electronic states as an argument.

$$S[\{f_i\}] = -k\sum_i [f_i \ln f_i + (1-f_i)\ln(1-f_i)]. \tag{2.36}$$

The electron density expressed by the Born rule is expanded in terms of its wavefunctions and the Fermi-Dirac probability function, f_i,

$$n(\mathbf{r}) = \sum_i^{\frac{N}{2}} f(\epsilon_i)\psi_i(\mathbf{r})\psi_i^*(\mathbf{r}), \tag{2.37}$$

which corresponds to the overall occupancy of the state in terms of the probability f_i of observable microstate,

$$f(\epsilon_i), = kT\left(\frac{\partial \ln(Z)}{\partial \mu}\right)_{V,T} = \frac{1}{\exp\left[\frac{\epsilon_i - \mu}{kT}\right] + 1}$$

minimis

2.3 Periodic Boundary Conditions and the Plane Wave Basis Set

2.3.1 Bloch's Theorem

The study of metallic systems can be readily simplified by exploiting the periodicity of perfect crystalline structures. Through this, the structural properties of large systems such as infinitely extended surfaces or bulk systems can be modelled through a periodic array of unit cells.

For a crystalline solid, the periodic array of underlying repeating units are described through the Bravais lattice. In this framework, each point in the Bravais lattice can be described through a three-dimensional position vector [85],

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{2.38}$$

where $\mathbf{a}_{1..3}$ are the primitive Bravais lattice vectors which maps any point in the Bravais lattice onto a periodic replica in directions $< a, b, c >$, and $n_{1..3}$ are integers describing the number of units moved in each lattice direction. The lattice vectors furthermore define the dimensions of the primitive unit cell. This cell is defined as the minimum enclosed volume consisting exclusively of unique points in space (i.e., a region where no points map to a periodic replica of itself). In principle, the Bravais lattice approach allows for the description of infinitely repeating unit cells for bulk crystals and surface facets. However, we have only presented the DFT framework in real space. Clearly, an infinitely large system cannot be calculated computationally, so we must formulate a set of basis functions which replicates the periodicity of the Bravais lattice. The bulk periodic system must therefore be described as a small unit cell under periodic boundary conditions.

Alternatively, the Bravais lattice can be defined in terms of reciprocal lattice, where the reciprocal lattice vectors are defined as,

$$\mathbf{G} = g_1\mathbf{b}_1 + g_2\mathbf{b}_2 + g_3\mathbf{b}_3 \tag{2.39}$$

where the $\mathbf{b_{1..3}}$ are the primitive reciprocal lattice vectors. These are defined in terms of the real space Bravais lattice vectors such that,

$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)},\ \mathbf{b}_2 = 2\pi\frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)},\ \mathbf{b}_2 = 2\pi\frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}, \tag{2.40}$$

where the product in the denominator corresponds to the volume of the primitive unit cell. Overall, the reciprocal lattice vector is defined such that when it is taken with a dot product of the real space lattice vector, it yields,

$$\mathbf{G} \cdot \mathbf{R} = 2\pi(g_1 n_1 + g_2 n_2 + g_3 n_3). \tag{2.41}$$

Using Bloch's theorem [86], the infinite system can be broken down into a sum of periodic plane waves ($e^{i\mathbf{k}\cdot\mathbf{r}}$),

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \tag{2.42}$$

where $\mathbf{k}$ is a wavevector, and $u_{\mathbf{k}}(\mathbf{r})$ is a function chosen to have the same periodicity of the Bravais lattice, such that,

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{\mathbf{k}}(\mathbf{r}). \tag{2.43}$$

The proof also follows that $\psi_{\mathbf{k}}(\mathbf{r})$ is an eigenstate of the translation operator $\hat{T}$, which is defined as shifting the selected function in question by $\mathbf{R}$. When applied to eigenstate $\psi_{\mathbf{k}}(\mathbf{r})$, we obtain,

$$\hat{T}_R \psi_{\mathbf{k}}(\mathbf{r}) = \psi_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot(\mathbf{r}+\mathbf{R})} u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \cdot [e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r})] = e^{i\mathbf{k}\cdot\mathbf{R}} \psi_{\mathbf{k}}(\mathbf{r}). \tag{2.44}$$

This has the benefit of making any two eigenfunctions differing by the inverse lattice vector $\mathbf{G}$ equivalent to each other. By considering $\mathbf{k}' = \mathbf{k} + \mathbf{G}$, one prove this outcome by transforming (2.42) for wavevector n (representing a band index) by the inverse lattice vector,

$$\psi_{n,\mathbf{k}'} = e^{i\mathbf{k}\cdot\mathbf{r}}[e^{i\mathbf{G}\cdot\mathbf{r}} u_{n,\mathbf{k}'}(\mathbf{r})] = e^{i\mathbf{G}\cdot\mathbf{r}} u_{m,\mathbf{k}}(\mathbf{r}). \tag{2.45}$$

The periodic function $u_{n,\mathbf{k}}(\mathbf{r})$ can be decomposed into a Fourier series,

$$u_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n,\mathbf{k}} e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{2.46}$$

where the Fourier coefficient couples to the solution of another electron state, m, which takes the values from the wavevector,

$$c_{n,\mathbf{k}+\mathbf{G}} = c_{m,\mathbf{k}} e^{-i\mathbf{G}\cdot\mathbf{r}} \tag{2.47}$$

These powerful results show that instead of considering an infinite number of wavefunctions across the periodic system, one only needs to consider a finite number of basis functions associated with each $\mathbf{k}$. As such, the wavefunctions can be defined through a finite basis of plane waves through $u_{n,\mathbf{k}}(\mathbf{r})$. Additionally, the periodicity of $\mathbf{k}$ means

that only the first Brillouin zone, defined as the minimum unit cell of reciprocal space, needs to be sampled to attain all the electronic information of the system.

2.3.2 Brillouin Zone Sampling

Calculating total energy with a plane wave scheme requires an integration with respect to the wave vector **k** over the Brillouin zone. In practice, solutions are obtained numerically, breaking down the expression into a sum of contributions for each point in **k**-space. However, since this would require the computation of an infinite sum, sampling schemes have been derived, so only a few **k** points need be considered when calculating the electronic properties of a system.

By far, the most common **k** point sampling method is the Monkhorst-Pack scheme [87], which constructs an NxNxN mesh of points defined along each reciprocal lattice.

$$\mathbf{k} = u_i \mathbf{b}_1 + u_j \mathbf{b}_2 + u_k \mathbf{b}_3 \tag{2.48}$$

$$u_r = \frac{2r - q - 1}{2q}, \tag{2.49}$$

where r is a sequence of integers up to a maximum, q, specified for the calculation ($r = 1, 2, 3, ..., q$).

As the Brillouin zone is defined within reciprocal space, the size of the real space lattice vectors and reciprocal lattice vectors are inversely linked. In practice, this leads to a decrease in the size of the Brillouin zone when using large simulation cells. As a result, the wavefunctions vary less across the inverse space, so fewer **k**-points are required to gain satisfactory sampling. This increases to the point where, with sufficiently large systems, only one **k**-point (known as the Γ point) needs to be considered. Current implementations of ONETEP all work under this approximation and only use Γ-point sampling.

In order to construct a mesh for a particular type of system, one must benchmark the number **k**-points required to reach the total energy convergence on a system-by-system basis. This is especially important for long-form studies on one type of system, where poor sampling of the Brillouin zone can lead to significant error, while using too many points involves unnecessary computational cost.

2.4 ONETEP

2.4.1 Linear Scaling DFT

Plane wave Kohn-Sham DFT poses many improvements compared to Hartree-Fock, and scales significantly better in terms of computational cost. Despite this, the delocalised KS orbitals extend across the whole simulation cell, meaning that the resources required to perform a calculation scale cubically. Therefore, there is a significant incentive to find schemes that reduce the workload required to solve the single-particle Kohn-Sham equations, ideally achieving a linear correlation between computational time and system size.

A basis for linear-scaling schemes is found in the near-sightedness of electronic matter observed by Walter Kohn [88]. This principle states that electrostatic interactions between the Kohn-Sham potential at $\mathbf{r}'$ and a charge $n(\mathbf{r})$ reduce to zero beyond short distances.

At the core of this approach is the reformulation of the charge density $\rho(\mathbf{r}, \mathbf{r}')$ in terms of a single-particle density matrix,

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{j=1}^{N} f_i \psi_i(\mathbf{r}) \psi_i(\mathbf{r}'), \tag{2.50}$$

where f_i is the occupancy of the orbitals and $\psi_i(\mathbf{r})$ is the orthogonal set of Kohn Sham states. This single body expression of the density corresponds to the interacting density via.

$$n(\mathbf{r}) = 2\rho(\mathbf{r}, \mathbf{r}'). \tag{2.51}$$

The original scheme by Kohn [89] stipulates that the single particle density must satisfy the idempotency condition $f_i^2 = f_i$ (that is, the occupancies are either 0 or 1), such that,

$$\rho^2(\mathbf{r}, \mathbf{r}') = \rho(\mathbf{r}, \mathbf{r}'), \tag{2.52}$$

which ensures that energy minimisation is performed in a stable manner, avoiding occupancies greater than 1 at lower energy levels. In principle, this method creates a route for minimisation schemes with $O(N)$ scaling, where evaluations of the Hartree potential can be performed in a small subspace rather than across the entire simulation cell. However, the long range nature of the orthogonalised Kohn-Sham orbitals conflicts with the requirement to localise the interactions of $\rho(\mathbf{r}, \mathbf{r}')$ [90]. However, by relaxing the orthogonality condition, one can use localised, non-orthogonal Wannier functions to represent the Kohn-Sham orbitals through a unitary transformation. In the next section, we will describe how this was achieved in the framework of ONETEP.

2.4.2 Linear Scaling DFT in ONETEP

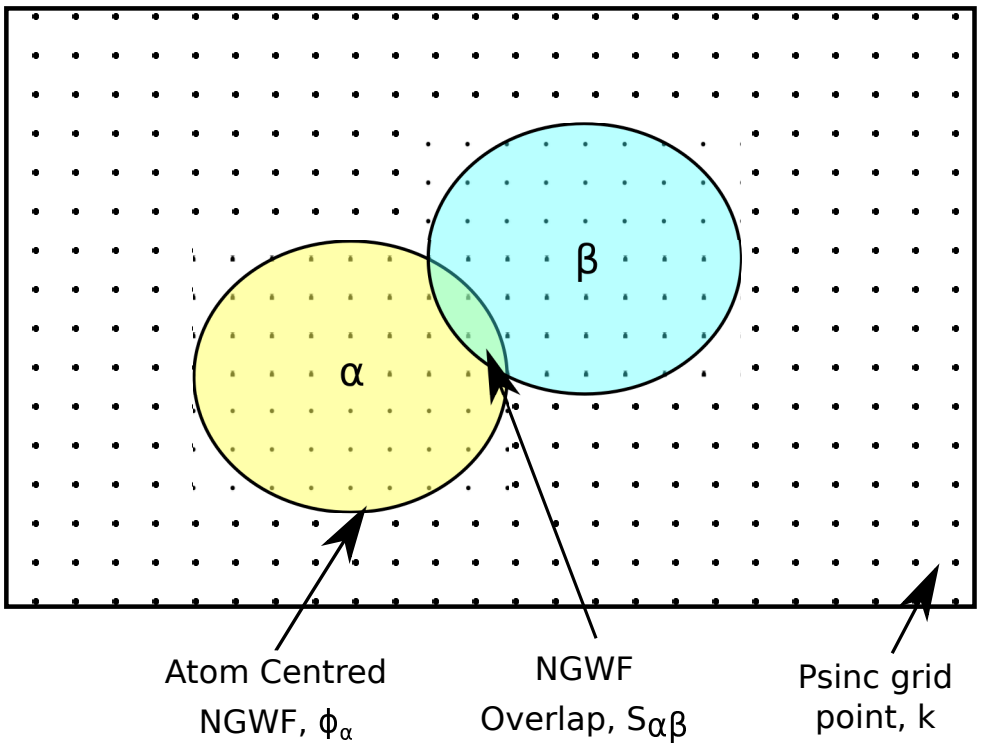

FIGURE 2.1: Demonstrates the NGWF basis as found in ONETEP. The simulation cell is constructed from an underlying set of localised psinc functions $D_k(\mathbf{r})$, which only take values within the sphere defined by r_c^α.

ONETEP [91] uses a set of atom-centred, localised, non-orthogonal Gannier wave functions (NGWFs), which allows for the sparse representation of matrix quantities such as the Hamiltonian and orbital overlap. This enables the use of sparse matrix algebra, which yields improvement in scaling over traditional plane wave codes that require cubically scaling dense diagonlisation procedures to solve the KS eigenvalue problem. NGWFs are localised by specifying a cut-off radius, r_c, for which $\rho(\mathbf{r},\mathbf{r}') = 0$ beyond $|\mathbf{r}-\mathbf{r}| > r_c$. By doing so, only interactions between atom centred functions within a limited range must be considered. This contrasts the traditional plane wave approaches, where computations are required between points which nominally do not interact with one another.

In the NGWF framework, the density is re-casted into an equivalent form,

$$\rho(\mathbf{r},\mathbf{r}') = \phi_\alpha(\mathbf{r})\boldsymbol{K}^{\alpha\beta}\phi^*_\beta(\mathbf{r}'), \tag{2.53}$$

where ϕ_α are the NGWFs and $\boldsymbol{K}^{\alpha\beta}$ represents the density kernel,

$$\boldsymbol{K}^{\alpha\beta} = \sum_n \boldsymbol{M}^\alpha_i f_i \boldsymbol{M}^{*\beta}_i, \tag{2.54}$$

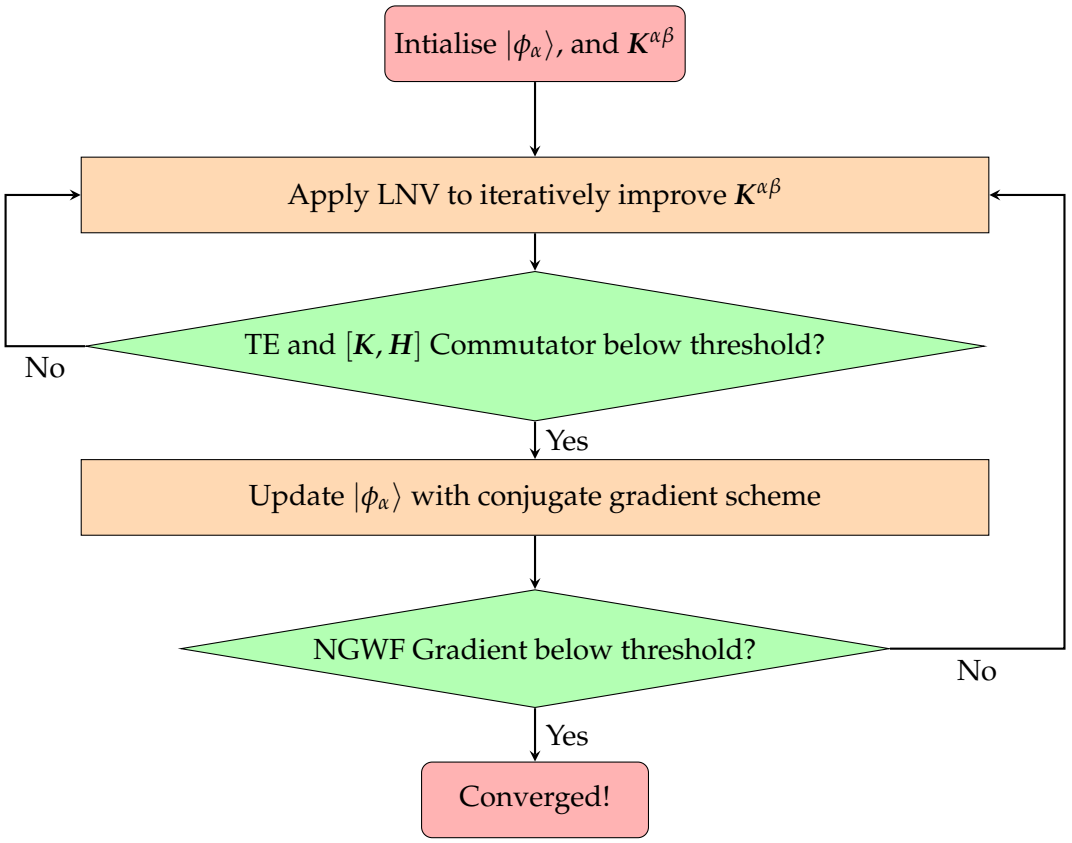

FIGURE 2.2: The two loop total energy optimisation scheme used in ONETEP.

where f_i represents the occupancy of state i, and $\boldsymbol{M}_i^\alpha$ are the rotation matrices which map the NGWFs into the orthogonal KS representation,

$$\phi_i(\mathbf{r}) = \phi_i(\mathbf{r})\boldsymbol{M}_i^\alpha. \tag{2.55}$$

The NGWFs are then strictly localised within a sphere defined by r_c, and expanded in terms of periodic psinc functions centred on a grid of defined psinc centres,

$$\phi_\alpha(\mathbf{r}) = \sum_k D_k(\mathbf{r})C_{k,\alpha}, \tag{2.56}$$

where D_k represents the psinc function at $\mathbf{r}$, $C_{k,a}$ are a set of expansion coefficients and k is the index defining the grid point in question. This scheme is demonstrated in Figure 2.1. Furthermore, the quality of the calculation can be tuned through two parameters, the cutoff radius defining the localisation sphere of each function and the spacing of the psinc grid, which is analogous to the kinetic energy cutoff found in traditional PW approaches. The minimisation of the total energy in ONETEP is achieved through a two-loop scheme. In the inner loop, $\boldsymbol{K}^{\alpha\beta}$ is iteratively improved in a fixed basis of NGWFs, while the outer loop updates the NGWFs through a conjugate gradient scheme. The minimisation scheme is shown in Figure 2.2. Convergence is reached for the inner loop when the commutator $[\boldsymbol{K}^{\alpha\beta}, \boldsymbol{H}_{\alpha\beta}]$ is below a specified threshold, and for the outer

loop when the NGWF RMS gradient is below 2×10^{-6} Ha. For insulating systems, the energy minimisation scheme for $K^{\alpha\beta}$ is achieved using the Li, Nunes and Vanderbilt (LNV) scheme [92]. However, for systems with small band gaps, such as metals, the idempotency condition cannot be satisfied. This is because many electronic minimisation schemes for metals require fractional occupancies for stable convergence. Therefore, the inner loop must be reformulated in terms of finite-temperature DFT, which is achieved in the ensemble DFT framework [83] through iterative improvement of the Hamiltonian. This scheme will be described in detail in Chapter 4.

By truncating the NGWFs and the density kernel beyond certain radii, we can perform DFT calculations with computational effort, which is reduced to $O(N)$ scaling as a function of the number of atoms. This takes practical advantage of the principle of near-sightedness of electronic matter [88]. Thus, ONETEP allows calculations with much larger scale than conventional DFT approaches, which is critical for systems where a large number of atoms are required. For example, ONETEP was used to obtain accurate energies for statistical block copolymers up to thousands of atoms [93]. These copolymers consist of donor and acceptor blocks, and are under ongoing investigation for applications in organic photovoltaics. Using conventional cubic scaling approaches, only short-chain polymers can be feasibly calculated [94], meaning studies regarding the ratio and composition of donor and acceptor blocks are limited. However, the linear scaling formulation in ONETEP enabled the calculation of long chain oligomers [93]. This study showed that varying ratios of donor and acceptor blocks has large effects on the electronic properties of the polymer, leading to variations of the HOMO-LUMO gap on a scale of approximately 0.1 eV.

2.4.3 Pseudopotentials and the PAW Method

In principle, the Kohn-Sham methods above can be applied to obtain the entire electronic structure of the system for all electrons. However, in practical calculations, it is advantageous to represent the core states using an approximate, smooth potential. There are two reasons within the framework of plane wave codes to make this approximation: a) most codes scale N^3 with the number of electrons included in the system, which is especially punishing for higher atomic number elements with many core states, b) the potential of the core states varies dramatically within a small span of space - as these quantities must be represented on a grid with finite spacing, one must use increasingly small grid sizes to represent the density of core states. In the plane wave formalism, this in effect corresponds to the increase in basis set size/the increase in the kinetic energy cut-off. Both of these factors mean in production calculations, explicit calculations including the core electrons become unfeasibly computationally expensive for larger systems.

In chemical applications, one is primarily interested in studying bond interactions, which occur predominantly between electrons within the valence states. However, the influence of neighbouring atoms on the core states is relatively small, meaning the all electron calculations dramatically increase computational costs in exchange for negligible increases in accuracy for most chemical studies. The pseudopotential provides a cheaper alternative in two ways. Firstly, the external potential due to the nucleus is described through its effective ionic charge ($Z_{eff} = Z_I - N_c|e|$, where N_c is the number of core electrons), which emulates the screening of the ionic charge by the core electrons without their inclusion in the total energy evaluation. By projecting out the core states, we significantly reduce the number of computations required to solve the N^3 scaling eigenvalue problem. Secondly, the valence states ($|\psi_v\rangle$) are converted into a set of smooth pseudo-waves ($|\psi_{ps}\rangle$). This simplifies the potential of the core regions into a smooth function, which reduces the grid spacing required to store plane waves.

However, the orthogonality condition of the valence states with respect to the core states must be maintained, ensuring that the eigenvalues of the valence states do not converge to the core state. Furthermore, ordering of the eigenstates must be maintained according to the *aufbau* principle. This is achieved through the Orthoganlised Plane Wave (OPW) method [95, 96], where the wavefunctions of the valence state are orthogonalised with respect to an expanded set of projection operators.

$$|\psi_{ps}\rangle = |\psi_{val}\rangle - \sum_{i}^{N_{core}} \left|\psi_c^i\right\rangle \left\langle\psi_c^i\middle|\psi_{val}\right\rangle, \tag{2.57}$$

where the latter terms enforces orthogonality through the relation,

$$\left\langle\psi_c^i\middle|\psi_{val}\right\rangle = -B_n^i, \tag{2.58}$$

where B_n^i is a coefficient applied to $|\psi_c^i\rangle$. Overall, these relations yield a modified set of valence wavefunctions, which contain either no or significantly fewer nodes in the core regions. Underpinning the pseudopotential method is the frozen core approximation, where the pseudowave functions are approximately orthogonal to the valence state, regardless of the chemical environment. In effect, this allows one to use a fixed set of core states and projectors, without requiring re-orthogonalisation of the core and valence states. Improvements in the OPW formalism were made by imposing the norm-conservation condition [97], where the charge of $|\psi_{ps}\rangle$ within a specified core region ($r_c < |\mathbf{r} - \mathbf{R_a}|$ for atomic centre a) are set equal to the all-electron wavefunction ($|\psi_{AE}\rangle$).

$$\int_{r_c<|\mathbf{r}-\mathbf{R_a}|} |\psi_{ps}(\mathbf{r})|^2 r^2 dr = \int_{r_c<|\mathbf{r}-\mathbf{R_a}|} |\psi_{AE}(\mathbf{r})|^2 r^2 dr \tag{2.59}$$

Imposing this condition ensures that the electrostatic potential in the valence region matches the all-electron form without resorting to renormalisation of the charge density.

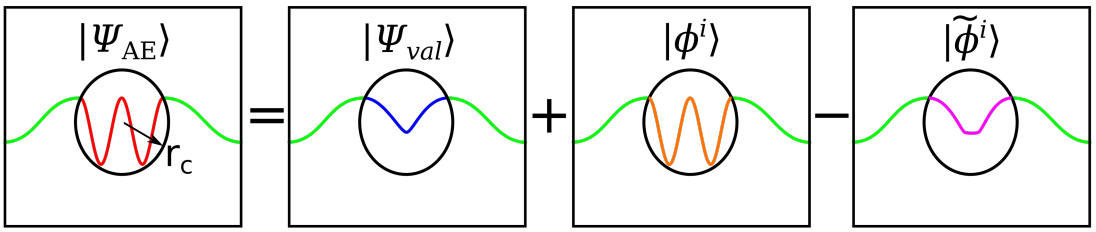

FIGURE 2.3: Simplified diagram of the PAW formalism. The unaltered valence wavefunctions $|\psi_{val}\rangle$ through a linear transformation operator $\hat{T}$ are transformed to the all-electron form through the addition of a set of all-electron partial waves $|\phi^i\rangle$ and the subtraction of auxiliary smooth waves $|\tilde{\phi}^i\rangle$ within the augmentation sphere, r_c. The partial waves match $|\psi_{val}\rangle$ outside the augmentation sphere.

However, for transition metals, the construction of pseudopotentials which follow the norm conservation is difficult, where the projections from $|\psi_{val}\rangle$ to the $|\psi_{ps}\rangle$ state result in still large spatial variations in the core regions. Furthermore, by generalising the core region into a smooth potential, electron density information close to the nuclei is lost. Later developed techniques such as the Projector Augmented Wave (PAW) approach [98, 99] form a bridge between all electron and pseudpotential methods by performing a linear transformation which effectively splits the valence wavefunction into an augmentation sphere which, matches the all-electron representation, and the valence region, which matches the smooth wavefunction, approximating the all-electron wavefunction ($|\psi_{AE}\rangle$) (Figure 2.3).

Formally, this linear transformation is described as,

$$|\psi_{AE}\rangle = |\psi_{val}\rangle + \sum_i (\left|\phi^i\right\rangle - \left|\tilde{\phi}^i_{val}\right\rangle) \left\langle p^i \middle| \psi_{val} \right\rangle , \tag{2.60}$$

for an augmentation region (r_c), where $\left|\phi^i\right\rangle$ are the all-electron partial waves and $\left|\tilde{\phi^i_{val}}\right\rangle$ are smooth auxiliary wavefunctions. The former are defined through the eigenstates of an all-electron calculation for an isolated atom, and the former is defined as a smooth extrapolation of the valence state into the core region. Both match the wavefunction of the unaltered valence states outside the augmentation sphere defined by the total energy calculation ($|\psi_{val}\rangle$). Within the augmentation sphere defined as $r_c \leq |\mathbf{r} - \mathbf{R}|$, $|\psi_{val}\rangle$ is transformed into its all-electron equivalent,

$$|\psi_{AE}\rangle = \sum_i \left|\phi^i\right\rangle \left\langle p^i \middle| \psi_{val} \right\rangle . \tag{2.61}$$

Outside of the augmentation sphere ($r_c \geq |\mathbf{r} - \mathbf{R}|$) where $\left|\phi^i\right\rangle - \left|\tilde{\phi}^i\right\rangle = 0$, ψ_{val} matches smooth valence states, in effect leaving these unaltered in the bonding regions,

$$|\psi_{AE}\rangle = \sum_i \left|\phi^i\right\rangle \left\langle p^i \middle| \psi_{val} \right\rangle = |\psi_{val}\rangle \,. \tag{2.62}$$

In effect, the summation of equation 2.60 calculates the difference between the unaltered valence wave functions and the all-electron form both outside and within the augmentation region. Core states are calculated in a similar way. By retaining the spatial variation of the all-electron wavefunctions in the core regions, the PAW formalism allows one to access properties such as the all-electron density and Hartree potential. Furthermore, as total energy evaluations within the augmentation regions are calculated separately within a specially defined radial grid, calculations in the valence region can be performed with reduced grid spacing. This allows for less computationally expensive total energy evaluations without the loss of accuracy.

We will also briefly mention Ultrasoft Pseudopotentials (USPPs) [100], which were developed in the same decade as PAWs and provide similar numerical advantages for total energy evaluations. These pseudopotentials are still commonly used throughout a range of plane wave codes and are in principle mathematically equivalent to PAWs [101]. However, as PAWs are used primarily throughout the work, the reader is referred to References [100, 101] for a more detailed discussion of USPPs.

2.4.4 Calculation of Forces and Geometry Relaxation

Beyond the electronic structure, *ab initio* simulations are often used to obtain structural information for molecules and crystals. Accurately describing the atomic positions of the system under study is vital for calculating binding strengths, crystal lattice parameters, and thermodynamic properties. However, without *a priori* knowledge of the precise atomic co-ordinates, it is necessary to perform geometry relaxation calculations to obtain the most energetically favoured structure. The goal of the geometry optimisation procedure is to traverse the potential energy surface (PES) by changes in the atomic positions, for which the local minima in the region of the starting geometry can be found. This is achieved through iterative procedures which minimise the energy in terms of well-conditioned changes in the atomic coordinates.

Evaluating the changes in energy with respect to the ionic positions (forces) is paramount in finding the structural local minima. Within the framework of single-particle wavefunctions, it can be shown that forces need only be calculated for ionic terms within the Hamiltonian. This is known as the Helmann-Feynman theorem [102], where the

explicit derivative of the energy in terms of the atomic position of atom L ($\mathbf{R}_L$) is expressed as,

$$\mathbf{F}_L = -\frac{dE[\psi_i(\mathbf{r}), \mathbf{R}_L]}{d\mathbf{R}_L}. \tag{2.63}$$

However, the wavefunctions are implicitly dependent on the ionic positions, meaning the full force term requires an expansion in terms of partial derivatives,

$$-\frac{dE[\psi_i(\mathbf{r}), \mathbf{R}_L]}{d\mathbf{R}_L} = -\frac{\partial E[\psi_i(\mathbf{r}), \mathbf{R}_L]}{\partial \mathbf{R}_L} - \frac{\partial E[\psi_i(\mathbf{r}), \mathbf{R}_L]}{\partial \psi_i(\mathbf{r})}\frac{\delta \psi_i(\mathbf{r})}{\delta \mathbf{R}_L}, \tag{2.64}$$

where the latter term includes the implicit dependency of the wavefunctions on atomic positions. This is known as the Pulay force [103]. However, if the iterative electronic minimisation scheme is successful, a stationary point of the energy is found such that,

$$\frac{\partial E[\psi_i(\mathbf{r}), \mathbf{R}_L]}{\partial \psi_i(\mathbf{r})} = 0. \tag{2.65}$$

As a result, for a complete basis set, $\mathbf{F}_L$ can be described in terms of the explicit derivative of the Hamiltonian. $\mathbf{F}_L$ is then calculated as the sum of the forces from the ionic cores (known as the Ewald force) and forces arising from the local and non-local energy components of the pseudopotentials. However, there are cases where the Helmann-Feynman theorem is weaker, such as where the basis set is not completely described. As the framework of ONETEP is based on a set of finite localised, atom centred basis functions dependent on the atomic position, we require an analytical expression of the Pulay forces to obtain an accurate description of $\mathbf{F}_L$ [104].

The geometry minimisation procedure itself is based on the quasi-Newton Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm [105]. This approach is related to the Newton class of numerical methods, which minimise an objective function (i.e., the energy $E[x_k]$) through successive updates of the atomic positions,

$$x_{k+1} = x_k - \alpha_k {\mathbf{H}_\mathbf{k}}^{-1} \nabla E[x_k], \tag{2.66}$$

where x_k is a generic descriptor of the structure co-ordinates at step k, α_k is the step length, ${\mathbf{H}_\mathbf{k}}^{-1}$ is the inverse of the Hessian $\nabla^2 E[x_k]$, and $\nabla E[x_k]$ is the vector of forces. Generally, α is defined through a line search, which sweeps through a range of step sizes, calculates $E[x_k]$ for each trial step length, and finds the optimal α_k through a polynomial fit. However, this scheme requires the construction and the inversion of the Hessian for every iteration, which is prohibitively expensive [106]. The BFGS algorithm circumvents these computations by performing iterative updates of ${\mathbf{H}_\mathbf{k}}^{-1}$ in addition to x_k, meaning only the forces are required to compute the iterative step of Equation 2.66. However, the memory usage of the BFGS scheme scales $O(N^2)$ with system size, which can be problematic for larger systems. Modified schemes such as Limited Memory

BFGS (L-BFGS) [107] reduce the memory scaling to $O(N)$ by implicitly representing $\mathbf{H_k}^{-1}$ as a set of m vectors, as opposed to a dense matrix.

The performance of the BFGS scheme can be improved by constructing the Hessian from the bulk properties empirically derived from the system [108]. Further improvements can be attained by applying a preconditioner to the Hessian. For example, the exponential preconditioner, which improves the conditioning of the Hessian by setting the frequencies between atoms beyond a cut-off radius r_{cut} to 0 [106]. This technique drastically improves the convergence rate for the BFGS algorithm, halving the number of iterations required to achieve converged geometries for metallic systems.

2.5 Implicit Solvent Models

In Chapter 1, we described how implicit solvent models provide a computationally inexpensive method for calculating the thermodynamics of solvation. This section details how this is achieved within quantum chemical codes and the specific details of the implementation in ONETEP prior to the developments in this thesis.

The implicit solvent model divides the free energy of solvation (ΔG_{solv}) into electrostatic phenomenon (polar) calculated through the bulk dielectric constant of the solvent (ΔG_{pol}), and non-electrostatic contributions such as the cavitation free energy and van der Waals energies (ΔG_{npol}). By doing so, the configurational degrees of freedom associated with the solvation shell are taken as an average, meaning the solvent is no longer represented atomistically. As a result, the scale of solvation calculations are drastically reduced.

The response to the bulk dielectric medium in quantum chemical codes is commonly calculated through a modified form of the Poisson-Boltzmann equation,

$$\nabla^2\phi_{vac}(\mathbf{r}) = -4\pi n_{vac}(\mathbf{r}) \rightarrow \nabla\cdot(\epsilon(\mathbf{r}))\nabla\phi_{solv}(\mathbf{r}) = -4\pi n_{solv}(\mathbf{r}), \tag{2.67}$$

where $\phi(\mathbf{r})$ is the electrostatic potential, $\epsilon(\mathbf{r})$ is the dielectric defined at position $\mathbf{r}$, and the subscripts denote whether the quantities are defined in solvent or vacuum. As $\epsilon(\mathbf{r})$ varies throughout the simulation cell, the solution to the Poisson equation can no longer be easily solved in reciprocal space as electrostatic potential calculations are calculated in a framework of a vacuum dielectric constant (where $\epsilon(\mathbf{r}) = 1$). As such, finite difference approaches in real space are applied [109], such as DL_MG multigrid Poisson-Boltzmann solver library [110]. However, later approaches such as those of *Andreussi et al.* [111] separate the total charge density into the solute ($n_{solute}(\mathbf{r})$) and the polarised density that responds to the dielectric ($n_{polar}(\mathbf{r})$). As the response to the dielectric is included in $n_{polar}(\mathbf{r})$, the standard vacuum form Poisson equation in reciprocal space can be applied to $n_{polar}(\mathbf{r})$ and $n_{solute}(\mathbf{r})$ separately to obtain $\phi_{tot}(\mathbf{r})$.

The overall ΔG_{pol} is calculated as the difference of total energy expressed in terms of $n_{solv}(\mathbf{r})$ and $n_{vac}(\mathbf{r})$.

$$\Delta G_{pol} = E[n_{solv}(\mathbf{r})] - E[n_{vac}(\mathbf{r})]. \tag{2.68}$$

In ONETEP, both quantities can be calculated in either fully periodic/open boundary conditions within the DL_MG Poisson Boltzmann solver.

One further consideration made in the ONETEP implementation of the implicit solvent is the treatment of the ionic charges. The point charge representation of the atomic cores leads to singularities in the solutions of real-space Poisson-Boltzmann solvers [112]. However, this can be overcome by replacing the point charge with Gaussian smeared charges expressed as,

$$\rho_I(\mathbf{r}) = -Z_I \pi^{-\frac{3}{2}} \exp\left\{ \left(-\frac{|\mathbf{r} - \mathbf{R}_I|^2}{\sigma_I^2} \right) \right\}, \tag{2.69}$$

for atom I, where σ_I^2 is the width of the Gaussian smearing and Z_I is the atomic charge. This allows for better representation of the ionic charges on the real-space grid, which leads to efficient solutions to the Poisson-Boltzmann problem.

The non-polar contribution (ΔG_{npol}) can be calculated in a variety of ways. The initial formulations used scaled particle theory (SPT) approximations of the free energy of cavitation (ΔG_{cav}) combined with the free energy of dispersion between the solute and the solvent (ΔG_{disp}) [113]. SPT is derived from the reversible work required to create a spheric cavity of volume V_{cav} based on the probability of displacing a number of solvent molecules through the ratio of the radii of the solute and the solvent. This results in the expression,

$$\frac{\Delta G_{cav}}{kT} - \ln(1-y) + \left(\frac{3y}{(1-y)} \right) R + \left[\frac{3y}{(1-y)} + \frac{9}{2} \left(\frac{y}{(1-y)} \right)^2 \right] R^2, \tag{2.70}$$

where y is the reduced number density of solvent molecules ($y = (4/3)\pi R_{solv}^3 n_{solv}$), R_{solv} is the radius of the solvent and n_{solv} is the density of the solvent. R is the ratio between the solvent radii of the solute and the solvent ($R = (R_{sol}/R_{solv})$). However, the value of ΔG_{cav} is highly sensitive to the input values of n_{solv} and R_{solv}. Therefore, the large variation in experimental values for the radius of water (1.36-1.46 Å) produces large errors in the calculation of ΔG_{cav}.

Later approximations of ΔG_{npol} use the Solvent Accessible Surface Area (SASA) approach, expressed as,

$$\Delta G_{npol} = \gamma S, \tag{2.71}$$

where γ is the surface tension of the solvent and S is the surface area of the dielectric cavity. However, this approximation systematically overestimates the nonpolar contributions to ΔG_{solv} [114], since the linear relationship in effect only calculates the

cavitation free energy. This ignores the dispersion-repulsion contributions of solvent, which have been observed to follow a similar linear relationship with respect to the surface area [115]. Therefore, greater precision is obtained in the implicit solvent model of ONETEP by applying a simple scaling factor of 0.281 to γ, which is obtained by a linear regression of the terms $\Delta G_{dis-rep}$ and ΔG_{cav} with respect to saturated hydrocarbons of increasing length and surface area. However, it has been noted that the SASA model, even with reparameterization, shows overall poor correlation with ΔG_{solv} for a range of different species [116–118] and has shown poor transferability even between different configurations of the same protein systems [119]. Alternative forms of ΔG_{npol} will be discussed in Chapter 3.

Throughout the electronic structure minimisation procedure, the value of $\rho(\mathbf{r})$ is updated at every step (the self-consistent (SC) cavity). Therefore, the charge density-based cavity function evolves throughout the energy minimisation calculation. Outside the energetic minimum, where $\frac{E[n_{solv}(\mathbf{r})]}{d\epsilon(\rho_{solv}(\mathbf{r}))} = 0$, additional terms must be introduced to the energy gradient to explicitly treat the energetic changes with respect to changes in the dielectric cavity,

$$\frac{dE[n_{solv}(\mathbf{r})]}{dn_{solv}(\mathbf{r})} = \phi_{solv} - \frac{1}{8\pi}|\nabla\phi_{solv}|^2 + \frac{\delta\Delta G_{npol}}{\delta n_{solv}(\mathbf{r})}, \tag{2.72}$$

where the second term corresponds to $\frac{E[n_{solv}(\mathbf{r})]}{d\epsilon(\mathbf{r})}$. However, as this term is small, the values represented on the grid are indistinguishable from numerical noise. As such, without representations of this term on very fine grids, the second term introduces numerical instability to the energy minimisation calculation. However, a finer grid spacing intrinsically leads to greater computational expense, and although defect correction schemes significantly reduce associated errors, grid spacings 2 to 3 times of standard ONETEP calculations were required for reliable convergence [114]. However, these errors can be overcome by fixing the cavity function (referred to the fixed cavity throughout) to the converged density of a vacuum calculation, $\epsilon(n_{vac}(\mathbf{r}))$. Especially for neutral molecules, the charge density only changes marginally in response to the dielectric, which means that the cavity shape constructed with $\epsilon(n_{vac}(\mathbf{r}))$ can adequately represent $\epsilon(n_{solv}(\mathbf{r}))$. This leads to only a minor increase in the RMS error vs. the SC cavity for a test set of neutrals, cations, and anions (15.9 kJ mol^{-1} vs. 17.1 kJ mol^{-1}) and leads to no deterioration in the correlation coefficient (r=0.83) [114]. Both values compare favourably compared to the Solvent Model based on Density (SMD) method, which includes a highly parameterized ΔG_{npol} term (RMS error: 14.2 kJ mol^{-1}, $r = 0.87$).

Construction of the dielectric cavity function $\epsilon(\mathbf{r})$ must also be considered. The original implementation of the implicit solvent model in ONETEP [114] follows the formulation of Fattebert, Gygi and Scherlis [120–122], where the dielectric cavity is described in terms of the electron density at the point $\mathbf{r}$, which defines a smooth switching function

from the vacuum to the bulk dielectric constant ($1 \rightarrow \epsilon_\infty$),

$$\epsilon(\rho(\mathbf{r})) = 1 + \frac{\epsilon_\infty - 1}{2}\left(1 + \frac{1 - (\rho(\mathbf{r})/\rho_0)^{2\beta}}{1 + (\rho(\mathbf{r})/\rho_0)^{2\beta}}\right), \tag{2.73}$$

where ρ_0 is the electron density isocontour where $\epsilon_\infty/2$ and β describes the slope of the switching function. The values ρ_0=0.0035 $\|e\|a^{-1}$) and $\beta = 1.3$ were selected with a parameter sweep to minimise the error in ΔG_{solv} for a range of neutral, cationic, and anionic species [114]. The minimal parameterization required for the Fattebert, Gygi and Scherlis (FGS) approach makes it a compelling cavitation model. However, the isodensity model has several shortcomings: a) significant ΔG_{solv} errors for anions with the default parameterization of ρ_0 and b) as shall be discussed in Chapter 3 and 5, the inability to simultaneously capture properties of metallic surfaces and light elements. These problems are a consequence of defining the cavity function for the whole system with a single value, ρ_0. For species such as anions and metals, the cavity must be adjusted while maintaining the default parameterization for neutral/light elements. In these cases, the single parameter approach means one must sacrifice accuracy by choosing either a compromise value of ρ_0, or ignoring the need for species specific cavities entirely.

To address this, the soft sphere cavity model of *Fisicaro et al.* [123] was implemented in ONETEP as part of our investigation into the metal/solvent interface. By defining the cavity function as a set of atom centered spheres, one can flexibly adjust the cavity for species which require special consideration. This model defines the cavity as a set of interlocking spheres with a set of distance-dependent h functions:

$$\epsilon(\mathbf{r}_i, \mathbf{R}_i) = (\epsilon_\infty - 1)\{\prod_i h(\{\xi\}; \|r - R_i\|)\} + 1, \tag{2.74}$$

where $\|r - R_i\|$ defines the distance to a point at position $\mathbf{r}$ from the atomic centre R_i, ϵ_∞ the dielectric permittivity of the bulk solvent, and ξ the parameters for a specific atomic centre. The h functions define a set of atom centred error functions (erf) smoothly varying from 0 to 1,

$$h(\mathbf{r_i}, \Delta, \|\mathbf{r} - \mathbf{R_i}\|) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\|r - R_i\| - r_i}{\Delta}\right)\right], \tag{2.75}$$

where the values of Δ define how quickly the dielectric function varies from bulk to vacuum permitivity values and r_i the distance defining the midpoint of the dielectric function ($\frac{\epsilon_\infty}{2}$) for the atomic centre i.

In the scheme of *Fisicaro et al.*, the parameters r_i and Δ were optimised using a parameter sweep over a test set of 13 molecules [123]. The cavity radii was varied by multiplying the original set of r_{vdW} by a scaling factor, f, where the van der Waals' radii were derived from the Universal Force Field (UFF) [124]. Values of the Minimum

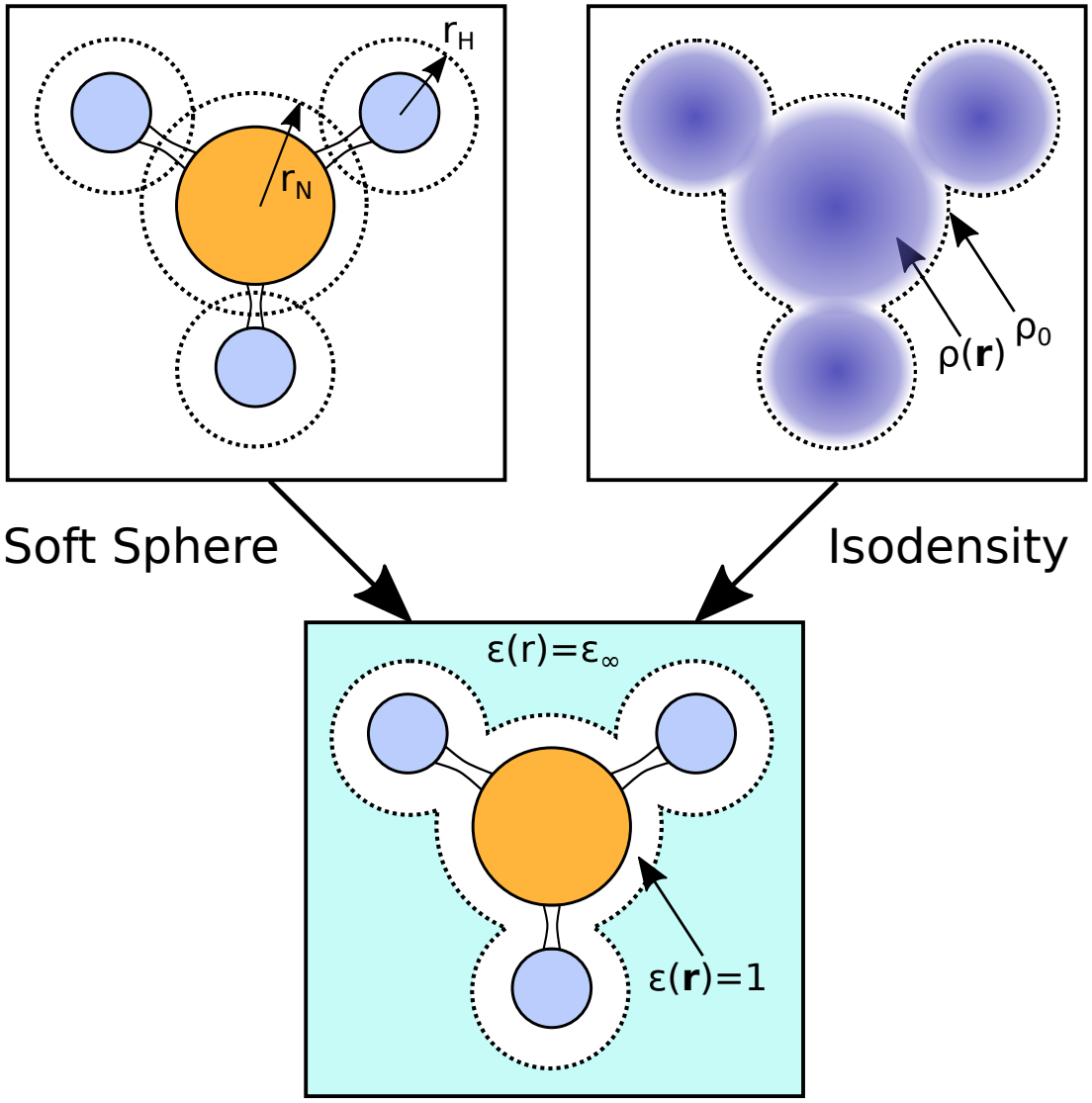

FIGURE 2.4: The construction of the dielectric cavity in terms of interlocking spheres based on the van der Waals' radii from the atomic centre (Soft Sphere) or based on isocontours of charge ρ_0 (charge based cavity).

Absolute Error (MAE) for ΔG_{solv} were measured relative to the Minnesota Solvation Database [125], where the lowest errors were measured with $f = 1.12$, where the MAE was 4.8 kJ mol^{-1}. Greater accuracy was obtained by substituting the atomic radii of problematic elements such as N, for which the original UFF model does not properly take into account the interaction of the N lone-pair with the H within H_2O hydrogen bonding. The soft sphere model compared favourably to charge density based cavity models for neutral molecules when extended to a larger, 274 molecule test set (MAE: 4.69 vs. 4.77 kJ mol^{-1}). Furthermore, the soft sphere model is notably more accurate for charged molecules, especially anions (MAE: 12.38 vs. 23.17 kJ mol^{-1}). The soft sphere model therefore provides a simple and accurate approach to constructing the dielectric cavity, without a loss of accuracy. Furthermore, the soft sphere model allows for the parameterization of species which require different dielectric cavity sizes within

the same simulation cell. As such, this model was implemented in ONETEP as a key part of the work presented, as described in Chapter 3.

Chapter 3

Extending the Implicit Solvent Model

As discussed in Chapter 2, the implicit solvent model provides a computationally inexpensive method for calculating the free energy of solvation, ΔG_{solv}. The original implementation in ONETEP [114] can compute ΔG_{solv} for neutral organic molecules with high accuracy, yielding RMS errors of only 7.5 kJ mol^{-1} with respect to experimental data. Combined with the linear scaling formulation of ONETEP, this allows computationally tractable calculations of solvation energies for larger molecules, which have been applied in calculations of the binding energies for assays of large protein-ligand complexes [112, 126].

However, the accuracy and flexibility of the implicit solvent model could be extended further in two ways. First, the SASA model for the calculation of ΔG_{npol} is known to correlate poorly with explicit solvent values (correlation coefficients of r=0.4 for water only and r=0.25 for a range of solvents [116]). Improvements in correlation were found by introducing a linear term for ΔG_{npol} based on the volume (SASA+SAV). We will discuss the procedure used to parameterize the SASA+SAV model, and discuss the benefits provided by describing surface tension (γ) as a free parameter.

Secondly, although the charge based cavity model allows for the construction of a dielectric cavity with minimal parameterization, it is sometimes desirable to maintain flexibility, so the cavity function can be defined separately for each atomic centre. This is especially important for systems with charged fragments or chemically distinct species (i.e. transition metals and light organics), where no single isodensity contour would produce an optimal cavity function for different parts of the same system.

As was introduced in Section 2.5, the soft sphere model by Fisicaro [123] provides a method for representing the atom centred, distance based cavity function compatible with electronic structure codes. Therefore, we will describe our implementation of this

model into the ONETEP code as part of this work. The soft sphere model will be especially useful in later chapters, where the ability to simultaneously parameterize the dielectric cavity for light organics and metallic atoms allows for the calculation of the free energy of adsorption in the aqueous phase.

In addition, we will determine the quality of the force calculations within the implicit solvent model. In principle, the force terms associated with the change in the dielectric cavity function with respect to position atomic position should be small. As such, analytic force terms for the cavity were not implemented into ONETEP. However, this is an untested assumption which requires validation for a range of systems with varying charges. To accomplish this, we will validate the force assumption by performing a set of finite difference force tests, which in principle gives the exact change in energy with respect to atomic position. This will be performed with both the fixed and self-consistent cavity update of the charge-based dielectric cavity method, where the latter should, in principle, yield fully converged forces without extra analytical terms in accordance with the Hellman-Feynman principle [102]. Furthermore, to facilitate future work conducted with ONETEP, we also extended these tests to systems using an electrolyte model [127, 128].

3.1 Background

3.1.1 Implementation and Parameterization of the Soft Sphere Cavity Model

As detailed in Section 2.5, the soft sphere model by Fisicaro [123] provides a dielectric cavity function ($\epsilon(\mathbf{r})$) which is defined in terms of the distance from the atomic centres,

$$\epsilon(\mathbf{r}_i, \mathbf{R}_i) = (\epsilon_\infty - 1)\{\prod_i h(\{\zeta\}; \|r - R_i\|)\} + 1. \tag{3.1}$$

This is a product of interlocking distance based functions,

$$h(\mathbf{r_i}, \Delta, \|\mathbf{r} - \mathbf{R_i}\|) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\|r - R_i\| - r_{vdw,i}}{\Delta}\right)\right], \tag{3.2}$$

, both of which are fully defined in Section 2.5.

Implementing this model requires calculations of h, where the products of Equation 3.2 are defined for each atomic centre across the underlying grid points of the simulation cell. The van der Waals radii ($r_{vdw,i}$) defined for each atomic centre are obtained from the van der Waals' model by *Alvarez* [129]. In principle, this set of radii should provide benefits over the set of Universal Force Field (UFF) radii used in the original implementation of the soft sphere method [130], as *Alvarez* used a distance histogram approach,

which measured the distribution of atomic distances from an extensive set of crystallographic structural data. In contrast, the UFF approach calculates the van der Waals' distance from the pairwise Lennard-Jones (LJ) potential [124], which ignores the effects of fine electronic features such as lone pairs. Furthermore, *Alvarez* uses the potential of oxygen as a probe for determining the van der Waals' radius, which provides a good approximation for the distance of H_2O from the atomic centre.

3.1.2 Implementation and Parameterization of the SASA+SAV ΔG_{npol} Term

In the SASA+SAV model, we describe non-polar free energy change as,

$$\Delta G_{npol} = \alpha * \gamma S + pV, \tag{3.3}$$

where α is a scaling factor, which allows the surface tension γ to be treated as a flexible parameter, V is the volume of the cavity, and p is the solvent pressure. Further models such as the dispersion model of *Wagoner* (SSA+SAV+WCA) [116] introduce Lennard-Jones (LJ) type attractive interactions to the cavity. However, the benefits provided by SSA+SAV+WCA over SAV are not particularly large (r=0.95 vs. 0.94), and the majority of the improvements in statistics are derived from the additional volume terms. Therefore, we have only implemented the volume based term in the solvent model.

In both the soft sphere and electron density-based cavity models, the volume was calculated using the product of the $h(\{\xi\}; \|r - R_i\|)\}$ inverted, such that the function switches from 1 inside the cavity to 0 outside. This can be obtained by the expression,

$$\theta[\epsilon(\mathbf{r})] = \frac{\epsilon_\infty - \epsilon(\mathbf{r})}{\epsilon_\infty - 1}. \tag{3.4}$$

The quantum volume is therefore defined as,

$$V[\epsilon(\mathbf{r})] = \int d\mathbf{r}\theta[\epsilon(\mathbf{r})]. \tag{3.5}$$

As the soft sphere model is formed from continuously differential spherical functions, the surface value can be computed in terms of the gradient of the dielectric cavity function [123],

$$S[\epsilon(\mathbf{r})] = \frac{1}{\epsilon_\infty - 1} \int d\mathbf{r}|\nabla\epsilon(\mathbf{r})|. \tag{3.6}$$

In contrast, electron density based cavity function is defined using an isosurface as opposed to a set of differentiable distance functions. Therefore, the original implementation of the quantum surface area is calculated through the difference in the surface area between two isosurfaces,

$$S = \int d\mathbf{r}\{\theta_{\rho_0 - \frac{\Delta}{2}}(\rho(\mathbf{r})) - \theta_{\rho_0 + \frac{\Delta}{2}}(\rho(\mathbf{r}))\} \times \frac{|\nabla(\rho(\mathbf{r})|}{\Delta}, \tag{3.7}$$

where $\theta_{\rho_0}(\rho(\mathbf{r}))$ is the cavity function (defined between 0 and 1, with 0.5 defined by isodensity ρ_0) for the electron density at position $\mathbf{r}$ ($\rho(\mathbf{r})$). The parameter Δ defines the cavity width between the two isosurfaces. However, as argued by *Andreussi et al.* [111], ρ_0 does not have a linear relation with the switching function, $\theta_{\rho_0}(\rho(\mathbf{r}))$. This means that the cavity functions defined by ρ_0 and $\rho_0 \pm \frac{\Delta}{2}$ have different topologies for an equivalent $\rho(\mathbf{r})$. Therefore, Equation 3.7 does not exactly calculate the difference between two isosurfaces for $\theta_{\rho_0}(\rho(\mathbf{r}))$ separated by $\frac{\Delta}{2}$, which results in approximately 20 % overestimates of the surface area.

This can be simply resolved by applying the parameter Δ to $\rho(\mathbf{r})$ itself, yielding:

$$S = \int d\mathbf{r}\{\theta_{\rho_0}(\rho(\mathbf{r}) - \frac{\Delta}{2}) - \theta_{\rho_0}(\rho(\mathbf{r}) + \frac{\Delta}{2})\} \times \frac{|\nabla(\rho(\mathbf{r})|}{\Delta}. \tag{3.8}$$

3.2 Methodology

3.2.1 Outline of the Test Set

To parameterize the ΔG_{solv} values of the soft sphere model, we followed the method of the original soft sphere paper [130] and the original density based cavity paper by *Andreussi et al.* [111]. Firstly, we parameterized ΔG_{pol} by sweeping through a range of f values (which uniformly scale the set of soft sphere radii), minimising the MAE with respect to another well parameterized code (in this case, the COMSO-SMD model [131] of NWChem [132]). This was to avoid the situation where in parameterizing ΔG_{solv} with respect to both ΔG_{pol} and ΔG_{npol}, a range of optimum values of f and γ can be found through a fortuitous cancellation of errors. Afterwards, we minimised the MAE of ΔG_{solv} with respect to changes in ΔG_{npol} by modifying the value of γ through an additional scaling factor α with a fixed value of f.

Parameter sweeps for f were performed with a small test set of 13 molecules, representative neutral molecules from the Minnesota Solvation Database [125]. Then, using a parameter sweep over α, overall ΔG_{solv} values are compared with the larger, 274 neutral molecule test set. Calculations were carried out using the Perdew-Burke-Ernzerhof (PBE) [133] functional, Open Boundary Conditions, 600 eV Kinetic Energy Cut-off, PAW pseudopotentials, and a 9.0 a_0 NGWF radii. NWChem calculations were performed using the PBE functional and the 6-31++G* Pople basis set [134].

3.2.2 Finite Difference Force Tests

Within the framework of ONETEP, the forces in vacuum for ion I are calculated as the derivative of the energy with respect to a change in position from $\mathbf{R}_I$ [135],

$$\mathbf{F}_I = -\frac{dE[K^{\alpha\beta}, \phi_\alpha(r)]}{d\mathbf{R}_I} = -\frac{\partial E}{\partial \mathbf{R}_I} - \sum_{\alpha\beta} \frac{\partial E}{\partial K^{\alpha\beta}} \frac{dK^{\alpha\beta}}{d\mathbf{R}_I} - \sum_{\alpha} \int \frac{\delta E}{\delta \phi_\alpha(\mathbf{r})} \frac{d\phi_\alpha(\mathbf{r})}{d\mathbf{R}_I} d^3r. \tag{3.9}$$

Within the two loop energy minimisation process, the partial derivatives of energy with respect to $K^{\alpha\beta}$ and $\phi_\alpha(\mathbf{r})$ tend to 0 at self-consistency, meaning the latter two terms are eliminated according to the Hellman-Feynman theorem [102]. As a result, forces can be expressed purely in terms of an explicit derivative of the ionic potential (Ewald forces) and local/nonlocal contributions from the pseudopotential.

In principle, we can obtain exact forces along a selected axis (x,y,z) atom I, using a finite difference method which can be calculated as,

$$\mathbf{F}_{I,x}^{FD} = \frac{E[\mathbf{R}_I + \Delta_x/2] - E[\mathbf{R}_I - \Delta_x/2]}{\Delta_x}, \tag{3.10}$$

where Δ_x represents a small step in the principle axis. The total magnitude of the force on atom I is therefore calculated as,

$$|F_I^{FD}| = \sqrt{\mathbf{F}_I^{FD,x} \cdot \mathbf{F}_I^{FD,x} + \mathbf{F}_I^{FD,y} \cdot \mathbf{F}_I^{FD,y} + \mathbf{F}_I^{FD,z} \cdot \mathbf{F}_I^{FD,z}}. \tag{3.11}$$

A separate total energy calculation is required for each move in the principle axis, meaning 6 single point calculations are required to obtain $|F_I^{FD}|$. Therefore, for each test system, we will select a probe atom for comparison rather than testing the forces for every atomic centre.

This approach enables a comparison with the implemented analytical force terms within the implicit solvent model. Tests are performed for a variety of systems, including phenol, the phenoxy anion, and the PF_6^- anion. This set of test systems will test the robustness of the implicit solvent force calculations for a range of charged and uncharged systems. Further validation is performed for the implementation of the electrolyte model within ONETEP, and for whether additional force terms must be implemented to obtain accurate analytical forces.

3.2.3 Electrolyte Model in ONETEP

In electrochemical simulations, the solvent environment contains a finite concentration of the electrolyte. Atomic representations of the electrolyte suffer many of the problems of the explicit solvent approach, including the requirement for large numbers of configurations to achieve statistically significant results [127]. Furthermore, very large

simulation cells are required to represent the diffuse concentration of electrolyte in the Stern layer, further increasing computational expense. Much like the implicit solvent, mean field approaches such as the Poisson-Boltzmann model provide a statistically averaged representation of the electrolyte charge density. As a result, significantly more tractable calculations for interfaces including electrolyte can be performed without extending the size of the simulation cell

Within the electrolyte model, the representation of the charge in the Poisson-Boltzmann model is divided into the quantum system in continuum solvent and the electrolyte ions in an ideal solution [110],

$$n_{PB}(\mathbf{r}) = n_{solv}(\mathbf{r}) + \sum_{i=1}^{p} z_i c_i(\mathbf{r}), \tag{3.12}$$

where z_i is the charge and $c_i(\mathbf{r})$ is the local concentration of the ion of the electrolyte $i = 1, .., p$.

Overall, this transforms the Poisson equation in solvent (Equation 2.67) into the Poisson-Boltzmann equation, which must be solved to include the charge of the Boltzmann ions,

$$\nabla \cdot (\epsilon(\mathbf{r})\nabla\phi_{solv}(\mathbf{r})) = -4\pi n_{solv}(\mathbf{r}) \rightarrow \nabla \cdot (\epsilon(\mathbf{r})\nabla\phi_{PB}(\mathbf{r})) = -4\pi(n_{solv}(\mathbf{r}) + \sum_{i=1}^{p} z_i c_i(\mathbf{r})). \tag{3.13}$$

As the number of particles varies in the simulation cell, the total energy must be described using a grand canonical ensemble ($\Omega[n_{solv}(\mathbf{r}, \{c_i\}, \phi_{PB}(\mathbf{r})]$), which is obtained as a Legendre transformation of the Helmholtz free energy, A [128]. Overall, this is expressed as,

$$\Omega[n_{solv}(\mathbf{r}), \{c_i\}, \phi_{PB}(\mathbf{r})] = E[n_{solv}(\mathbf{r})] - TS[f_k] + \Omega_{mf}[n_{solv}(\mathbf{r}), \{c_i\}, \phi_{PB}(\mathbf{r})] + \Omega_{nmf}[A, V], \tag{3.14}$$

where the first term is the total energy at 0 K, the second is the electronic entropy determined from a Fermi-Dirac smearing scheme (Equation 2.36), the third is the mean field free energy functional which accounts for the energetics of the electrolyte, and the fourth which is the non-mean field energy (which is equivalent to ΔG_{npol}). The energy minimisation of $\Omega[n_{solv}(\mathbf{r}), \{c_i\}, \phi_{PB}(\mathbf{r})]$ is obtained using a three loop scheme, in terms of ionic concentration ($\{c_i(\mathbf{r})\}$), the density kernel ($K^{\alpha\beta}$), and the set of NGWFs ($|\phi_\alpha(r\rangle$)). A more detailed account of these terms is given by *Dziedzic et al.* [127].

However, analytical force terms have not been derived for the grand canonical energy expression. Although energy minimisation performed in terms of concentration should, in theory, eliminate the implicit dependence of ionic concentration with changes in atomic position, tests must still be performed to ensure that no additional analytical force terms are required in order to perform geometry optimisations.

In force tests carried out for the electrolyte model, calculations were performed in open boundary conditions, with an $LiPF_6$ electrolyte at a concentration of 1 M unless otherwise specified. To calculate accurate forces, energy minimisation was performed using 850 eV kinetic energy cut-off (a value justified by increasing the kinetic energy until meV accuracy was achieved), and the Ensemble DFT [83] module of ONETEP. PAW pseudopotentials were used to represent the core states.

3.3 Results

3.3.1 Comparison of Soft Sphere and Parameterizing SASA+SAV

TABLE 3.1: Comparison between the SASA and SASA+SAV models in the soft sphere model ($f = 1.201$) for ΔG_{solv} between the ONETEP's implicit solvent and experimental data of the Minnesota Solvation Database. Performed for 13 molecule test set.

Cavity Method	Parameterization Technique	α	r	Max Error kJ mol^{-1}	RMSE kJ mol^{-1}	MAE kJ mol^{-1}
Soft Sphere	Opt α - SASA+SAV	0.808	0.91	20.47	5.52	4.22
	Opt α - SASA	0.344	0.89	23.03	6.14	4.94
	Disrep α - SASA	0.281	0.90	17.98	7.15	6.01
Charge Cavity	Opt α - SASA+SAV	0.730	0.86	28.16	6.73	4.84
	Opt α - SASA	0.211	0.88	26.94	6.38	4.86
	Disrep α - SASA	0.281	0.86	33.24	8.71	6.87

Figure 3.1 shows the parameterization of the soft sphere cavity model in terms of ΔG_{pol} and ΔG_{tot} through the cavity scaling factor f and the nonpolar scaling factor, α. Over the range of f, we obtain a minimum error of $\Delta G_{pol} = 3.57$ kJ mol^{-1} with respect to NWChem with $f = 1.208$. The work of *Fisicaro et al.* [123] shows that the variation of ΔG_{pol} around the minimising value of f ($f = 1.201$) is nonparabolic, where the variation of ΔG_{pol} is much larger for smaller values of f. This arises due to the interaction of the dielectric with the larger electron density closer to the atomic centre. Therefore, a higher order polynomial (3^{th} or 4^{th} order) is required to produce an accurate fit.

Due to the low computational cost of evaluating α, we parameterize the nonpolar term using the entire 274 molecule test set, and sweep over small intervals of α (0.01). In contrast to the ΔG_{pol} term, ΔG_{npol} is calculated by a set of linear additive terms. As such, the fit appears to be more linear around the minimum. For the soft sphere cavity SASA+SAV non-polar term, the minimum MAE is achieved with $\alpha = 0.808$. Forgoing the procedure for parameterizing ΔG_{pol} for the isodensity model and simply using the default cavity parameters ($\rho_0 = 0.00035$e Å^{-3} and $\beta = 1.3$), we find that the optimum α values are marginally different with respect to the soft sphere model, indicating that the values of ΔG_{pol}, surface area and volume of the soft sphere approach are comparable to the electron density-based cavity model.

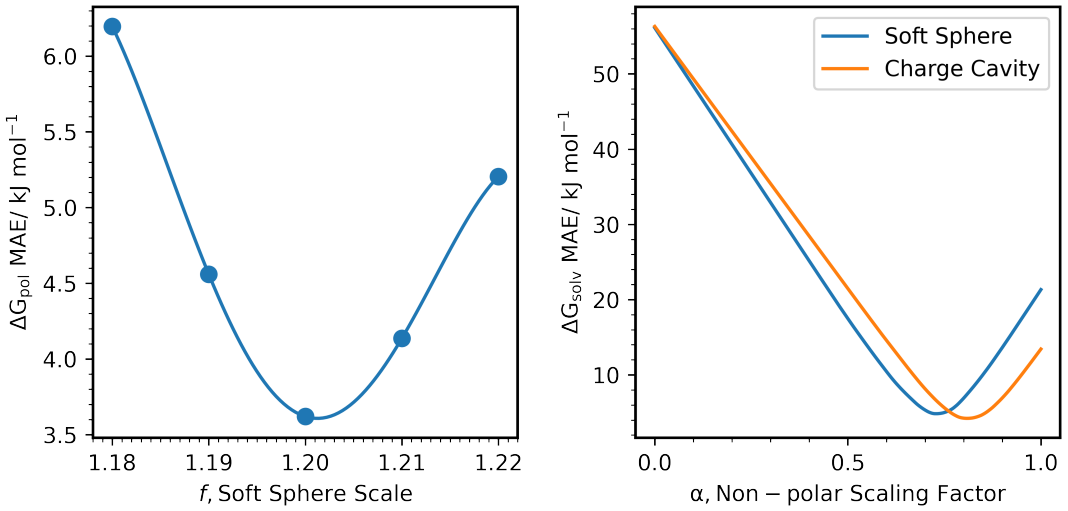

FIGURE 3.1: Parameterization of ΔG_{pol} with a 13 molecule test set compared to COMSO-MSD of NWChem with variations of the soft sphere scaling factor, f fit to a fourth order polynomial (left), and the parameterization of ΔG_{solv} with respect to α compared to the Minnesota Solvation Database (right).

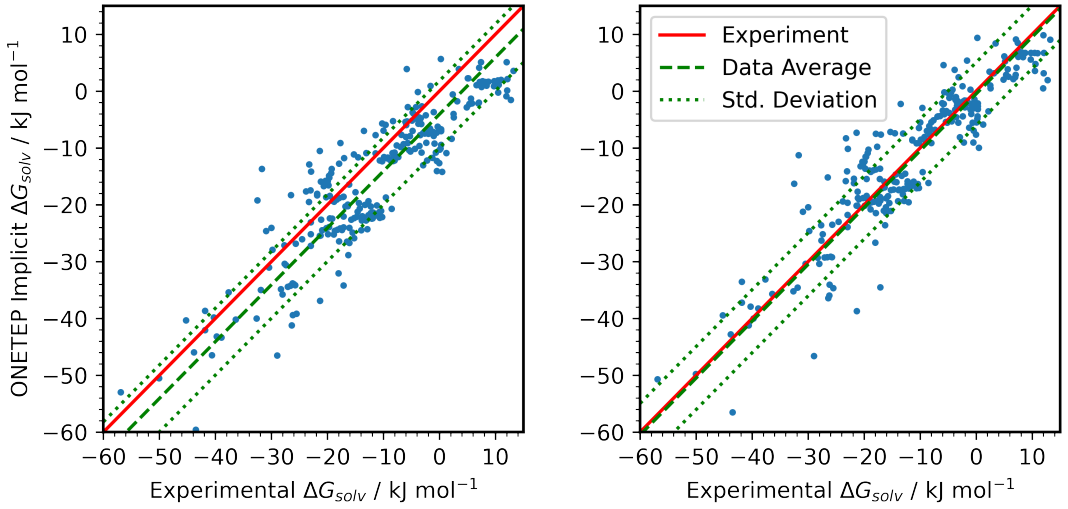

FIGURE 3.2: Comparison of ΔG_{solv} between the original SASA model ($f = 1.201, \alpha = 0.281$) (left) and the SASA+SAV models in the soft sphere model ($f = 1.201, \alpha = 0.808$) (right) for ΔG_{solv} between the ONETEP's implicit solvent and experimental data of the Minnesota Solvation Database. Performed for full 274 molecule test set. Errors with respect to the mean signed error are shown to one standard deviation (SASA: -4.08 ± 5.88 kJ mol^{-1} and SASA+SAV: -0.49 ± 5.54 kJ mol^{-1}).

Comparing the errors across the 274 molecule test set with the original implementation of the nonpolar term (SASA, where $\alpha = 0.281$ corresponds to the dispersion-repulsion corrected van der Waals' non-polar term) to the parameterized SASA+SAV model, we find that the latter model provides moderate improvements in accuracy (Table 3.1). In the soft sphere model, parameterizing α in the SASA+SAV model yields lower errors

with respect to the experimental test set by 1.8 kJ mol^{-1} compared to the original implementation, but only provides a slight improvement in correlation. However, there are a number of outliers between $\Delta G_{solv} = -30$ to -20 kJ mol^{-1} not present in the original SASA approach. We obtain a MAE for the parameterized SASA model (1.1 kJ mol^{-1}), but there is a marginal decrease in correlation.

Evaluating the parameterization of ΔG_{npol} for the charge cavity based approach, we observe a larger improvement in the MAE for the parameterized α SASA and SASA+SAV models. However, the SASA approach shows better performance in terms of correlation, maximum error and the RMSE. From this, we can infer that the slight change in topology between the soft sphere and charge-based cavities leads to the surface area alone being a better descriptor for the latter cavity model.

3.3.2 Implicit Solvent Model Force Tests

Figure 3.3 shows the difference between the analytical forces and the exact finite difference forces. Here, we tested the oxygen atom of the neutral phenol and phenoxy anion species and the P atom of $[PF_6]^-$. In our tests, $||F_I^{FD}| - |F_I^{An}|| < 0.001$ Ha Bohr^{-1} is considered an acceptable degree of error, as it is significantly below the threshold used for geometry optimization calculations of $|F|_{max} < 0.003$ Ha Bohr^{-1}. The error in vacuum *versus* analytical forces is less than 0.0005 Ha Bohr^{-1} for all measured systems, which acts as a further reasonable bound for the force error. Across all systems tested, we measure significantly higher force errors where the electron density-based dielectric cavity is kept fixed from a converged vacuum calculation. In contrast, the force errors

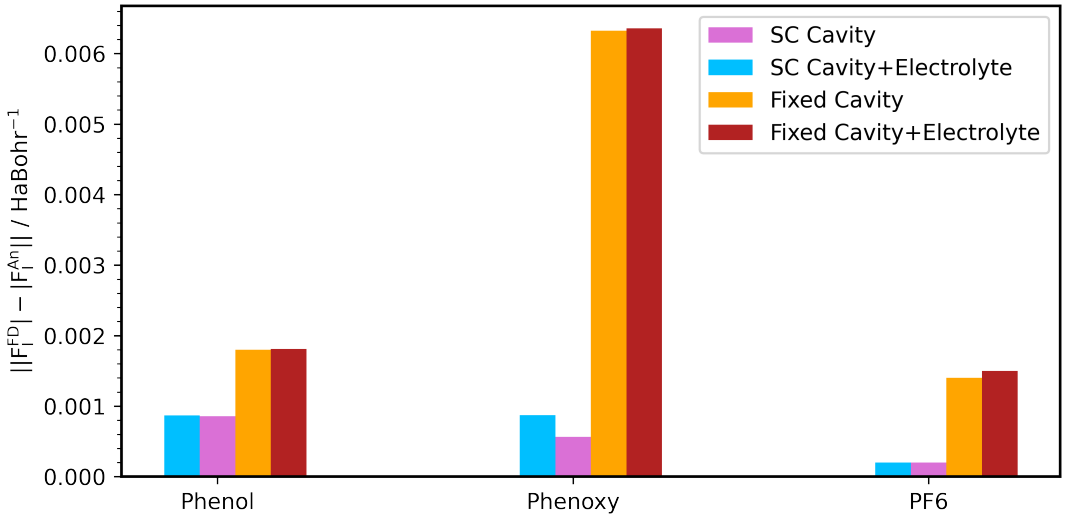

FIGURE 3.3: Errors in the analytical forces compared to the finite difference forces described by Equation 3.10. Forces calculated for the O atom of phenol and phenoxy anion, and P of $[PF_6]^-$

TABLE 3.2: Variation in force error with low and high concentrations of $[PF_6]^-$ for charged and uncharged species.

	$\|\|F_I^{FD}\| - \|F_I^{An}\|\|$ / Ha Bohr^{-1}	
Bulk KPF_6 Concentration	Phenol O Atom	Phenoxy O Atom
Fixed Cavity – 1M	0.0069	0.0018
Fixed Cavity – 6M	0.0023	0.0020
SC Cavity – 1M	0.0003	0.0009
SC Cavity – 6M	0.0007	0.0009

associated with the self-consistent cavity method are significantly smaller. This is explained by the satisfaction of the Hellman-Feynman principle, where the self-consistent cavity method implicitly minimises the energy with respect to the cavity function (i.e. $\partial E/\partial \epsilon(\mathbf{r}) = 0$). In contrast, leaving the dielectric cavity fixed introduces a large degree of error associated with $\partial E/\partial \epsilon(\mathbf{r})$, meaning this term must be computed explicitly to obtain accurate forces. This further implies that, for the soft sphere cavity model, one would expect force errors on the scale of the fixed cavity, as the electronic energy is similarly not minimised with respect to the dielectric cavity function.

We further remark that the force errors for the oxygen atom of the negatively charged phenoxy anion are significantly larger than those for the neutral phenol. As charged species are significantly more polarised by the dielectric, we expect charged species to deviate further from the $\delta E/\delta \epsilon(\mathbf{r}) = 0$, leading to larger errors in the analytical forces which do not take this term into account.

The electrolyte model does not appear to introduce a large degree of error to the analytical forces, with $||F_I^{FD}| - |F_I^{An}||$ being of equal magnitude to the corresponding cavity scheme. This holds for both charged and neutral species. We have further tested this for higher electrolyte concentrations (1 M vs 6 M KPF_6) with the oxygen atom of the phenol and the phenoxy anion (Table 3.2). This ensures that this effect is not an artefact of using a dilute electrolyte with a weak response to the solute. Even at high concentrations, the force error is well below the force tolerance used for geometry optimizations, and is on the scale of the error for forces calculated in vacuum. The largest error, consistent with Figure 3.3, occurs with the use of the fixed cavity for both species. Therefore, the results presented show that the analytical forces can be used unmodified for systems in the electrolyte, provided that the self-consistent cavity update scheme is used.

3.4 Chapter Summary

In this chapter, we have described a range of implementations for the implicit solvent model of ONETEP which will enable us to perform both more flexible parameterization

of the aqueous organic/metallic interface, and allow for the accurate calculation of the ΔG_{npol} term.

Within the wider context of ONETEP's solvent model, the soft sphere model provides an accurate alternative to the isocontour approach. Furthermore, we found that by introducing a dependency on the dielectric cavity volume to the ΔG_{npol} term (SASA+SAV) and parameterizing the original surface tension term γ, we obtain lower errors overall ΔG_{solv} compared to the original SASA model. Therefore, we have demonstrated a successful implementation of the soft sphere cavity model into ONETEP, which compares favourably to the density-based cavity model, enabling its application to further studies.

Furthermore, the finite difference force tests show that the fixed cavity dielectric leads to significant force terms. In order to eliminate this spurious force error, ONETEP would require the implementation of an analytical expression, or achieve energy minimisation with respect to the dielectric cavity through the SC cavity approach. Interestingly, contrary to initial findings [114], self-consistent cavity calculations achieve convergence without finer grid spacing. As a result, such calculations can be performed at computational costs similar to those of the fixed cavity method.

The extra functionalities introduced in this chapter will be instrumental throughout the rest of the work presented. They will be especially useful in Chapter 6, where the electron density dielectric model yields large errors in the calculations of ΔG_{solv}^{ads}.

Chapter 4

Accelerating Ensemble DFT in ONETEP

In Chapter 2, we recounted the underlying theory of DFT as applied to metallic systems (see: Section 2.2.3). In practical calculations, particular care must be taken to ensure calculations are performed in an efficient and timely manner. Throughout this work, we calculate the ground state energies and geometries for large metallic systems, which require large amounts of computational resources. Therefore, the numerical procedures used to minimise the electronic energy should be designed with efficiency in mind. This is achieved by minimising the number of iterations in the self-consistent field scheme, and the computational cost of each iteration is kept small.

The original implementation of Ensemble DFT (EDFT) [83] is based on the two-loop scheme of *Marzari et al.* [136]. This approach is a robust and variational method to minimise the total energy of metallic systems. However, as we demonstrate in this chapter, the linear order of convergence leads to slow convergence within the inner loop of ONETEP. This is especially apparent for larger systems, where even with many iterations, the inner loop often fails to achieve convergence. Furthermore, as a line search routine included in the original implementation, the polynomial fit must be populated with at least two total energy values to find the best step size. The required total energy evaluations and diagonalisations are in themselves computationally expensive procedures. Therefore, the trade-off in computational effort for accuracy should be evaluated to ensure that the EDFT scheme avoids unnecessary resource-intensive subroutines.

Several approaches have been attempted to address these issues. In the following chapter, we will recount the two most successful outcomes of these studies:

1. The application of Pulay Mixing algorithms to replace the previously used damped fixed point algorithm.

2. Replacing the line search algorithm used to determine an optimum step size for each inner loop step with a user determined step size.

This account will outline the underlying numerical methods used in the old and new EDFT implementations in ONETEP, the relevant shortcomings of each approach, and the improvements in computational efficiency achieved with these new implementations.

4.1 Background

4.1.1 Self-Consistent Field Approaches

In the general iterative framework, the density output by the solution to the Kohn-Sham eigenvalue problem is known as the Kohn-Sham map,

$$\boldsymbol{K}[n^{in}(\mathbf{r})] = n^{out}(\mathbf{r}), \tag{4.1}$$

which is used to define the Hartree and exchange-correlation potentials of the next iteration. Section 2.2.2 describes the Kohn-Sham formulation of DFT, where Equations 2.28, 2.33 and 2.34 are the Kohn-Sham equations. The goal of DFT is to find the optimal solution to these equations, which is defined by the potential V^{in} and density $n^{in}(\mathbf{r})$ which minimises the Kohn-Sham functional (equation 2.30). Self-consistency is said to be achieved when $V^{out} = V^{in}$ and $n^{out}(\mathbf{r}) = n^{in}(\mathbf{r})$, where both the potential and the electron density are said to be in their respective ground states. This approach is known as the self-consistent field, which arises from the early work by Hartree [137].

Outside of the minima, the density produced by the eigenvectors for a particular eigenvalue problem generates a different potential to that of the input density (i.e. $n^{out}(\mathbf{r}) - n^{in}(\mathbf{r}) \neq 0$). However, self-consistency is said to be achieved when the solution to the eigenvalue problem generates the same density and potential to their corresponding inputs. When this condition is not satisfied, we define this difference between the input and output densities as the residual,

$$\boldsymbol{R}[n^{in}(\mathbf{r})] = n^{out}(\mathbf{r}) - n^{in}(\mathbf{r}). \tag{4.2}$$

Self-consistency can be achieved through a variety of techniques. The most simple of these is the fixed-point approach shown in Figure 4.1. This is formally defined as,

$$n^{in}_{p+1}(\mathbf{r}) = n^{in}_{p}(\mathbf{r}) + \boldsymbol{R}[n^{in}_{p}(\mathbf{r})], \tag{4.3}$$

where p is an index that defines the iteration step. Here, an initial density generates a potential for which one successively solves the Kohn-Sham eigenvalue problem

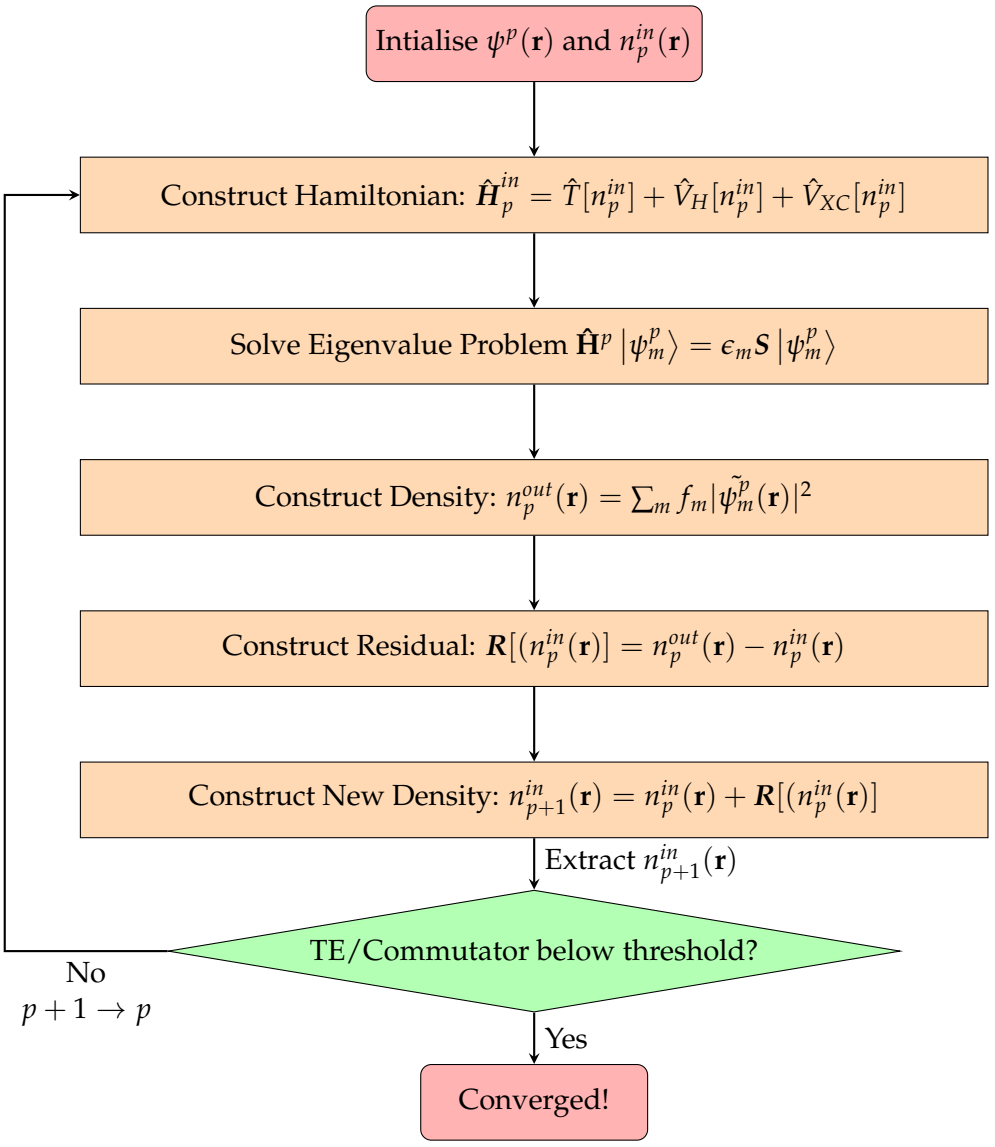

FIGURE 4.1: The fixed-point update scheme for achieving the self-consistent solution to the Kohn-Sham energy functional.

with the input potential V^{in}. The resultant eigenfunctions $\psi(\mathbf{r})$ generate a new density $\mathbf{K}[n^{in}(\mathbf{r})]$, which then defines a new potential, V^{out} for which we must solve again the eigenvalue problem until self-consistency is achieved. However, outside of trivial cases, this approach is insufficient to perform stable minimisations of the Kohn-Sham equations.

This can be justified by considering the output density $n_{out}^p(\mathbf{r})$ sufficiently close to the ground state density, $n_0(\mathbf{r})$ [138, 139]. Defining the corresponding potentials as $V[n_{in}(\mathbf{r})]$ and $V[n_0(\mathbf{r})]$, the deviation between the ground state and the output density is expressed as $\delta n_{out}(\mathbf{r}) = n_{out}(\mathbf{r}) - n_0(\mathbf{r})$. When $\delta n_{out}(\mathbf{r})$ is small, we can linearise this

expression (i.e., taking the Taylor expansion to the first order around $n_0(\mathbf{r})$) such that

$$\delta n^p_{out}(\mathbf{r}) \approx \frac{\delta n^p_{out}(\mathbf{r})}{\delta n^p_{in}(\mathbf{r})} \delta n^p_{in}(\mathbf{r}), \tag{4.4}$$

where the derivative can be further expanded by the chain rule to include the dependence of $V(n^p_{in}(\mathbf{r}))$ on the densities and vice versa,

$$\frac{\delta n^p_{out}(\mathbf{r})}{\delta n^p_{in}(\mathbf{r})} = \frac{\delta n^p_{out}(\mathbf{r})}{\delta V^p_{in}(\mathbf{r})} \frac{\delta V^p_{in}(\mathbf{r})}{\delta n^p_{in}(\mathbf{r})}. \tag{4.5}$$

These equations can be better understood as a response function [82], $\epsilon(\mathbf{r}, \mathbf{r}')$, where for a potential $V^p(\mathbf{r}) = V^p_H(\mathbf{r}) + V^p_{xc}(\mathbf{r})$ [139],

$$\epsilon(\mathbf{r}, \mathbf{r}') = \boldsymbol{I} - \int d\mathbf{r}'' \frac{\delta V^p_{in}(\mathbf{r})}{\delta n^p_{out}(\mathbf{r}'')} \frac{\delta n^p_{in}(\mathbf{r}'')}{\delta V^p_{in}(\mathbf{r}')} = \boldsymbol{I} - \int d\mathbf{r}'' \left(\frac{1}{|\mathbf{r} - \mathbf{r}''|} + \frac{\partial^2 E_{XC}}{\partial n_{in}(r) \partial n_{in}(r'')} \right) \chi_0, \tag{4.6}$$

where the terms of the integral corresponds to the coulomb and exchange correlation kernels respectively. The linear response of the density is described by the dielectric susceptibility matrix, χ_0 defined by the relation,

$$\chi_0 = \frac{\delta n^p_{out}(\mathbf{r})}{\delta V^p_{in}(\mathbf{r})}. \tag{4.7}$$

In order to describe an ideal step within the iterative scheme, an exact expression of $\epsilon(\mathbf{r}, \mathbf{r}')^{-1}$ for the system under study would be required. This would effectively cancel the linear response induced by the change of density at iteration i, therefore allowing convergence to be achieved within a single step,

$$n_0(\mathbf{r}) = n^{p+1}_{in}(\mathbf{r}) = n^p_{in}(\mathbf{r}) + \epsilon(\mathbf{r}, \mathbf{r}')^{-1} \boldsymbol{R}[n^{in}_p(\mathbf{r})]. \tag{4.8}$$

However, $\epsilon(\mathbf{r}, \mathbf{r}')^{-1}$ is both highly system dependent and expensive to calculate [139]. As such, this term is represented by an approximate ansatz of $\epsilon(\mathbf{r}, \mathbf{r}')^{-1}$, which we will describe as $\boldsymbol{J}^{-1}$, or the inverse Jacobian in the language of Newton algorithms. We can therefore describe the ideal construction of J^{-1} as one that is fully equivalent to $\epsilon(\mathbf{r}, \mathbf{r}')^{-1}$,

$$\boldsymbol{I} = \epsilon(\mathbf{r}, \mathbf{r}') \epsilon(\mathbf{r}, \mathbf{r}')^{-1} \approx \epsilon(\mathbf{r}, \mathbf{r}') \boldsymbol{J}^{-1}. \tag{4.9}$$

In iterative schemes, $\boldsymbol{J}^{-1}$ should be constructed such that successive residuals approach 0 in a finite number of steps (that is, $\boldsymbol{R}[n^{p+1}(\mathbf{r})] < \boldsymbol{R}[n^p(\mathbf{r})]$). The fixed point iteration scheme simply represents the ansatz Jacobian as $\boldsymbol{J}^{-1} = \boldsymbol{I}$. However, it can be shown that the eigenvalues γ_i of $\epsilon(\mathbf{r}, \mathbf{r}')$ are always positive and greater than 1 [82, 139]. As a result, the value of $\boldsymbol{J}^{-1} = \boldsymbol{I}$ within the fixed-point framework necessarily leads to divergent iteration. This means minimisation cannot be achieved outside the simplest systems, and rationalises why the self-consistent field problem cannot be solved within

a single step.

However, we can improve this scheme by building a value of $\boldsymbol{J}^{-1}$ that better reflects $\epsilon(\mathbf{r},\mathbf{r}')^{-1}$. This is most simply achieved by expressing $\boldsymbol{J}^{-1}$ as a scalar value applied to $\boldsymbol{R}[n^{in}(\mathbf{r})]$. This scheme is known as the damped fixed-point or linear mixing, which reduces the size of the iterative step through a scalar prefactor, λ.

$$n_{in}^{p+1}(\mathbf{r}) = n_{in}^{p}(\mathbf{r}) + \lambda \boldsymbol{R}[n_{in}^{p}(\mathbf{r})], \tag{4.10}$$

where $\lambda < 1$. A value of λ which guarantees a minimising step can be expressed by considering the maximum eigenvalue of $\epsilon(\mathbf{r},\mathbf{r}')$ [82],

$$\lambda \leq \frac{2}{\max(\gamma_j)}. \tag{4.11}$$

This approach necessarily improves the robustness of the SCF algorithm by ensuring $\max(\gamma_j) < 1$. As this method does not modify or introduce approximations to the form of the linear response interaction, this approach is also globally convergent for an appropriate value of λ [139]. However, the eigenspectra of $\epsilon(\mathbf{r},\mathbf{r}')$ change with the size and species composition of the system. As the difference between $\max(\gamma_j)$ and $\min(\gamma_j)$ increases, the rate of convergence for energy levels associated with $\min(\lambda_j)$ decreases. Because of this, for insulators with a fixed band gap, the number of iterations increases by order N with system size. The rate of convergence for metals worsens due to divergences in the eigenspectra of $\epsilon(\mathbf{r},\mathbf{r}')$. *Annett et al.* and *Woods et al.* give a more complete discussion of these divergences [82, 139]. Therefore, on top of the N^3 computational cost of operations in DFT, we must perform more iterative steps to minimise the energy. This leads to more expensive calculations, which we will demonstrate in the results section.

This is further exacerbated by the linear order of convergence in the linear mixing approach [139, 140]. For a sequence of iteratively improved densities $n^{p+1}(\mathbf{r})$, the rate of convergence in linear mixing is described by the relation:

$$|n_{out}^{p+1}(\mathbf{r}) - n_0(\mathbf{r})| \leq c|n_{out}^{p}(\mathbf{r}) - n_0(\mathbf{r})|, \tag{4.12}$$

where $c < 1$. According to this inequality, $n^{p+1}(\mathbf{r})$ will only be smaller than $n^{p}(\mathbf{r})$ by a constant value. Depending on the complexity of the problem involved and the distance from the stationary point, many iterations would be required to achieve convergence with this approach. Improved rates of convergence can be achieved with superlinear type algorithms, where convergence is instead described as,

$$|n_{out}^{p+1}(\mathbf{r}) - n_0(\mathbf{r})| \leq t_p|n_{out}^{p}(\mathbf{r}) - n_0(\mathbf{r})|, \tag{4.13}$$

where t is such that $t^{p+1} < t^p$. Over an infinite number of iterations, this results in a convergence towards 0,

$$\lim_{p\to\infty} \frac{|n_{out}^{p+1}(\mathbf{r}) - n_0(\mathbf{r})|}{|n_{out}^{p}(\mathbf{r}) - n_0(\mathbf{r})|} = 0. \tag{4.14}$$

Superlinear algorithms, therefore, achieve convergence appreciably faster than linear algorithms, as each iteration produces a smaller error with respect to the ground state density than the previous.

To improve the rate of convergence, one requires iterative techniques which both increase the order of convergence, and reduce the ill-conditioning inherent in the Kohn-Sham problem. This can be achieved by considering superlinear convergent methods, such as the quasi-Newton class of iterative schemes, which include the Broyden scheme. We can further improve convergence by using multisecant methods, which use a weighted history of iterations to generate the density of the next iteration [139]. The most successful and well-known in the computational chemistry community is the Pulay mixing method [141].

4.1.2 Pulay Mixing

The Pulay mixing algorithm (also known as Direct Inversion of the Iterative Subspace) is a quasi-Newton iterative scheme with superlinearly convergent characteristics. Quasi-Newton techniques are characterised by the following expression in the framework of iteratively improving charge densities,

$$n^{p+1}(\mathbf{r}) = n^p(\mathbf{r}) + \boldsymbol{J}_p^{-1}\boldsymbol{R}[n^{in}(\mathbf{r})], \tag{4.15}$$

where $\boldsymbol{J}^{-1}$ is the inverse Hessian defined as

$$\boldsymbol{J} = \frac{\delta n^{p+1}(\mathbf{r})}{\delta n^p(\mathbf{r})} \approx \frac{n_{out}^{p+1} - n_{out}^{p}}{n_{in}^{p+1} - n_{in}^{p}} \tag{4.16}$$

However, the Newton method has two drawbacks: firstly, the Jacobian ($\boldsymbol{J}$) must be computed, updated and stored at each iteration and secondly, the Jacobian must be inverted. Although the Broyden class of methods can circumvent the latter problem by updating the J^{-1} directly, this large matrix must still be stored and is difficult to compute. Later developed Pulay [141, 142] schemes avoid these issues by implicitly calculating the inverse Jacobian using a mixed history of residual vectors. This method attempts to construct an optimal residual, $\boldsymbol{R}[n^p(\mathbf{r})]$, by using a linear combination of previous residuals, accelerating convergence compared to the simple linear mixing approach. In the Pulay method, the optimal residual is defined by,

$$\boldsymbol{R}[n_{in}^{p}(\mathbf{r})] = \sum_{j=p-m+1}^{m} c_j\, \boldsymbol{R}[n_{in}^{j}(\mathbf{r})] \tag{4.17}$$

subject to the Lagrangian constraint,

$$\sum_{j=p-m+1}^{m} c_j = 1 \tag{4.18}$$

The Pulay method obtains the optimal coefficients c_j by minimising the Frobenius norm of the residual history. This is achieved by solving the following system of linear equations,

$$\begin{pmatrix} B_{11} & B_{12} & ... & B_{1m} & 1 \\ B_{21} & B_{22} & ... & B_{2m} & 1 \\ ... & ... & ... & ... & ... \\ B_{m1} & B_{m2} & ... & B_{mm} & 1 \\ 1 & 1 & ... & 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ ... \\ c_m \\ \nu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ ... \\ 0 \\ 1 \end{pmatrix} \tag{4.19}$$

wherein $B_{jk} = \left\langle \boldsymbol{R}[n_{in}^{j}(\mathbf{r})] \middle| \boldsymbol{R}[n_{in}^{p\ k}(\mathbf{r})] \right\rangle$, with the braket notation indicating the inner product of the error vectors and ν the Lagrangian multiplier enforcing the constraint of equation 4.18. The overall update of the $n_{in}(\mathbf{r})$ is then expressed as,

$$n_{in}^{p+1}(\mathbf{r}) = \sum_{j=m-p+1}^{m} [c_j\, n_{in}^{j}(\mathbf{r}) + c_j \boldsymbol{R}[n_{in}^{j}]]. \tag{4.20}$$

4.2 Direct minimisation Approaches

4.2.1 Marzari and Freysoldt Ensemble DFT

Self-consistent field approaches such as density mixing [143] are the mainstay of many plane wave DFT codes. Pulay mixing schemes and appropriate models, which screen long range perturbations such as Kerker preconditioning [144], make these approaches comparatively robust. However, the preconditioning approach presupposes the physics of the system under study. For example, the Kerker preconditioner is parameterized for homogeneous metallic systems, and is less efficient for inhomogeneous systems which include conducting and insulating species [145]. An analysis performed with an array of systems shows that the most robust self-consistent approach tested (Pulay mixing with Kerker preconditioning) achieves convergence in 84% of 54 test cases [139]. However, to avoid the effects of ill-conditioning, one must either develop more sophisticated preconditioning schemes [145] or use direct minimisation approaches [136, 146].

Direct minimisation algorithms are a more robust, universal alternative to self-consistent field methods. One such algorithm is the Ensemble DFT method, implemented as the two-loop minimisation scheme of *Marzari et al.* [136], or the simultaneous update

scheme of *Freysoldt et al.* [146]. In essence, the Marzari scheme directly minimises the Kohn-Sham orbitals through a conjugate gradient scheme for the outer loop, while the inner loop iteratively improves the occupation numbers for the associated orbitals. This approach is advantageous over the conjugate gradient scheme of *Fürthmuller et al.* [81] in that it decouples the optimization of occupancies from the orbitals by reformulating the Helmholtz free energy functional (Equation 2.36) as,

$$A[T;\{\psi_i\}] = \min_{\{f_{ij}\}} A[T;\{\psi_i\},\{f_{ij}\}], \tag{4.21}$$

where f_{ij} are the occupancies of the eigenstates. This scheme performs a minimisation of the occupancy matrix, which includes the terms for the unitary transformation of the orbitals within the iterative improvement of the occupancies. This was achieved by calculating the implicit derivative of the Helmholtz free with respect to the change in occupancy, which reduces to zero under the minimisation conditions ($\frac{\partial A\{\psi_i\}}{\partial f_{ij}} \frac{\delta f_{ij}}{\delta d\psi_i}$, where $\frac{\delta A\{\psi_i\}}{\delta f_{ij}} = 0$). Therefore, this approach removes the implicit dependency of $A[T;\{\psi_i\}]$ on f_{ij} in accordance with the Hellman-Feynman theorem [102]. As a result, only the explicit derivative of $\frac{dA\{\psi_i\}}{d\psi_i}$ must be calculated.

Furthermore, one must consider the unitary transformations which map ψ_i into the occupied subspace, which is required to obtain a diagonalised Hamiltonian [81]. In systems with fractional occupancy, the original Helmholtz free energy operator, which treats the occupancy matrix as diagonal, is no longer invariant to unitary rotations in the occupied subspace [147, 148], expressed as,

$$|\psi_i\rangle \rightarrow \sum_i \boldsymbol{U}_{ji}|\psi_j\rangle. \tag{4.22}$$

Therefore schemes must be applied which variationally decrease the energy according to the dependency based on U_{ij}. The Marzari scheme circumvents this issue by expanding the occupancy matrix to include the unitary transformations,

$$f_{ij} = \sum_k \boldsymbol{U}_{ik} f_k \boldsymbol{U}_{kj}, \tag{4.23}$$

while simultaneously applying the same transformation to the orbitals according to Equation 4.12. The ill-conditioned variations due to subspace rotations are treated implicitly by the minimisation of the free energy functional with respect to the occupancies. As a result, this method drastically improves the convergence for systems with small band gaps.

The Freysoldt scheme uses the same principle as the Marzari approach, but instead of minimising the Helmholtz free energy with respect to the f_{ij}, a pseudo-Hamiltonian ($\boldsymbol{H_{i,j}}$) constructed from f_{ij} and U_{ik} is used instead, producing a new projected Helmholtz

free energy functional of,

$$A[T] = \min_{H_{ij},\psi_i} A[T;\{\psi_i\},\{\boldsymbol{H}_{ij}\}], \tag{4.24}$$

where explicit gradient terms are derived for A with respect to H_{ij} and ψ_i, and both quantities are updated simultaneously. As expressed by *Aarons et al.* [149], minimising the energy directly via. the orbital energy improves the ill-conditioning of the Marzari scheme, where the occupancies $f_i \to 0$ map to a large range of orbital energies.

4.2.2 ONETEP Implementation of EDFT

Ruiz-Serrano et al. [83] implemented EDFT into ONETEP, utilizing elements of both the Marzari and the Freysoldt approaches - taking the two loop minimisation scheme of Marzari, but performing the minimisation of A with respect to H_{ij} instead of f_{ij}. The two-loop minimisation matches well with the scheme already implemented within ONETEP [91], while using the Hamiltonian as the iterate avoids the ill-conditioning of f_{ij} mentioned in the previous section. However, the expressions from the plane wave DFT formulation of EDFT must be redefined in terms of the localised NGWF framework of ONETEP.

The Helmholtz free energy functional in terms of the NGWFs and the $\boldsymbol{M}_i^{\alpha}$ expansion coefficients is expressed as,

$$A[T;\{|\phi_\alpha\rangle\}] = \min_{\boldsymbol{H}_{\alpha\beta}} A[T;\{\boldsymbol{H}_{\alpha\beta}\},\{|\phi_\alpha\rangle\}], \tag{4.25}$$

where the Hamiltonian $\boldsymbol{H}_{\alpha\beta}$ is defined in terms of the NGWFs $|\phi_\alpha\rangle$,

$$\boldsymbol{H}_{\alpha\beta} = \langle\phi_\alpha|\,\hat{H}\,|\phi_\beta\rangle = \boldsymbol{S}_{\alpha\gamma}\sum_{i}^{N_b}(\boldsymbol{M}_i^{\gamma}\epsilon_i\boldsymbol{M}_i^{\dagger\delta})\boldsymbol{S}_{\delta\beta}. \tag{4.26}$$

Within this minimisation scheme, the inner loop iteratively improves $H_{\alpha\beta}$ with a fixed set of NGWFs until convergence is achieved, then the NGWFs are improved using a conjugate gradient scheme. As the improvements implemented in this work only involve the inner loop, we will exclude the outer loop from this discussion for the sake of brevity. For a more detailed description of the whole scheme, see reference [83].

The inner loop of EDFT is structured as follows. An initial density, $n(\mathbf{r},\mathbf{r}')$ (Equation 2.53) is used to construct the potential and exchange-correlation components of $H_{\alpha\beta}^{p}$. The Hamiltonian is then diagonalised according to the eigenvalue problem,

$$\boldsymbol{H}_{\alpha\beta}^{p}\boldsymbol{M}^{\beta}{}_{i}^{p} = \boldsymbol{S}_{\alpha\beta}\boldsymbol{M}^{\beta}{}_{i}^{p}\boldsymbol{\epsilon}_i^{p}. \tag{4.27}$$

The resulting eigenvalues ϵ_i^p are then tested for degeneracy and reorthonormalised according to the Löwdin orthonormalisation scheme [150]. The returned values of $\boldsymbol{M}^{\alpha}{}_i{}^p$ and ϵ_i^p allow for the construction of the smeared occupancies determined by the Fermi-Dirac distribution,

$$f_i^p(\epsilon_i^p) = \left(1 + \exp\left[\frac{\epsilon_i^p - \mu^p]}{k_B T}\right]\right)^{-1}. \tag{4.28}$$

In turn, the entropy term for the minimisation of the free energy functional may be calculated as,

$$S^p[\{f_i^p\}] = -k_B \sum_i [f_i^p \ln f_i^p + (1 - f_i^p) \ln 1 - f_i^p]. \tag{4.29}$$

The density kernel with fractional occupancies is given by,

$$\widetilde{\boldsymbol{K}}_p^{\alpha\beta} = \sum_p \boldsymbol{M}^{\alpha}{}_i{}^p f_i^p \boldsymbol{M}_i{}^{\dagger\beta p}. \tag{4.30}$$

The diagonalised Hamiltonian at step p ($\tilde{H}_{\alpha\beta}^p$) is then used to generate the next Hamiltonian at step $p+1$ by building the residual,

$$\boldsymbol{R}[\tilde{\boldsymbol{H}}_{\alpha\beta}^p] = \tilde{\boldsymbol{H}}_{\alpha\beta}^p - \boldsymbol{H}_{\alpha\beta}^p, \tag{4.31}$$

$$\tilde{\boldsymbol{H}}_{\alpha\beta}^p = \boldsymbol{H}_{\alpha\beta}^p + \lambda \boldsymbol{R}[\tilde{\boldsymbol{H}}_{\alpha\beta}^p], \tag{4.32}$$

where the value of λ is an optimal trial step constructed through a line search routine. This is calculated by sweeping through a range of λ values between $[0, 1]$ and evaluating the Helmholtz free energy functional obtained from each value of λ. Including the energy at $\lambda = 0$, the step size that results in the optimal free energy is obtained via. a polynomial fit. Typically, only two trial steps and a quadratic fit are required to find the best value of λ. The convergence is then measured according to the tolerance thresholds of the change in free energy ($< 10^{-6}$ Ha per atom) and the commutator $[\boldsymbol{H}_{\alpha\beta}^{p+1}, \widetilde{\boldsymbol{K}}^{\alpha\beta(p)}]$ (10^{-5} Ha). Otherwise, the scheme is repeated from Equation 4.27 until convergence is achieved.

This scheme can be recognised as the simple damped fixed-point iteration approach discussed in Section 4.1.1, except that the Hamiltonian is used as the iterate instead of the charge density $n(\mathbf{r})$. Although this approach is robust, many of the flaws described in the linear mixing regime appear in this algorithm. Particularly for metallic systems where the minimisation scheme is ill-conditioned, large numbers of iterations are required to achieve convergence. As we will show in Section 4.4 with large metallic systems, the number of iterations required to achieve convergence increase dramatically with system size. This is undesirable, as individual steps not only become more costly due to the intrinsic expense of computing the densities and energies for a larger system, but must also perform more steps to obtain convergence.

Furthermore, although the line search routine guarantees safe steps along the search

direction (if they are available) by calculating the optimal λ, this approach introduces large computational costs. Each point generated for the trial λ involves costly calculations of the free energy, where at least two total energy evaluations are required to produce the quadratic fit, and a third to perform the optimal step according to Equation 4.32. Evaluations of $A[T;\{H_{\alpha\beta}\},\{|\phi_\alpha\rangle\}]$ involve many costly computations, such as the calculation of the density and evaluations of the Hartree and exchange-correlation potentials. For the system sizes used in this work, where the N^3 scaling dense matrix diagonalisation is relatively cheap, these density and potential calculations form the majority of computational effort. Therefore, the computational cost of each inner loop iteration is dramatically increased in exchange for a marginal improvement in accuracy - in regimes where λ fluctuates by $< 10\%$ between iterations, this approach does not result in increased accuracy.

By addressing the two issues described above, the remainder of this chapter outlines the improvements made to the EDFT implementation within ONETEP. First, instead of a damped linear mixing approach, a Pulay mixing algorithm was implemented, replacing the density with $\boldsymbol{H}_{\alpha\beta}$ as iteration (see Sections 4.2.1 and 4.1.2). Secondly, we have tested replacing the line search algorithm with a simple fixed step size (λ) approach. In this case, the user specifies an approximate λ value that sufficiently decreases $A[T;\{\boldsymbol{H}_{\alpha\beta}\},\{|\phi_\alpha\rangle\}]$ at each step. Although conceptually simple, we demonstrate that this approach significantly reduces computational costs by reducing the number of energy evaluations at each step, while minimally impacting overall accuracy. However, this also relaxes the variational nature of the original implementation, meaning that uphill energy steps may appear if a poor step size is selected. Therefore, we have also implemented a safety measure where the line-search routine is performed if successive energy increases are calculated in each inner loop iteration. From observations, the step size value of the first 1-3 iterations is significantly larger than that of later steps. As such, the line search routine is maintained for the first 2 steps before proceeding with the user input value of λ. Combined with the Pulay mixing approach, which displays superlinear convergent behaviour, convergence of the inner loop can be performed up to three times faster than the original implementation for large systems. As a result, this scheme is capable of achieving significantly tighter convergence thresholds with a reasonable number of iterations.

4.2.3 Hamiltonian Pulay Mixing in ONETEP

Originally, Pulay mixing in ONETEP was implemented using $K^{\alpha\beta}$ as the iterate. This approach was implemented in ONETEP [104] in the form of kernel mixing for insulators. The authors made attempts to extend this framework to metals as part of this work by including smeared occupancies. However, this approach was numerically unstable, even with the implementation of a real-space Kerker preconditioner [144,

151]. *Ruiz-Serrano* speculated this instability results from breaking the *aufbau* principle, which results from constructing the density kernel from a history of previous iterations [149, 152]. Furthermore, since the density kernel is analogous to the unitary transformation of the fractional occupancy matrix described in the original Marzari scheme (Equations 4.30 & 4.23), it inherits the same ill-conditioning problems associated with mapping small occupancies onto a large span of eigenvalues. However, Hamiltonian mixing not only circumvents the ill-conditioning of the occupancy-based density kernel through the logic of the Freysoldt scheme [146], it also ensures that the occupancies satisfy the ordering of eigenvalues in accordance to the *aufbau* principle. Therefore, we will limit our discussion to the successful implementation of the Hamiltonian mixing scheme. Much of the framework developed in the kernel mixing method [152] can be transferred to the EDFT module, which means that implementation can be achieved relatively easily.

The scheme is in principle very similar to that outlined in Section 4.2.2, except the residual $\boldsymbol{R}[\boldsymbol{H}^p_{\alpha\beta}]$ is constructed according to the Pulay mixing scheme of Section 4.1.2. The scheme is summarised in Figure 4.2. In principle, this technique should produce step directions that converge superlinearly towards the ground state.

The residual of Equation 4.31 from the previous number of steps, defined via. $p-m+1$, where m defines the number of previous Hamiltonians used to form a system of linear equations expressed as,

$$\boldsymbol{R}_{opt}[\boldsymbol{H}^p_{\alpha\beta}] = \sum_{j=p-m+1}^{p} c_j \, \boldsymbol{R}[\boldsymbol{H}^p_{\alpha\beta}], \tag{4.33}$$

subject to the constraint,

$$\sum_{j=p-m+1}^{p} c_j = 1. \tag{4.34}$$

This linear system of equations is solved according to the matrix shown in Equation 4.19. However, the inner product of the residuals $B_{jk} = \left\langle \boldsymbol{R}[\boldsymbol{H}^j_{\alpha\beta}] \middle| \boldsymbol{R}[\boldsymbol{H}^k_{\alpha\beta}] \right\rangle$ must be calculated such that the tensorial nature of the Hamiltonians are correctly accounted for,

$$B_{jk} = \left\langle \boldsymbol{R}[\boldsymbol{H}^j_{\alpha\beta}] \middle| \boldsymbol{R}[\boldsymbol{H}^k_{\alpha\beta}] \right\rangle = tr[\boldsymbol{R}^j \boldsymbol{S} \boldsymbol{R}^k \boldsymbol{S}] = \boldsymbol{R}^j_{\alpha\beta} \boldsymbol{S}^{\beta\gamma} \boldsymbol{R}^k_{\gamma\zeta} \boldsymbol{S}^{\zeta\alpha}, \tag{4.35}$$

where the inner products are taken as the product between the covariant Hamiltonians of the metric tensor $\boldsymbol{S}^{\alpha\beta}$, where $\boldsymbol{S}^{\alpha\beta} = \boldsymbol{S}^{-1}_{\alpha\beta}$ [153]. Overall, this yields a rank-zero tensor.

From the resultant optimal residual vector, an optimal update step is performed according to,

$$\boldsymbol{H}^{p+1}_{\alpha\beta} = \sum_{j=p-m+1}^{p} [c_j \, \boldsymbol{H}^j_{\alpha\beta} + \lambda c_j \boldsymbol{R}[\boldsymbol{H}^j_{\alpha\beta}]]. \tag{4.36}$$

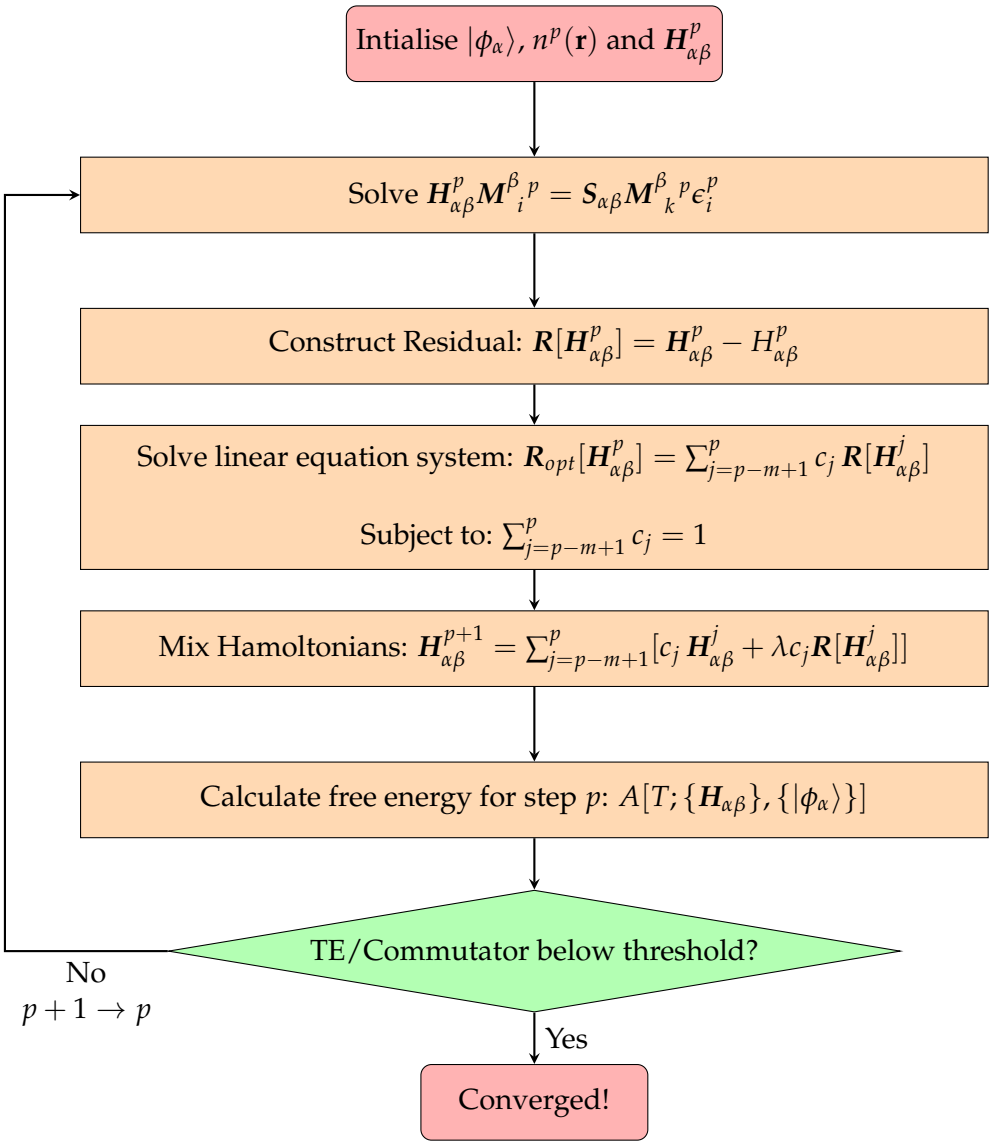

FIGURE 4.2: Proposed Pulay mixing scheme for the Hamiltonians for the inner loop of the ONETEP Ensemble DFT scheme.

We note that the damping λ step size between [0,1] is still required to obtain a minimising step along the search direction. Although the Pulay mixing method indirectly provides an approximated inverse Jacobian $\boldsymbol{J}^{-1}$, which allows for superlinear convergence, it is still subject to the ill-conditioning introduced by the response of long range changes in density with respect to potential (Equation 4.6). Similar to the original implementation with the damped fixed-point approach, one can perform a line search along search direction ($\sum_{j=p-m+1}^{p} \lambda c_j \boldsymbol{R}[\boldsymbol{H}^{(i)}_{\alpha\beta}]$) for λ [0,1], but the fixed λ approach can also be utilised to reduce the number of energy evaluations performed at each step.

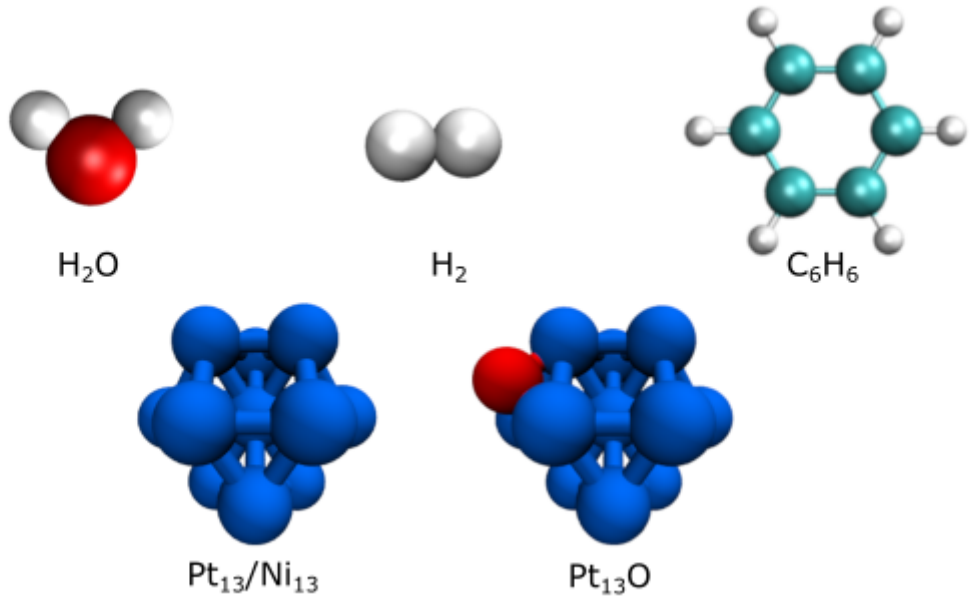

FIGURE 4.3: A selection of small molecules which form a small test set used to validate the Pulay mixing and fixed λ appraoches.

4.3 Calculation Set-up

To provide an assessment of the speed-ups associated with both the fixed λ approach and the Pulay mixed Hamiltonians, we performed a set of comparisons for a range of systems (both insulators and conductors) shown in Figure 4.3. The discussion is then extended to a set of large metallic facets (Pt(111)) representative of the systems used throughout the rest of this work (Figure 4.6). We will discuss these structures in the body of the text. The convergence of structural and electronic properties for the Pt slab (with respect to **k**-point sampling and number of Pt layers) are shown in Appendix C.

Calculations are performed in ONETEP with a 600 eV kinetic energy cutoff and a 9.0 a_0 NGWF cutoff radius for small systems and 850 eV and 12.0 a_0 for large Pt slabs. These values were selected by increasing both the NGWF cut-off radius and kinetic energy cut-off until meV accuracy was achieved. All calculations are performed with the PBE functional [133]. The Pt structures are generated with an experimental FCC lattice constant of 3.92 Å [154]. The initial pseudoatomic orbitals defining the initial NGWFs were constructed using the electronic structure corresponding to the *aufbau* principle of each atomic species. PAW pseudopotentials are used for all systems.

4.4 Results

4.4.1 Convergence Behaviour of Inner Loop for Pt_{13}

First, it will be useful to analyse the inner convergence behaviour of a simple example system within the first outer loop. In this case, we have used the original line search technique, which calculates optimal λ values through a polynomial fit. We also investigate how varying the history length (m of Equation 4.33) affects convergence. Figure 4.4 shows a range of convergence criteria for a Pt_{13}.

Considering the change in $A[T; \{\boldsymbol{H}_{\alpha\beta}\}, \{|\phi_\alpha\rangle\}]$ per step, we see that the Pulay mixing scheme satisfies the convergence criteria within 10-15 iterations. In contrast, the linear mixing routine satisfies this criterion at iteration 25. This is also reflected in the commutator, where Pulay mixing achieves convergence within the same window as the free energy, but damped linear mixing fails to converge until iteration 45. Therefore, we can infer that the Pulay mixing scheme converges more rapidly to a given threshold. This allows the use of tighter convergence criteria, which can improve the overall precision of calculated energies. In contrast to the damped fixed point approach, Pulay mixing occasionally performs uphill energy steps, even when the optimal λ value is found. Although convergence is eventually attained after this step, this shows that the latter minimisation method is not strictly variational, as is typical of residual minimisation methods.

To better understand this instability, we can inspect the convergence behaviour of different Hamiltonian history lengths (m). Overall, we observe relatively little variation between $m = 3$ and $m = 10$. However, the convergence with respect to the commutator and the free energy are marginally more unstable for $m = 10$, as characterised by the uphill energy step at iteration 8. This also occurs to a lesser degree for $m = 5$, but is entirely eliminated for $m = 3$. Although convergence is achieved in spite of these uphill steps, in the case of $m = 10$, we note that more iterations are required to correct the uphill step and satisfy the convergence criteria. In the wider discussion of Pulay mixing methods, the solution to the system of linear equations can become ill-conditioned [155], leading to stagnation in the iterative subspace of previous residuals [156]. The use of shorter histories has been demonstrated to marginally improve the robustness of the Pulay mixing routine [139]. Furthermore, more sophisticated approaches, such as periodic Pulay mixing, have demonstrated that efficiency can be further improved by flushing the Hamiltonian/density history within a set number of iterations [156].

Furthermore, the optimal λ for the Pulay mixing results are significantly more variable than linear mixing. Generally, the steps performed within the former scheme are larger, while the latter continually switches between step sizes of $\lambda = 0.029$ and $\lambda = 0.073$. Comparatively, the larger λ values of Pulay mixing further enable accelerated convergence, which results from the better conditioning caused by the approximated J^{-1}.

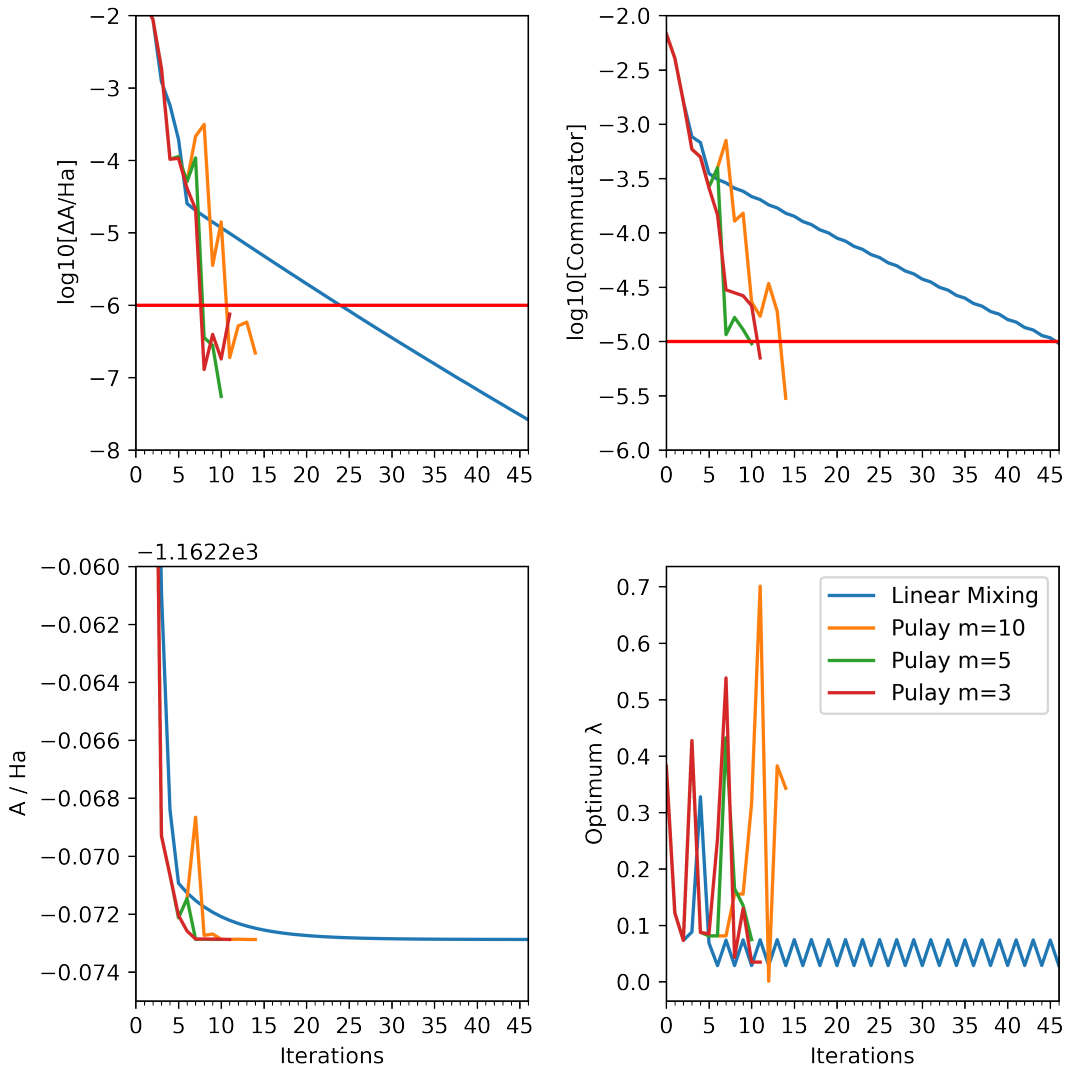

FIGURE 4.4: The relevant convergence criteria for the inner loop for the first NGWF iteration for the Pt_{13} nanoparticle. From top left to bottom right: The logarithm for the change in the free energy per atom, the logarithm of the commutator $[H_{\alpha\beta}^{p+1}, \widetilde{K}^{\alpha\beta(i)}]$, the total free energy and the optimal λ value obtained through the line fitting routine, The red horizontal lines represent the thresholds corresponding to convergence.

Beyond iteration 10 of the linear mixing plot, the order of convergence is linear with respect to the total energy, where the order of convergence is calculated as,

$$\frac{|A^{p+1} - A_\infty|}{|A^n - A_\infty|} = t_p, \tag{4.37}$$

where A_∞ is approximated for a very tight convergence threshold for a Pulay mixing scheme of $m = 5$. The values for this equation are shown in Figure 4.37. For iterations beyond 10, $t_p \approx 0.85$, where a fixed value of $0 < t_p < 1$ characterises linear convergence. A similar analysis for Pulay mixing is nontrivial - the total energy changes inconsistently from step to step, driven in part by the large variation in the optimal λ

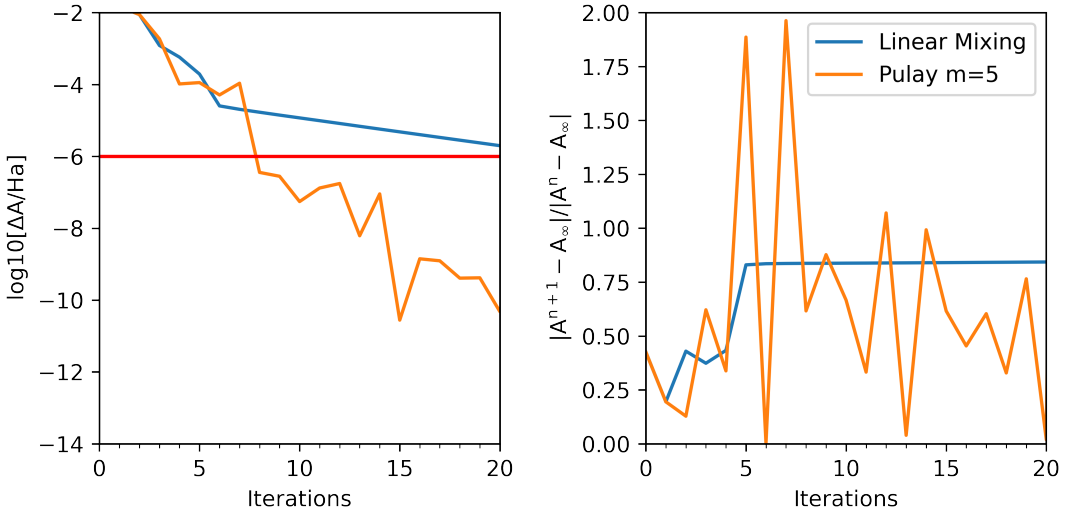

FIGURE 4.5: Shows the convergence bevaviour for 20 iterations with the Pulay mixing and linear mixing routines. The left plot shows the log of the free energy change per step, and the right plot shows the relative size of each free energy step relative to the last.

value. Furthermore, evaluating the algorithm in the limit of $\eta \to \infty$ (where $n = 80$), the machine errors intrinsic to computational methods dominates for energy differences of $\sim 1 \times 10^{-11}$ Ha. This leads to oscillations in the calculation of total free energy at $t \to \infty$. Overall, this means the order of convergence cannot be easily calculated in our analysis. However, from Figure 4.37 we observe that the error in the free energy varies dramatically from step to step, occasionally taking large steps towards the converged final energy. This occurs more often after an uphill step, where $\frac{|A^{n+1}-A_\infty|}{|A^n-A_\infty|} > 1$. On average, larger steps are observed than the linear mixing algorithm, as evidenced by both the overall lower $\frac{|A^{n+1}-A_\infty|}{|A^n-A_\infty|}$ values and the occasional large steps, where $\frac{|A^{n+1}-A_\infty|}{|A^n-A_\infty|} < 0.005$.

4.4.2 Convergence of Total Energy for Insulators and Conductors

As shown in section 4.4.1, the Pulay mixing scheme applied to Hamiltonians confers numerical advantages over the linear mixing scheme within the inner loop. Next, we considered the convergence behaviour for a set of single-point energy calculations, which includes optimisation of the NGWFs in the outer loop. We also introduced the fixed and optimised (line search derived) λ approach, where for the former, a value of $\lambda = 0.075$ was used. The small test set includes a set of small conducting and insulating species to ensure that the Pulay mixing scheme is capable of treating systems with large as well as small band gaps. Further tests were performed with a $Pt_{13}O$ nanoparticle as a representation of inhomogeneous systems, with metallic and non-metallic components. Calculations were performed on the IRIDIS 5 High Performance Computer (HPC), using 1 node composed of 40 2.0 GHz Intel Skylake processors. The number of

TABLE 4.1: Differences in calculation efficiency between the original implementation of EDFT and the Pulay mixing routine for a mixture of small systems with large band gaps and metallic systems with small band gaps. Calculated for the fixed λ and line search (optimised λ) approach for Pulay mixing. The maximum number of inner loop iterations set to 50.

Pulay Mixing Method	System	# Inner Loop Difference	Total Energy Difference / Ha	Calculation Time Difference
Fixed λ	Pt_{13}	-174%	1.50E-06	-172%
	Ni_{13}	-86%	-0.00014	-214%
	H_2O	18%	-2.06E-08	-5%
	$Pt_{13}O$	-153%	-3.04E-06	-181%
	H_2	0%	-	-3%
	C_6H_6	12%	-1.08E-07	4%
Optimised λ	Pt_{13}	-191%	1.50E-08	-91%
	Ni_{13}[2]	-95%	-0.00272	-72%
	H_2O	0.00%	-2.06E-10	5%
	$Pt_{13}O$	-138%	-3.04E-08	-76%
	H_2	0%	-	7%
	C_6H_6	-4%	-1.08E-09	6%

OMP (Open Multi-Processing) threads was set to 10, with 4 MPI processes[1]. As shown in Table 4.1, the optimised and fixed λ with Pulay mixing achieves convergence with significantly fewer overall inner loop iterations than linear mixing with optimised λ (i.e., the original implementation). This translates into significantly shorter calculation times, especially when the λ value is fixed. In spite of using an approximated λ, which moderately increases the number of inner loop iterations compared to using the optimised λ at every step, significantly reduced calculation times are obtained. For metals, convergence is achieved up to three times faster than the original implementation. Furthermore, the final energies are almost identical for most systems, indicating that the Pulay mixing scheme converges to the same local minima of the original scheme. The overall low variation in final energies shows that the convergence is robust for an appropriately specified λ. This means that the fixed-λ method can be used to significantly improve computational efficiency, without introducing large inaccuracies. However, Ni_{13} shows a small variation in energy between the optimised and fixed λ approaches. This results from the greater number of NGWF steps taken in the former method, leading to convergence to another local minimum. It is possible that a tighter convergence threshold for the outer loop would lead to greater alignment in the energies between the tested methods.

Improvements in computational efficiency are less apparent for insulators, where we observe either more modest decreases in calculation time or even slight increases. As discussed in Section 4.1.1, the iterative steps for systems with large band gaps are better conditioned compared to the metallic species with small band gaps. Therefore, the

[1]Except for H_2O and H_2, where the number of processes was set to 1, owing to the restriction in ONETEP where the number of MPI processes cannot exceed the number of atoms within the system

overall number of inner loops are significantly smaller in the original scheme, leaving less room for improvement in the Pulay scheme. Furthermore, the value of $\lambda = 0.075$, which is appropriate for smaller metallic systems, may limit the step sizes for small insulating molecules. We observe optimal step sizes of $\lambda > 0.5$ with line search throughout the calculation, meaning in practice, that benzene could achieve convergence faster with larger step sizes. However, the overall total differences are again small for this subset of systems, meaning the Pulay Mixing routine performs adequately for small molecules.

4.4.3 Convergence of Total Energy Calculations for Large Pt(111) Slabs

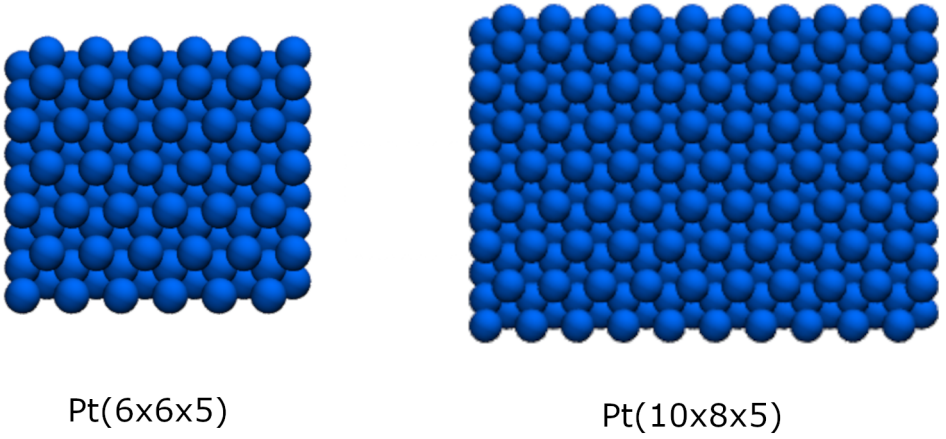

FIGURE 4.6: The Pt (6x6) and 10x8) super cells of the Pt(111) facet based on repetitions of the (2x1) unit cell.

The Pulay mixing algorithm was then tested for larger metallic Pt slabs, which are representative of the systems used throughout the rest of this work. For this, calculations were performed on two orthogonal unit cells of Pt(111) facets constructed as (6×6) and (10×50) supercells (which are periodic repetitions of a primitive (2x1) unit cell) with 5 layers (Figure 4.6). These slabs are separated by 10 Å of vacuum in the z-direction. Calculations were performed on the IRIDIS 5 High Performance Computer (HPC), using 16 nodes composed of 40 2.0 GHz Intel Skylake processors. Each calculation uses 4 OMP threads, which translates to 160 MPI processes. We further show the performance of the fixed λ approximation compared to the line search optimised λ values. The fixed λ values were selected from the average of the inner loop step values calculated through the line search routine after the first 5 steps were performed.

In Table 4.2, overall, the Pulay mixing method with a fixed value of λ leads to considerable savings in calculation time for large metallic systems. Compared to the initial implementation of EDFT (Linear-opt), a factor of ~ 2 speed-up was achieved for both the (6×6×5) and (10×8×5) slab calculations. However, we observe that Pulay mixing

TABLE 4.2: Energy minimisation performed for a large metallic slabs, showing the convergence behaviour for a range of properties. The total energy is expressed as the average of the Helmholtz free energy and the total energy without the entropy components to recover the approximate Total Energy at 0 K.

System	SCF Method	Time / h	# Inner Loop Iterations	# Outer Loop Iterations	# FD Smearing TE Evaluations	Total Energy / Ha
Pt(6x6x5)	Linear - Opt λ	2.90	208	8	630	−16106.66702
	Linear - Fixed λ	1.79	231	7	253	−16106.66691
	Pulay - Opt λ	2.65	139	7	562	−16106.66648
	Pulay - Fixed λ	1.30	132	7	156	−16106.66652
Pt(10x8x5)	Linear - Opt λ	8.78	208	7	629	−35792.3217
	Linear - Fixed λ	6.29	260	9	251	−35792.22859
	Linear - Opt λ [a]	13.90	347	7	1046	−35792.32668
	Linear - Fixed λ [a]	7.56	368	7	361	−35792.29476
	Pulay - Opt λ	8.26	134	7	594	−35792.32599
	Pulay - Fixed λ	4.68	162	8	199	−35792.32627

[a] 45 maximum inner loop steps. All other calculations performed with 25 maximum inner loop steps.

calculations with the line search routine, despite achieving convergence with fewer inner loop steps, is markedly slower than the linear mixing routine performed with fixed λ. This can be justified by considering the number of Fermi-Dirac smearing operations performed, which reflects the number of total energy evaluations and diagonalisations performed within the inner loop. For both systems, the number of energy evaluations within the optimised λ Pulay mixing are comparable to linear mixing with optimised λ, and significantly higher than all calculations performed with fixed λ. This occurs because, as observed for the inner loop of the simple Pt_{13} system (Figure 4.4), λ varies significantly between each iteration. As a result, the line search routine must scan over a larger number of λ values when constructing the A *vs.* λ curve, often performing 3 and up to 6 energy evaluations for the polynomial fitting. In contrast, λ in the linear mixing scheme shows far less variation past iteration 5. As a result, only two points A vs. λ (energy evaluations) are required to find the optimal value of λ. On the contrary, the fixed λ routines completely circumvent the difficulties in performing the line search for Pulay mixing. Although we observe an overall increase in the number of inner loop steps for Pulay fixed λ vs. optimised λ Pulay, the reduced number of total energy/diagonalisation evaluations by avoiding the line search routine leads to considerable savings in computational time. Furthermore, the Pulay mixing routine calculates total energies within 0.0005 Ha of the linear mixing algorithms. These values are within the error limits for the 2×10^{-6} NGWF RMS gradient convergence criteria of ONETEP. An exception to this is total energy for fixed-λ linear mixing for Pt($10 \times 8 \times 5$) with 25 maximum inner loop iterations, where the total energy is ~ 0.1 Ha higher than the corresponding calculations performed with the Pulay mixing routines. The relatively higher energies for the linear mixing routines are justified by considering the convergence of the inner loop throughout the outer loop steps. For each NGWF iteration, the

inner loop fails to achieve full convergence, especially for the first NGWF iterations where the commutator is two orders of magnitude above the threshold of 5×10^{-6} Ha. Overall, this reduces the quality of the NGWF gradient required for updating the NGWFs in the outer loop. Although for later outer loop iterations the linear mixing routine almost reaches the commutator convergence threshold, these calculations still perform the maximum number of inner loop steps at each NGWF step without reaching convergence.

In contrast, the Pulay mixing routine reliably achieves inner loop convergence for each outer loop iteration. This ensures the calculation of accurate NGWF gradients throughout the energy minimisation. The convergence of linear mixing can be marginally improved by increasing the number of maximum inner loop iterations, but we note that the total energy for fixed λ linear mixing is still 7.9 kJ mol^{-1} higher than the other minimisation approaches shown. Although the increased number of maximum inner loop iterations leads to overall lower energies, the time taken to perform the calculation increases dramatically. This results from the relatively large steps taken in the first outer loop NGWF iterations for the fixed Hamiltonians. Between successive outer loops, the optimal Hamiltonians for the NGWFs at each iteration are significantly different from each other. Although the Hamiltonian of the previous outer loop step is transferred to the next, much of the computational effort required to converge the inner loop for one set of NGWFs must be performed again for the next iteration. This is exacerbated by the slow, linear type convergence of the linear mixing routine. However, this is less problematic as we approach convergence for the outer loop, as the optimised Hamiltonians from NGWFs in each outer loop iteration are largely similar to one another, therefore requiring fewer inner loop iterations to reach convergence.

4.5 Chapter Summary

In general, the Pulay mixing routine with a fixed λ value provides significant advantages over the initial implementation of EDFT in ONETEP. Furthermore, improved rates of convergence are enabled by the estimated inverse Jacobian. This allows for larger minimisation steps, controlled by the λ step size, which significantly reduces the number of iterations required to achieve convergence. Further computational effort was saved by eschewing the line search routine to optimise λ. Instead, a fixed, user input approximation of λ was used. Although this can increase the required number of inner loop iterations, this option significantly reduces the overall number of costly total energy evaluations and diagonalisations across the calculation. This was achieved without a trade-off in terms of accuracy for the final energy, where the differences in the Helmholtz free energy between the fixed and optimised λ approaches were negligible for small systems. Furthermore, for larger systems, the Pulay mixing routine leads to increased accuracy by achieving convergence within a reasonable number of inner loop

iterations. This contrasts with the linear mixing routine, where a large number of inner loop steps must be specified to achieve convergence at each outer loop step. Although linear mixing would eventually converge with a large enough number of inner loop iterations, doing so dramatically increases the calculation time. Especially considering much of the optimization work is lost between the initial NGWF updates, this makes the initial outer loop steps incredibly computationally costly compared to the Pulay mixing scheme.

The resulting speed-ups facilitated by the Pulay mixing algorithm increase the scope of systems that may be studied. Not only can greater accuracy in energies and forces be achieved by utilising tighter total energy thresholds, the greater efficiency of the Pulay mixing algorithm allows studies which include both larger and greater numbers of systems. The implementations described here and in the preceding chapter are particularly important in Chapter 6, where the ΔG_{ads} for a broad range of large metallic/absorbate interfaces must be calculated for a number of coverages.

Chapter 5

Parameterizing the Implicit Solvent Model for Metallic Work Functions

In this chapter, we study how the implicit solvent method can be parameterized to obtain the work function and potential of zero charge for the Pt(111) surface. Furthermore, by tuning dielectric cavity parameters, we explore how one can calculate the enthalpies of adsorption for phenol on a solvated platinum surface with implicit solvent models. The material in this chapter was published in the Journal of Chemical Theory and Computation [157]:

G. Bramley, M. T. Nguyen, V. A. Glezakou, R. Rousseau and C. K. Skylaris, *Journal of Chemical Theory and Computation*, 2020, **16**, 2703–2715.

This work was performed in collaboration with our colleagues at the Pacific Northwest National Laboratory (PNNL): Dr. Roger Rousseau, Dr. Vassiliki-Alexandra Glezakou, and Dr. Manh-Thuong Nguyen. Data pertaining to the AIMD simulations of the Pt(111)/H_2O were performed by Dr. Manh-Thuong Nguyen. Sections describing the AIMD simulations carried out in CP2K were written by Dr. Nyugen. I have written all other elements of the text. Post-processing of these frames to obtain the work function and potential of zero change values was performed by myself in ONETEP. The above authors contributed to the manuscript with corrections and ideas, and the project was supervised by Prof. Chris-Kriton Skylaris.

5.1 Background

The study of metallic surfaces has garnered significant interest in the modern chemistry and materials communities. Platinum catalysts have received especially close attention given their important role in many emergent technologies such as the conversion of crude bio-oils to transport fuels [18, 33] and the proton-exchange membrane fuel cells

(PEMFCs) [158]. In addition to empirical studies, computational simulations of these systems can provide insight into the mechanisms and properties that underlie these processes, guiding design improvements that increase both yields and turnover frequencies [159]. However, reactivity and selectivity in these applications are strongly influenced by the solvent environment [1, 45, 160], whether through site-blocking at the catalytic surface [161], stabilization of polar transition states or the introduction of side processes involving the solvent [162]. Given these complex effects, one can find numerous examples where the solvent (usually aqueous) environment can either hinder or aid chemical/industrial processes. For example, the rate of hydrodeoxygenation of ketones on MoO_3 catalysts markedly decreases at high H_2O concentrations [161]. In contrast, aqueous solvent activates the oxidation reaction of ethanol in Au nanoparticles, where the initial O_2 dissociation step is more thermodynamically favoured by increases in the binding energy of Au-O, and kinetically favoured by a reduction in the activation energy [163]. Therefore, computationally efficient techniques must be designed to capture the complex phenomena of solvation at the solvent/metal interface, in order to accurately measure the energetics of adsorption which underlie heterogeneous catalytic processes.

A well-known effect of the aqueous environment on metallic surfaces is the reduction of the work function [164], which increases the reductive strength of heterogeneous catalysts [3, 165]. Using DFT simulations of the water monolayer / multilayer at the surface of Pt (111) *Tripkovic et al.* [166] showed that solvent-induced changes in work function can be decomposed into polarization and orientation contributions, as defined in the model of potential zero charge (PZC) model of *Trassati*[167]. The orientation contribution is the work function change produced by the net dipole moment of molecular water, while the polarization contribution corresponds to the electric dipole moment induced by charge transfer/Pauli repulsion [168] effects at the metallic/water interface. However, static configurations fail to represent the dynamic nature of water. In reality, the rotational freedom of solvent causes the overall dipole moment to evolve over time, meaning no single water structure can represent the work function. Although the experimental potential of zero charge is measured as a macroscopic average over a large system, computed values across small simulation cells require an ensemble average of the work function from multiple configurations. *Sakong et al.* [2, 169] performed *ab initio* molecular dynamics (AIMD) simulations of the water/Pt(111) interface, showing that the mean work function decreases to 5.0 eV compared to the vacuum value of 5.6 eV, with a variance of approximately 0.23 eV. Further work by *Le et al.* [170] demonstrated that for a range of metal/water interfaces, the majority of the change in the work function is induced by the Pauli repulsion charge pushback (-1.3 eV for Pt), while the net orientation of water provides only a small contribution across a 10 ps simulation ($\sim$0.0-0.2 eV).

Although AIMD studies are able to reproduce the experimental electrochemical properties of the metallic interface, they are computationally costly. This is especially limiting when considering techniques such as microkinetic modelling, where many (tens to hundreds) elementary steps are used to calculate the global kinetics of a complex reaction mechanism [171]. Additional concerns surround the ability of particular GGA functionals (e.g. PBE, or other GGA methods) to model the structure of water and solvent metal bond lengths [172].

In principle, continuum approaches can emulate the equilibrium properties of solvent/metal interfaces without extensive configurational sampling. Early work showed that the Poisson-Boltzmann continuum model was able to capture the qualitative decrease in the work function [173]. However, the original parameterization of the density dependent self-consistent continuum solvation model underestimates the PZC as well as the differential capacitance (DC) for metallic interfaces [173, 174]. *Hörmann et al.* [175] provided an alternative parameterization of the continuum solvent cavity, where the cavity size was reduced to obtain the PZC of the Pt(111)/H_2O interface obtained with AIMD simulation. However, the computational PZC values from the explicit solvent simulation include the orientational contribution of the work function changes, which cannot be represented in the continuum model owing to the absence of atomistic water. Consequently, the dielectric continuum can only induce work function changes through the polarization contributions (i.e. electron density changes), which can be over/underestimated depending on the dipole moment of the water layer and the structure of the surface under study. Several publications have suggested including a monolayer / multiple layers of explicit water combined with the dielectric continuum (hybrid implicit/explicit methods) in order to better represent the solvent/metal interface [176, 177]. Improvements are apparent for metal oxides, where strong interactions between bonding centers and the atomistic aqueous phase constrain the rotational dynamics of water, which creates a significant dipole moment, thereby introducing a potential offset of up to 1 V to the work function [177]. However, the hybrid method reintroduces the effect of the rotational dynamics of the solvent on the surface, which significantly increases the configuration space that must be sampled to obtain the adsorption enthalpies of adsorbates under aqueous conditions [55]. In contrast to metal oxides, water at noble metal surfaces maintains orientational freedom at room temperature, meaning the potential offset due to the overall dipole contribution is far smaller (between -0.1 to 0.2 V) [170, 178]. However, as will be discussed later, this value can be overestimated by up to 0.2 V, depending on the method used to measure the computational PZC due to the overestimation of the surface potential of water. We show that PZC values for the Pt/continuum interface can be brought closer in line with the experiment by parameterizing against the polarisation contribution to the atomistic work function change. This is demonstrated by performing an AIMD simulation of the water/Pt(111) interface and extracting the changes of the work function due to changes in the electric dipole moment at the surface. This result is then used to reparameterize the

dielectric cavity of the implicit solvent model, representing the work function changes in the absence of orientation effects. We also highlight the quantitative improvements provided by nonlocal van der Waals' (vdW) functionals such as VV10 [179] for the PZC. Furthermore, we compare the electronic structure changes of implicit and explicit interfaces through electronic projected d-band density of states (PDOS) for the atoms of the metallic surface.

A further point that we wish to address is the use of isodensity cavitation models, which define the dielectric cavity along a specified isocontour of electronic charge. Although these models require the optimization of relatively few parameters, we will demonstrate that no single-value of the isodensity can simultaneously give accurate solvation energies for small molecules and metallic work function changes. This is a consequence of homogeneously applying the isodensity parameter across the whole system, combined with the large deviations from the default cavity parameters necessary to obtain accurate values of the PZC. [175] Such changes in the shape of the cavity compromise the precision of the free energy of solvation ΔG_{solv} for species containing light elements, since the default isodensity values are selected to minimise the error of the changes in solvation energy for common organic molecules. However, cavitation schemes such as the soft sphere radii of *Fisicaro et al.* [123] allow one to define separate cavity radii for surface metal species and first-row elements. Using these capabilities, we demonstrate for the first time the ability of the soft sphere model to capture simultaneously the aqueous enthalpy of adsorption of phenol on Pt(111), as well as the solvent work function of the metallic surface.

5.2 Methodology

5.2.1 ONETEP Calculations

Calculations were performed with ONETEP [91], a linear-scaling DFT code. In ONETEP simulations of metallic systems were carried out with the finite-temperature Kohn-Sham DFT approach by *Mermin* [180] which is implemented as a variant of the Ensemble DFT method by *Marzari* [136], which has been reformulated for the localised NGWF framework of ONETEP [83].

The PBE [133] GGA exchange-correlation functional and the non-local vdW correction including the functional rVV10 [179] were used throughout this work. The core states were represented by Projected Augmented Waves (PAW) [98], obtained from the GBRV pseudopotential library [181]. The valence states of the Pt atoms were represented using 12 NGWFs in spherical regions of radii 9.0 a_0, with an electronic configuration of $5p^6 6s^1 5d^9 6p^0$, following the work by *Verga et al.* [182]. The psinc kinetic energy cut-off was set to 850 eV for total energy calculations. Geometry relaxations were conducted

with the BFGS algorithm [183], using a convergence threshold of 5×10^{-3} E_h a_0^{-1} for the maximum force. A Fermi-Dirac smearing width of 0.09 eV is used throughout for electronic occupancies. For systems including explicit water, the slab type Coulomb cut-off approach was used [135, 184, 185] to eliminate electrostatic interactions perpendicular to the slabs. The cutoff point is applied to both the electronic density and the long-range tail of the local pseudopotentials, while the ion-ion Coulomb energy is calculated through a quasi-2D Ewald summation. Projections of the Density of States (DOS) are performed with an angular momentum resolved basis set of pseudoatomic orbital (i.e., the basis functions of NGWFs used to initialise an ONETEP calculation) [149]. Projection is performed up to the d angular momentum channel, and plots are produced with 0.2 eV smearing.

Comparisons to previous benchmark properties for both the Pt fcc bulk and Pt(111) surface in vacuum are provided in the Appendix C.

5.2.2 CP2K Calculations and *ab initio* Molecular Dynamics

CP2K [186] electronic structure calculations were conducted using the PBE GGA exchange-correlation functional [133] corrected with D3-type van der Waals potentials [187] for dispersion interactions. We employed the GPW hybrid basis set scheme [188] in which the Gaussian basis sets of Double Zeta Valence Polarised (DZVP) (for the bulk Pt and the surface of Pt without water) and (Triple Zeta Valence Polarised) TZVP (for a film of water on Pt(111)) quality were used to expand the Kohn-Sham states and a plane wave energy cut-off of 450 Ry was used for electrostatic calculations.

An AIMD simulation was conducted within the NVT ensemble, with the temperature enforced with a velocity-rescaling thermostat [189]. We used modified masses (5 au for H and 10 au for O) and a time step of 1 fs to integrate the equations of motion. We used a frame of a system containing 64 Pt atoms and 52 water molecules, published in a previous work [3]. The system was initially mixed with a 2 ps run at T=1000 K, and subsequently cooled down to 330 K. This was followed by a production run of 24 ps, at T = 330 K. The temperature of 330 K avoids overstructuring of the water layer, which is otherwise known to arise in simulations of liquid water using the PBE functional [190]. We should note that, due to the use of modified masses, which enables better sampling on a shorter trajectory, the dynamics quantities considered do not correspond to the actual correlation time of water at the Pt(111) surface. Rather, they serve as a measure of the convergence of the calculations with respect to the characteristic time constant of the system. Our use of CP2K was motivated by a desire to maintain consistency with our previous simulations of the metal / water interfaces [3]. The resulting structures were eventually used as explicit solvent systems to study the electronic structure and calculate the work function. These calculations were performed in ONETEP, allowing for a consistent comparison between the explicit and implicit solvent models.

5.2.3 Continuum Solvent Model

The continuum solvent model used in this work utilises the Poisson-Boltzmann method formulated by Fattebert, Gygi, and Scherlis (FGS) [120–122] This was implemented with additional dispersion and repulsion terms in ONETEP by *Dziedzic et al.* [114], providing quantitative improvement for the free energy of solvation for neutrals, anions, and cations. Solutions to the Poisson-Boltzmann problem are obtained with the parallel multigrid DL_MG library [110]. For a full explanation, please refer to Section 2.5.

At this stage of the project, we had not yet developed the full SASA+SAV scheme. The following represents an older method for parameterizing the soft sphere model in terms of the SASA-only. The parameters $\boldsymbol{r_i}$ and Δ are determined using a similar error minimisation procedure as outlined in the original soft-sphere scheme [123]. We use the vdW radii by *Alvarez* [129], $\mathbf{r_{vdW}}$, and multiply this set of radii by a scaling factor, f, in order to uniformly increase/decrease the soft sphere radii ($\boldsymbol{r_i} = f\boldsymbol{r_{vdW}}$). The scaling value was varied from f = 1.0 to 1.4 to achieve a minimum mean absolute error (MAE) for the free solvation, ΔG_{solv}. This was performed over a set of 20 neutral molecules containing a range of common functional groups, and results were compared to experimental data obtained from the Minnesota Solvation Database [125]. The lowest MAE (1.09 kcal mol^{-1}) was attained for Δ = 0.5 with a scaling factor of f = 1.33.

To demonstrate the sensitivity of the work function of a Pt(111) surface to variations in cavity size, the values of ρ_0 and $\boldsymbol{r_i}$ (for Pt) were varied between 0.0035-0.04 e/Å^3 and 3.90-5.20 a$_0$ for each respective cavitation model. To ensure that solvent dielectric was excluded from the inside of the Pt slab, we manually set $\varepsilon(\boldsymbol{r})$ to 1.0 from the bottom layer to the top layer of the metallic slab.

5.2.4 Work Function Changes in Solvent

The work function of the bare and solvated metallic slabs can be expressed in terms of absolute potentials (5.1) [176, 190]:

$$\varphi^M \equiv U_M = -\frac{\mu^M}{e_0} + \chi^M, \tag{5.1}$$

$$\varphi^{M|S} \equiv U_{M|S} = {}_M\Delta_S\varphi - \frac{\mu^M}{e_0} + \chi^S, \tag{5.2}$$

where ${}_S\Delta_M\varphi$ is the Galvanic potential difference between the solvent and the metal, $\frac{\mu^M}{e_0}$ is the standard potential of the electron in the metallic species, e is the elementary charge of an electron, and $\chi^{S/M}$ is the surface potential at the solvent/vacuum interface.

The inner (Galvanic) potential of the solvent and the metallic bulk (φ^S and φ^M) are taken as the average of the single particle electrostatic potential through the xy-plane

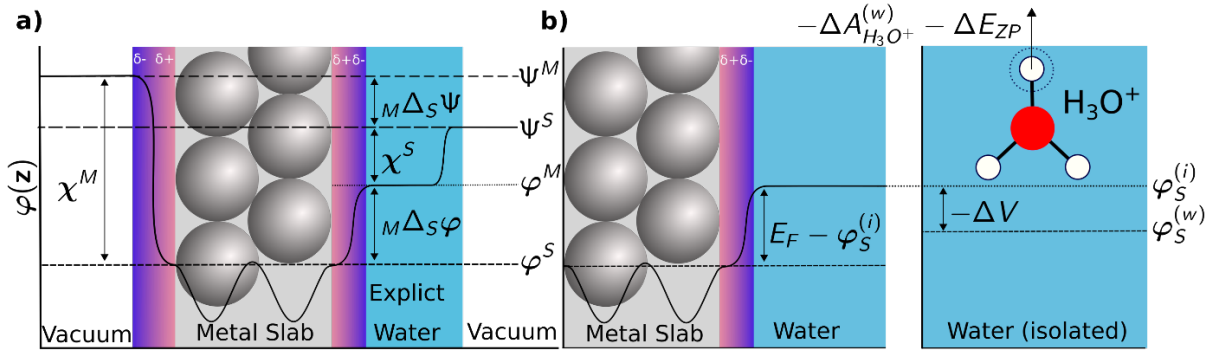

FIGURE 5.1: The variation of the electrostatic potential averaged in the xy-plane ($\varphi(\mathbf{z})$) for a metallic slab interfaced with solvent and vacuum. φ and Ψ shows the inner potential and outer potentials respectively. Two methods of calculating the PZC are: a) the work function method, where the difference between the outer (Volta) potentials of the solvent (ψ^S) and metallic surface (ψ^M) represents the work function change. χ is the surface potential of the indicated phase, and b) the Computational Standard Hydrogen Electrode method, where the Fermi level (E_F) is referenced to the inner potential of water $\varphi_S^{(i)}$. The electrode is converted to the SHE scale by coupling $E_F - \varphi_S^{(i)}$ to the free energy of deprotonation for the hydronium cation calculated in a pure water box, taking into account the zero-point energy of the H-OH_2^+ bond ($\Delta_{DP}A^{(w)}_{H_3O^+} - \Delta E_{ZP}$). Defects of the Periodic Boundary Condition between the two cells are eliminated by subtracting the difference between the bulk electrostatic potentials of water (ΔV).

for a selected section of the phase, while the outer (Volta) potential ($-e_0\psi$) represents the single particle electrostatic potential averaged across a plane vacuum outside the phase indicated by the superscript. The surface potential for either phase (shown as α), χ^α therefore represents the work required to move a particle with elementary charge through the interface between the measured phase and vacuum:

$$\chi^\alpha = \psi^\alpha - \varphi^\alpha, \tag{5.3}$$

From this, the total change in the work function ($\Delta\phi_{tot} = \phi^{M|S} - \phi^M$) can be expressed as the change in the outer potential in the vacuum region of both phases:

$$\Delta\phi_{tot} \equiv {}_M\Delta_S\varphi + \chi^M - \chi^S = \psi^M - \psi^S \equiv {}_M\Delta_S\psi \tag{5.4}$$

In addition to the surface potential of the solvent/vacuum interface χ^S, the solvent layer introduces an additional potential step between the metal and the bulk of water ($_M\Delta_S\varphi$). This represents the contribution to $\Delta\phi_{tot}$ from: a) electron density changes produced by charge transfer or Pauli repulsion mechanisms at the surface; or b) The intrinsic dipole of the solvent at the surface.16 According to the theory of Trasatti [167, 191], these two contributions give the total work function change,

$$\Delta\phi_{tot} = \phi^{Pt|S} - \phi^{Pt} = \Delta\phi_{pol} + \Delta\phi_{orient}, \tag{5.5}$$

where $\Delta\phi_{\text{orient}}$ arises from the net dipole moment of the solvent at the metallic surface,

and $\Delta\phi_{\text{pol}}$ arises from changes of the electronic structure of the metallic surface. $\Delta\phi_{\text{pol}}$ was obtained by subtracting $\Delta\phi_{\text{orient}}$ from the total change in the work function. For individual configurations of the AIMD simulation, $\Delta\phi_{\text{orient}}$ is calculated by performing a separate single point calculation of the frozen water configuration isolated from the metallic slab and taking the Volta potential difference between the vacuum regions either side of the bulk water.

5.2.5 Calculating the Potential of Zero Charge

The potential of zero charge of a surface vs. the standard hydrogen electrode, $U_{PZC}\,(SHE)$ is calculated by coupling the electrode under study,

$$M\,(aq) + e^{-}\,(vac) \rightarrow M^{-}\,(aq)\,, \tag{5.6}$$

to the redox half reaction of the proton,

$$H^{+}\,(aq) + e^{-}\,(vac) \rightarrow \frac{1}{2}H_2\,(g)\,, \tag{5.7}$$

To give the overall electrode reaction,

$$M^{-}\,(aq) + H^{+}\,(aq) \rightarrow \frac{1}{2}H_2\,(g) + M^{\cdot}\,(aq)\,. \tag{5.8}$$

In *ab initio* calculations, the overall potential of this equation can be calculated through: a) the work function method or b) the Computational Standard Hydrogen Electrode (CSHE).

In the work function method, $U_{PZC}\,(SHE)$ is calculated as the difference between the absolute potentials of the two half equations (Equations 5.6 and 5.7) referenced against vacuum [190],

$$U_{PZC}\,(SHE) = U_{M|S}\,(abs) - U^{o}_{H^{+}/H_2}(abs), \tag{5.9}$$

where,

$$U_{M|S}(abs) = {}_{M}\Delta_{S}\varphi - \frac{\mu^{M}}{e} + \chi^{S}_{calc}, \tag{5.10}$$

and,

$$U^{o}_{H^{+}/H_2}\,(abs) = \frac{\mu^{o}_{H^{+}} - \frac{1}{2}\mu^{o}_{H_2}}{e} + \chi^{S}_{exp}, \tag{5.11}$$

where $U^{o}_{H^{+}/H_2}(abs)$ is the absolutely potential of the standard hydrogen electrode, taken as the experimentally determined value of Trasatti [191], 4.44 V. $U_{M|S}(abs)$ is taken as the work function of the metallic/water interface, and χ^{S}_{calc} and χ^{S}_{exp} are the *ab initio* and experimentally determined surface potentials of water.

In contrast, the CSHE approach [170, 190, 192] calculates $U_{PZC}\,(SHE)$ by directly computing the free energy difference between the redox equations (Equations 5.6 and 5.7), where Equation 5.6 is expressed as two steps,

$$H^+\,(aq) \rightarrow H^+\,(g)\,, \tag{5.12}$$

$$H^+\,(g) + e^-\,(vac) \rightarrow \frac{1}{2}H_2\,(g)\,. \tag{5.13}$$

The electrode potential vs. the SHE is then calculated as,

$$eU^o_{PZC} = -E_F - e(\varphi^{(w)}_{wat} - \varphi^{(i)}_{wat}) + \Delta_{DP}A^{(w)}_{H_3O^+} - \mu^{g,o}_{H^+} - \Delta E_{ZP}, \tag{5.14}$$

where E_F is the Fermi Level of the metallic slab and φ_{wat} is the average electrostatic potential of bulk water, with $\varphi^{(i)}_{wat}$ and $\varphi^{(w)}_{wat}$ measured for a water box in the presence of the metallic slab and a pure water box respectively ($\varphi^{(w)}_{wat} = 0$). $\Delta_{DP}A^{(w)}_{H_3O^+}$ (15.35 eV) [192, 193] is the free energy of deprotonation of the hydronium cation calculated *via* Free Energy Perturbation for the insertion of a proton into bulk water, $\mu^{g,o}_{H^+}$ (15.81 eV) [194] is the standard chemical potential of the H^+ ion, and ΔE_{ZP} (0.35 eV) [195] is the zero-point energy of the H-$OH_2{}^+$ bond. $e(\varphi^{(w)}_{wat} - \varphi^{(i)}_{wat})$ is required to correct the uncertainty in the potential reference (Hartree Potential Shift) of $\Delta_{DP}A^{(i)}_{H_3O^+}$ in the metallic/water system relative to the measured value of $\Delta_{DP}A^{(w)}_{H_3O^+}$, where

$$\Delta_{DP}A^{(i)}_{H_3O^+} = \Delta_{DP}A^{(w)}_{H_3O^+} + e(\varphi^{(w)}_{wat} - \varphi^{(i)}_{wat}). \tag{5.15}$$

In terms of the redox equations, $\Delta_{DP}A^{(w)}_{H_3O^+} - \mu^{g,o}_{H^+} - \Delta E_{ZP}$ now corresponds to the change in free energy of Equations 5.12 and 5.13, and $-E_F - e(\varphi^{(w)}_{wat} - \varphi^{(i)}_{wat})$ corresponds to the vertical ionisation energy of the metallic surface corrected for the Hartree Potential Shift.

To allow comparison of $\Delta\phi_{tot}$ and $\Delta\phi_{orient}$ between the work function and the CSHE methods, we can recover $\phi^{Pt|S}$ with the following expression:

$$\phi^{Pt|S} = (eU^o_{PZC} + U^o_{H^+/\,H_2}(abs)), \tag{5.16}$$

which converts the potential vs. the SHE to the absolute potential scale. Furthermore, we obtain $\Delta\phi_{orient}$ by assuming that $\Delta\phi_{pol}$ is equal between the two methods and re-adding its contribution to $\Delta\phi_{tot}$.

5.3 Results and Discussion

5.3.1 Explicit Solvent Work Function Changes

In order to correctly parameterize the work function changes of Pt(111) with the implicit solvent model, we first obtained the value of $\Delta\phi_{\text{tot}}$ in explicit solvent. This was achieved by averaging 90 evenly spaced snapshots (sampled every 0.25 ps) from an equilibrated 24 ps AIMD run performed at T = 330 K (Table 5.1). We analyzed the molecular density and orientation dynamics of water molecules within the water layers to validate the convergence of our simulations (details provided in Appendix C). Simulations were performed on a (6x6) orthogonal unit cell of Pt(111) with four layers.

The key difference between the work function and computational SHE methods is the treatment of the surface potential, χ^S. It is assumed the two values of χ^S exactly cancel (i.e., $\chi^S_{calc} - \chi^S_{exp} = 0$) in both experiment and the CSHE method. However, in the work function method (Equations 5.9-5.11), the value of χ^S calculated implicitly in $U_{PZC}(abs)$ (χ^S_{calc}) and measured experimentally in $U^o_{H^+/H_2}(abs)$ (χ^S_{exp}) no longer exactly match, as χ^S_{calc} is determined by the dipole moment of the water/vacuum interface in the AIMD simulation. χ^S_{calc} varies between 0.15-0.33 V in the literature [175, 196] whereas χ^S_{exp} is measured between 0.13-0.14 V [167, 197]. Therefore in the work function method, the contribution from $\Delta\phi_{orient}$ to $\Delta\phi_{tot}$ can be overestimated by up to 0.3 V, which arises through the additional contribution to $\Delta\phi_{orient}$ from $\chi^S_{calc} - \chi^S_{exp}$. In contrast, the CSHE method removes the influence of χ^S by referencing the Fermi Level to the bulk electrostatic potential of water as opposed to the vacuum potential. This results in a lower value of U_{PZC}(SHE) for the CSHE method compared to the work function method (0.21 V vs. 0.43 V for PBE), where $\Delta\phi_{orient}$ was +0.5 eV and +0.3 for the work function and CSHE methods respectively. We therefore draw a similar conclusion to *Le et al.*, where the contribution of $\Delta\phi_{orient}$to U_{PZC} is relatively small compared to $\Delta\phi_{pol}$.

The value of U^o_{PZC} calculated with the work function method coincides within 0.1 V of previous studies using the same approach but is systematically larger than the computational SHE value by up to 0.2 V for both PBE and rVV10. As each study is well statistically sampled, this suggests a fundamental discrepancy between the methods used to calculate U^o_{PZC}.

In contrast, $\Delta\phi_{pol,}$ is independent of the reference used for the electronic energies, as evidenced by the small quantitative differences between the two methods calculating U_{PZC}(SHE). Furthermore, $\Delta\phi_{pol}$ exhibits a small variance (σ = 0.01 eV) compared to $\Delta\phi_{tot}$ (σ = 0.43 eV). This is a consequence of the rotational freedom of the water molecules across the simulation, which drives the variation of $\Delta\phi_{orient}$, whereas

TABLE 5.1: Interface work function (WF) values, $\phi^{Pt|S}$, where the values in this work are obtained from an average of 90 snapshots from a 24 ps AIMD calculation. $\Delta\phi_{tot}$ is calculated relative to the work function of the system in vacuum found in each respective work. The work function due to charge rearrangement, $\Delta\phi_{pol}$ is measured as the change in dipole due to the change in electron density between the vacuum and solvent system. U_{pzc} is calculated relative to the absolute SHE of *Trasatti* [167], $U^{o}_{H^+/\,H_2}(abs) = 4.44\ V$. Errors are calculated as 95% confidence intervals.

	ϕ^{Pt} /eV	$\phi^{Pt\|S}$ /eV	$\Delta\phi_{tot}$ / eV	$\Delta\phi_{orient}$ / eV	$\Delta\phi_{pol}$ / eV	U_{PZC} / V
WF Method (rVV10)	5.83	5.00 ± 0.13	-0.83 ± 0.13	$+0.49 \pm 0.03$	-1.32 ± 0.03	0.56 ± 0.13
CSHE (rVV10)	5.83	4.80 ± 0.08	-1.03 ± 0.08	$+0.29 \pm 0.08$	-1.32 ± 0.34	0.36 ± 0.08
WF Method[a]	5.51	4.96 ± 0.01	-0.55	0.7	-1.25	0.52 ± 0.01
CSHE[b]	5.8	4.7[c]	-1.1	0.2	-1.3	0.2
Expt.	5.9					0.3[d]/0.33[e]

[a] *Sakong et al.* [169] ($Pt(6 \times 6)/(H_2O)_{144}$), 298 K, RPBE+D3. [b] *Le et al.* [170], 330 K, PBE+D3. [c] Value obtaining by adding $U^{o}_{H^+/\,H_2}(abs)$ to calculated U_{PZC}(SHE). [d] *Cuesta et al.* [164], CO-adsorption technique. [e] *Gomez et al.* [198], CO-adsorption technique.

$\Delta\phi_{pol}$ largely varies as a result of distance of the Pt-H_2O bonds as argued by *Tripkovic* [166].

The interfacial layer and the bulk liquid exchange water molecules at a rate of once every 0.25 ps, indicating these events are accessible on the 24 ps timescale of the current simulation. On the other hand, water dipole orientation relaxation occurs on the $\sim$10.7 ps timescale such that we are able to observe more than 2 relaxation events per molecule during our simulation. This indicates that our current simulation accounts for the critical timescale at which the metal/water interface structure changes and impacts upon $\Delta\phi$.

The dynamics in CP2K were performed with the PBE functional with the Grimme D3 dispersion correction, while the snapshots were assessed with nonlocal and empirical vdW correction schemes (i.e., the rVV10 functional). The empirical approach provides no correction for the nonlocal density changes due to dispersion, meaning density dependent properties (i.e., the work function) assessed with pure PBE and PBE+D3 would be identical to one another for the same structure. However, including the rVV10 functional demonstrates that the nonlocal vdW correction systematically increases the work function and U_{PZC}(SHE) for the Pt(111) surface. However, the description of the structure and dynamics of water varies slightly between these two vdW corrections, particularly at the vacuum air interface [199]. The use of rVV10 as opposed to PBE+D3 for the dynamics may, therefore, have some impact on the value of $\Delta\phi_{orient}$ calculated with rVV10 unaccounted for in this work.

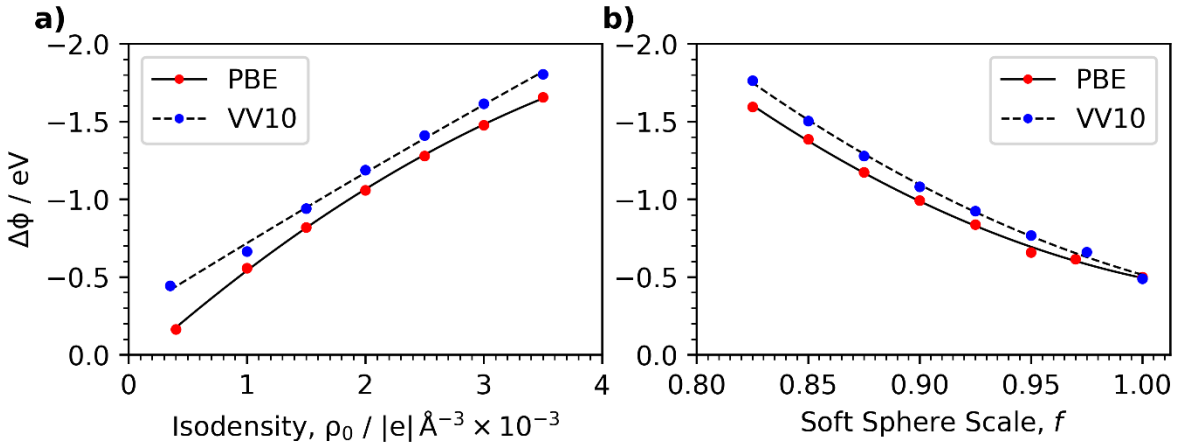

FIGURE 5.2: Variation of $\Delta\phi$ with differing cavity parameters for a Pt(7×6×4) slab: a) Isodensity of the FGS model and b) Scaling factor of the soft sphere model with respect to the default Pt radius of 2.29 Å. Line fitting is performed with a second degree polynomial.

5.3.2 Implicit Solvent Work Function Changes

Previous studies show that water at the Pt surface forms a 5 Å bi-layer with a small orientational dipole at the PZC [170, 178]. Therefore, provided the contribution from $\Delta\phi_{orient}$ remains small, the continuum solvent approach can represent the overall work function change in solvent with reasonable accuracy, with an error similar in scale to AIMD studies. This section will outline the parameters required to reproduce the work function changes of molecular water with the continuum model. This will be achieved using two cavitation models outlined in the Methods section.

The default parameterization of the isodensity model reproduces the qualitative decrease in the work function (-0.16 eV). This is significantly smaller than the 1.23 eV decrease calculated from the explicit solvent systems using PBE. However, by defining the isodensity, ρ_0 or the soft sphere radii such that the bulk dielectric is drawn closer to the atomic centers, the work function decrease can be brought into closer agreement with the explicit approach. This occurs through increased charge re-arrangement towards the surface, which decreases the dipole moment of the metallic centres and the diffuse electron region, resulting in greater reductions of the work function. Careful tuning of the cavity parameters and polynomial fitting then obtain quantitative agreement with the average work function of the explicit solvent systems (Figure 5.2 and Table 5.2). We show that this is achieved with ρ_0 = 0.0024 $e\text{Å}^{-3}$ for the isodensity approach and a scaling factor of 0.87 for the soft sphere of Pt (corresponding to $\mathbf{r}_{vdW}$ = 1.99 Å) for PBE [175, 190].

Previous studies parameterized the cavity sizes against the PZC of AIMD simulations using the work function method, yielding a PZC of 0.5 V vs. SHE for Pt(111) [2, 169]. Excluding the overestimated orientation effects decreases the calculated PZC, where

TABLE 5.2: Default parameters of the isodensity and the soft sphere model cavity model, and the values from a second-order polynomial fitting of the cavity parameters for a Pt(7×6×4) slab (Figure 5.2) corresponding to the average electronic change in the work function ($\Delta\phi_{pol}$ = -1.23 eV (PBE) / -1.32 eV (rVV10)) of the AIMD snapshots, for an initial work function of 5.71 eV (PBE) and 5.91 eV (rVV10). The corresponding values of the PZC are also presented against the absolute SHE scale.

Cavity Model	Soft Sphere Scale (f)	FGS (ρ_0)	U_{PZC}(SHE) / V
Default	1.33	0.00035	-
Re-fitted (PBE)	0.867	0.00237	0.04
Re-fitted (rVV10)	0.872	0.00213	0.15
Hörmann et al.[175]	-	-	0.52
Expt.[164]	-	-	0.29

results obtained with rVV10 result in greater quantitative agreement with experiment (Table 5.2). Furthermore, the use of a larger slab to parameterize these results leads to an upward shift of 0.1 eV relative to the explicit solvent results.

We note that by using this parameterization scheme, the value of $U_{PZC}\,(SHE)$ for the dielectric interface should be offset by approximately -0.2 to -0.3 V from experiment/CSHE and up to -0.5 V with respect to the work function method, depending on the calculated value of $\Delta\phi_{orient}$ [175]. Although one can represent the orientation contribution by including an explicit layer at the metallic surface, one would require dynamic sampling to obtain an accurate value of χ^S_{calc}, which would correctly cancel χ^S_{exp}.

5.3.3 Dispersion Including Functionals and the Work Function

The PBE GGA functional systematically underestimates the vacuum work function compared to the experimental measurements (5.7 vs. 6.0 eV for Pt(111)). However, the dispersion including functional rVV10 systematically increases the work function of the Pt(111) slab in vacuum, bringing the value more in line with experimental LEED studies [200]. By analyzing the change in electron density $\Delta\rho(\mathbf{r})$ for a simple Pt surface, as shown in Figure 5.3, we observe that the rVV10 functional redistributes charge towards the vacuum. This increases the surface dipole, which in turn increases the work function. In vacuum, the work function of the Pt(7×6×4) surface increases from 5.7 eV to 5.9 eV, similar to observations made by *Patra et al.* [201]. Values of $U_{PZC}\,(SHE)$ obtained with the PBE functional for the implicit solvent interface are significantly underestimated with respect to experiment (0.0 vs. 0.3 V). However, Figure 5.3 shows that the work function changes induced by the implicit solvent for PBE and rVV10 closely align. As a result, the inclusion of van der Waals interactions functional increases the calculated values of $U_{PZC}\,(SHE)$ by approximately 0.2 V, thereby improving agreement with experiment. Furthermore, meta-GGAs in conjunction with the rVV10 correction

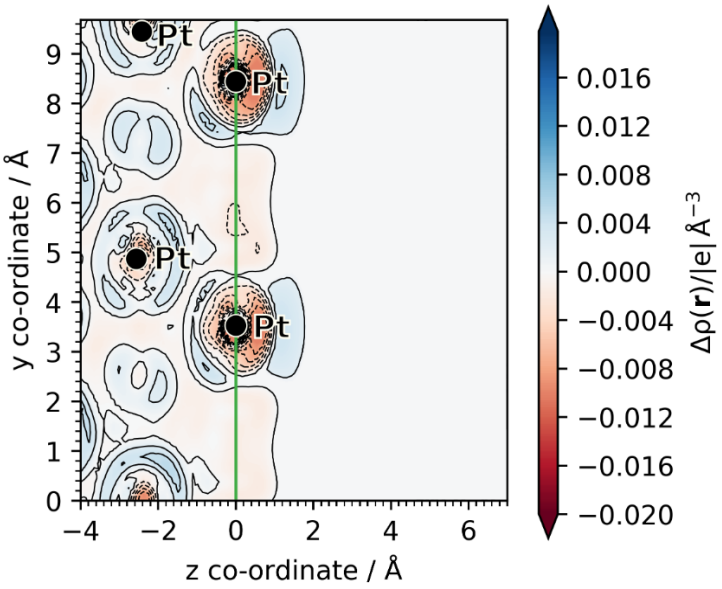

FIGURE 5.3: Electron density difference profile between the rVV10 and PBE functionals for the Pt(111) surface (taken from a single snapshot from the 24 ps AIMD simulation). Blue regions represent electronic charge accumulation and red charge depletion. First layer of the Pt atoms is taken as z = 0. The green line represents the [111] plane at the vacuum interface.

lead to greater quantitative increases in the work function [201], therefore we anticipate that functionals such as SCAN+rVV10 would further improve the description of the diffuse electron region.

5.3.4 Energetics of Adsorption for Phenol on Pt(111)

The key advantage of the soft sphere model over the isodensity model is ability to define the cavity size for individual atoms. The importance of this is highlighted when considering the free energies of solvation of simple molecules on the metallic surface. Without considering the entropy of adsorption, this parameterization of the implicit solvent model provides a simple route for obtaining semi-quantative ΔH_{ads}^{solv}.

As shown in Table 5.3, increasing ρ_0 values lead to worsening errors for the solvation enthalpy ΔH_{solv} of phenol. We note that ΔH_{solv} is calculated as ΔG_{solv} in ONETEP, which includes both cavitation and dispersion repulsion contributions [135]. These terms contain a significant portion of the solvation enthalpy which cannot be trivially separated from the relatively small entropy contribution. Therefore, these terms are included and calculated values are quoted as ΔH_{solv} for ease of comparison with experimental enthalpy values. This ambiguity is resovled in Chapter 6, where we explicitly consider the entropy changes for the constituent elementary steps of adsorption. Considering that the isodensity values are parameterized to minimise the solvation free energy error for neutral, cationic and anionic species, it is unsurprising that deviations

TABLE 5.3: Solvation enthalpy ΔH_{solv} for phenol and Pt(111) with the isodensity and soft sphere models, with the refitted cavity parameters obtained in Section 3.4. All obtained with the rVV10 functional.

Cavity Model	ΔH_{solv} Phenol / kJ mol^{-1}	ΔH_{solv} Pt(111) (per atom) / kJ mol^{-1}
Isodensity (Default)	−31.7	−0.24
Isodensity (Refitted)	−206.6	−27.66
Soft Sphere (Default)	−22.3	0.61
Soft Sphere (Refitted)	−22.3	−16.90
Expt.[a]	−28.9	−9.61[b]

[a] Minnessota Solvation Database [125]. [b] Calculated using the energy of adhesion of solid water to the Pt surface and the surface energy of water (See Appendix C).

from the default ρ_0 value leads to worsening errors. As a consequence, accurate values of both solvation free energies and the metallic work function cannot be computed with the isodensity model. In contrast, the soft sphere model radii can be separately defined for metallic and organic species cavity functions, meaning these quantities can be calculated simultaneously.

The lack of solvation free energy data for transition metal surfaces limits the comparisons that can be made with experiment. However, using a simple modification of the bond additivity model [202], one can obtain an approximation of ΔH_{solv} for a single Pt atom of Pt(111) (See Appendix C and *Singh et al.* [48]). This allows for qualitative assessment of the refitted cavity parameters (Table 5.3). For Pt, the default cavity parameters significantly underestimate the polarisation contribution to the solvation free energy. As a result, in the soft sphere model, unfavorable cavitation/dispersion-repulsion effects (See Section 2.5) dominate over polarisation, leading to a positive ΔH_{solv} value. By drawing the cavity closer to the metallic atoms, the effect of the dielectric continuum is significantly increased. However, both cavitation methods oversolvate the surface Pt atoms, where the refitted soft sphere parameter facilitates agreement to within 7 kJ mol^{-1}. The refitted isodensity parameter significantly overestimates the solvation enthalpy by 18 kJ mol^{-1}, despite yielding the same value of the work function as the soft sphere model. This suggests that the topology of the cavity has significant effect on the solvation energy changes. However, in the case of the soft sphere model, a compromise cavity radius may be found to give adequate agreement with both solvation and work function changes.

To demonstrate the flexibility of the soft sphere approach, we performed a series of calculations to obtain the enthalpic changes for the adsorption of phenol on the Pt(111) surface under aqueous and vacuum conditions (Figure 5.4). Enthalpy changes are compared to the bond-additivity analysis conducted by *Singh and Campbell* [48] (Table 5.4).

Enthalpies of adsorption for phenol (ΔH_{ads}^{vac}) were significantly underestimated by the PBE functional (-122.88 kJ mol^{-1} vs. ∼-200 kJ mol^{-1} [72]). This leads to qualitatively

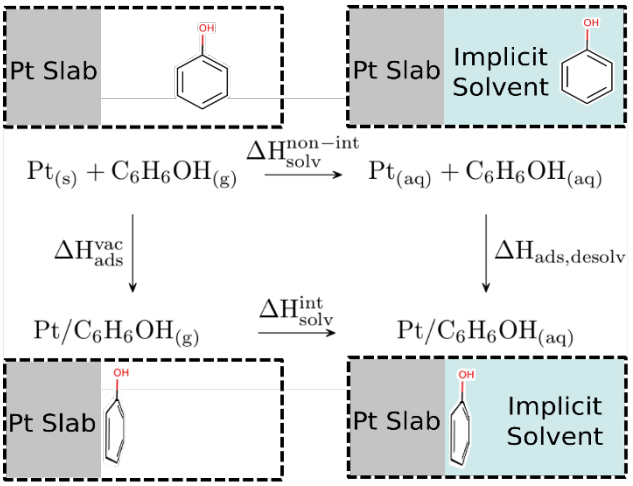

FIGURE 5.4: Thermodynamic cycle for the adsorption of phenol on a Pt(111) (6x6x4) facet under implicit aqueous conditions and in vacuum. ΔH_{solv}^{int} and $\Delta H_{solv}^{non-int}$ describes the enthalpy of solvation for the interacting and non-interacting metal/phenol systems. Values found in Table 5.4.

incorrect outcomes, where the low Pt/phenol bond energy results in an endothermic enthalpy for $\Delta H_{ads,desolv}$. Calculations performed with rVV10 substantially improved agreement with experimental values of ΔH_{ads}^{vac} (-188.69 kJ mol^{-1} vs. $\sim$ -200.0 kJ mol^{-1}), which in turn restores the correct exothermic enthalpy of $\Delta H_{ads,desolv}$.

We aim to compare computed values of ΔH_{ads}^{aq} to values obtained through bond-additivity analysis. To achieve this, $\Delta H_{ads,desolv}$ was decomposed into enthalpy changes from three sources: (1) adsorption of phenol to Pt(111) under aqueous conditions, $-\Delta H_{ads}^{solv}$ (2) desolvation of the occupied Pt atoms, $-\Delta H_{desolv}^{slab}$ and (3) desolvation of one face of phenol, $-\Delta H_{desolv}^{phenol}$,

$$-\Delta H_{ads,desolv} \approx -\Delta H_{ads}^{solv} - \Delta H_{desolv}^{slab} - \Delta H_{desolv}^{phenol}, \tag{5.17}$$

where $-\Delta H_{desolv}^{slab} - \Delta H_{desolv}^{phenol} = -\Delta H_{desolv}^{phenol+slab}$ (Table 5.4). Values of $-\Delta H_{desolv}^{phenol+slab}$ were obtained through two approximations. Similar to the explicit system, adsorbed phenol and its dielectric cavity 'displaces' the implicit solvent from the metallic surface. As phenol adsorbs planar to the Pt surface, the solvation energy lost for the metallic slab can be approximated as a proportion of the metallic slab the dielectric cavity of phenol occupies in the xy-plane multiplied by the total solvation of the metallic slab,

$$-\Delta H_{desolv}^{slab} \approx \frac{area\ of\ phenol\ cavity}{area\ of\ Pt\ surface} \times \Delta H_{solv}^{slab}, \tag{5.18}$$

A further approximation can be made for $\Delta H_{desolv}^{phenol}$ by considering the flat lying adsorption geometry of the molecule with only one side facing the implicit solvent, therefore desolvation can be simplified as $-\Delta H_{desolv}^{phenol} \approx \frac{1}{2} \times \Delta H_{solv}^{phenol}$. Bond additivity analysis calculated a $\sim$57 kJ mol^{-1} discrepancy between experimental values of ΔH_{ads}^{vac} and derived values of ΔH_{ads}^{solv}. However, it has been suggested phenol adopts high local

TABLE 5.4: Enthalpy changes outlined in Figure 5.4 for the adsorption of phenol to the Pt(111) (6x6x4) facet under aqueous conditions. Calculations were performed with the re-fitted cavity parameters obtained in Section 3.4 for both soft sphere and isodensity methods. $\Delta H_{desolv}^{phenol+slab}$ shows the enthalpy change associated with the desolvation of the slab and phenol upon adsorption. ΔH_{ads}^{solv} corresponds to the heat of adsorption under aqueous conditions.

Cavitation Model	ΔH_{ads}^{vac} kJ mol^{-1}	$\Delta H_{ads,desolv}$ kJ mol^{-1}	$\Delta H_{desolv}^{phenol+slab}$ kJ mol^{-1}	ΔH_{ads}^{aq} kJ mol^{-1}
PBE (Isodensity)	−122.88	36.88	−317.36	−280.49
PBE (Soft Sphere)	-	29.09	−134.03	−104.94
rVV10 (Isodensity)	−188.69	−4	−318.38	
rVV10 (Soft Sphere)	-	−35.51	−141.42	−176.93
Bond-Additivity Model	−200.0[a]	−21.05[b]	−116.0	−148.4[c]

[a] *Carey et al.*, SCAC 140 K [58]. [b] *Akinola et al.* [49]. [c] *Singh et al.* [48].

coverages on Pt(111) under aqueous conditions [203]. In vacuum, higher coverage reduces adsorption strength from -200 kJ mol^{-1} to -142 kJ mol^{-1} [58], which suggests the difference between ΔH_{ads}^{solv} and ΔH_{ads}^{vac} results from coverage effects rather than an intrinsic weakening of Pt/phenol bonding. As our calculations are performed with low phenol coverage, one should expect ΔH_{ads}^{solv} to be approximately equal to ΔH_{ads}^{vac}. Calculations using the soft sphere model support these findings, where the difference between ΔH_{ads}^{solv} and ΔH_{ads}^{vac} is correspondingly small (+17.95 kJ mol^{-1} and +11.76 kJ mol^{-1} for PBE and rVV10 respectively). In contrast, the same difference obtained with isodensity cavitation is substantially larger (−157.60 kJ mol^{-1} (PBE) and −129.70 kJ mol^{-1} (rVV10)), largely as a result of the overestimated value of $\Delta H_{desolv}^{phenol}$. This is a consequence of readjusting the ρ_0 cavity parameter, which places the dielectric too closely to the phenol molecule. The soft sphere model circumvents this by maintaining the default cavity size of phenol, while using the newly parameterized cavity radii for Pt from Section 3.4.

It should be noted that the soft sphere model overestimates the value of $\Delta H_{desolv}^{phenol+slab}$ for both functionals and cavitation models. For the preferred model (rVV10 with soft sphere), this leads to a misleading agreement with the high coverage experimental value of $\Delta H_{ads,desolv}$ (-35.51 kJmol^{-1} vs. -21.0 kJ mol^{-1}), where one would expect a significantly more negative enthalpy in accordance with observations of phenol under low coverage. The likely source of this error is the oversolvation of the Pt(111) surface, leading to a comparatively large value of $\Delta H_{desolv}^{phenol+slab}$. The overestimated solvation energy for the Pt surface is moderately reduced when one considers that the implicit solvent solvates a smaller surface area than the explicit solvent (corresponding to 7.5 Pt atoms in the soft sphere model as opposed to 9 Pt atoms for the phenol footprint in the explicit model), which still results in a discrepancy of approximately 40 kJ mol^{-1} for ΔH_{desolv}^{slab}. Conversely, the use of the solvation free energy as opposed to the heat of solvation reduces the solvation estimate of phenol compared to the bond additivity analysis (16 kJ mol^{-1} vs. 25 kJ mol^{-1} for $\Delta H_{desolv}^{phenol}$). Error cancellation between

$\Delta H_{desolv}^{phenol}$ and ΔH_{desolv}^{slab} reduces the error between the bond additivity analysis and the rVV10 soft sphere model for $\Delta H_{desolv}^{phenol+slab}$ to approximately 30 kJ mol^{-1}. To achieve closer agreement with the experimental enthalpy values, further parameterization of the Pt cavity must be performed to align the values of $\Delta H_{desolv}^{phenol+slab}$. This emphasizes a need within the catalysis community to find accurate values of free energy changes of solvation for transition metal surfaces. Values of the work function change will also need to be taken into consideration, as we anticipate that parameterizing to achieve more accurate solvation enthalpies will lead to deviations from the ideal value of $\Delta\phi_{pol}$ obtained in Section 3.6. However, the scheme shown here provides an outline as to how the implicit solvent model with the soft sphere cavitation method can be used to obtain qualitative agreement with experimental enthalpy changes under aqueous conditions.

In the wider context of this work, the work function parameterization method acts as a first guess for the cavity parameters used to calculate ΔH_{ads}^{solv}. In order to obtain more precise energetics of adsorption, it is necessary to minimise the errors of ΔG_{solv}. This will give a more accurate value of ΔH_{desolv}^{slab}, which will in turn better capture ΔH_{ads}^{solv} according to the Hess cycle.

5.3.5 Electronic Structure of the Metal/Solvent Interface

The previous section shows that the implicit solvent model can accurately represent the work function changes for Pt(111) induced by solvent. However, this does not guarantee the electronic structure of the Pt(111)/H_2O interface is accurately represented. Figure 5.5 shows the density difference profiles $\Delta\rho(\mathbf{z})$ averaged across the 90 selected snapshots for both the implicit and explicit solvent models. Overall, implicit solvent replicates the electron displacement of the explicit model, with depletion of electron density from the diffuse region and an accumulation at the metallic surface. This corresponds to the 'pillow effect' model of charge displacement, where the electron density of the surficial metal atoms is repelled away from the adsorbate through a Pauli repulsion mechanism. However, the implicit model does not completely capture certain electronic features of the explicit molecular approach. Most notably, compared to the explicit water layer, the implicit model shifts the peak appearing at $\sim$ 3 Å to 2 Å into single peak. This likely arises due to the absence of electron density from the H_2O bi-layer, meaning there is a smaller quantity of charge that can be displaced from the interface region. Furthermore, the implicit solvent method fails to capture the charge accumulation/depletion region in the subsurface region (-2 Å to 0 Å), while also removing the double-peak character of the electron density change. Further differences are evidenced in Figure 5.6, which shows a yz-plane of the Pt/water interface. In the explicit interface, two dominant bonding modes of water feature in the selected plane [204]: H-up (y $\approx$ 8.5 Å) where bonding takes place through the oxygen lone-pair and; H-down (y $\approx$ 4.6 & 1.3 Å), where water and Pt form a 3-centre bond. In both bonding modes,

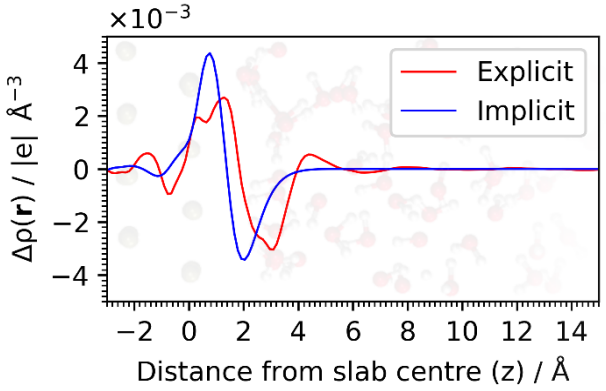

FIGURE 5.5: Density difference profile, $\Delta\rho(\mathbf{z})$, averaged across the 90 snapshots for both the explicit and implicit solvent, with z = 0 taken as the first Pt layer.

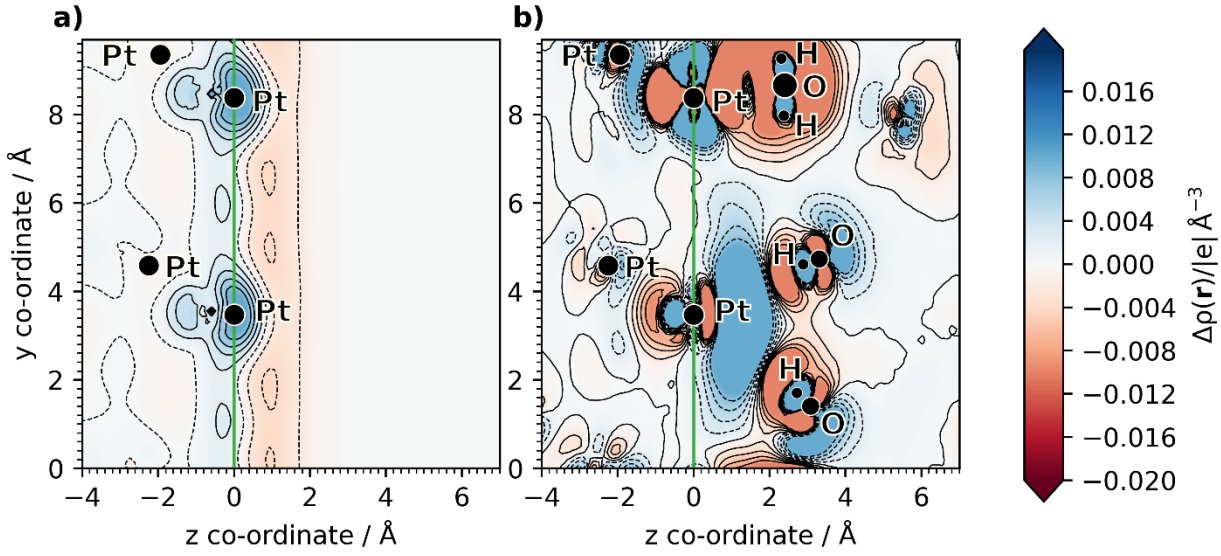

FIGURE 5.6: Charge density difference, $\Delta\rho(\mathbf{r})$ across the yz-plane at x = 4.86 Å, for a selected snapshot of: a) the bare slab with implicit solvent and b) the slab with the explicit solvent system. In these models, the first Pt layer is taken as z = 0. Red and blue regions correspond to electron depletion/accumulation respectively. The green line represents the [111] plane at the solvent interface.

chemisorption induces significant and complex pattern of charge re-arrangement. The H-up arrangement results in a displacement of charge from the interface due to Pauli repulsion between the oxygen lone pair and the metallic electron density [204]. Conversely, the H-down configuration results in charge accumulation between water and Pt, which contrasts with the Pauli repulsion model of the Pt/H_2O interface. Previous AIMD simulations show that the H-up configuration at the Pt(111) surface dominates [178], which justifies the negative value of $\Delta\phi_{\text{pol}}$.

In contrast, the implicit solvent approach displaces the electronic density through a simple charge polarisation mechanism. This results in homogenous charge displacement towards the surface, with no re-hybridization of the surface states taking place

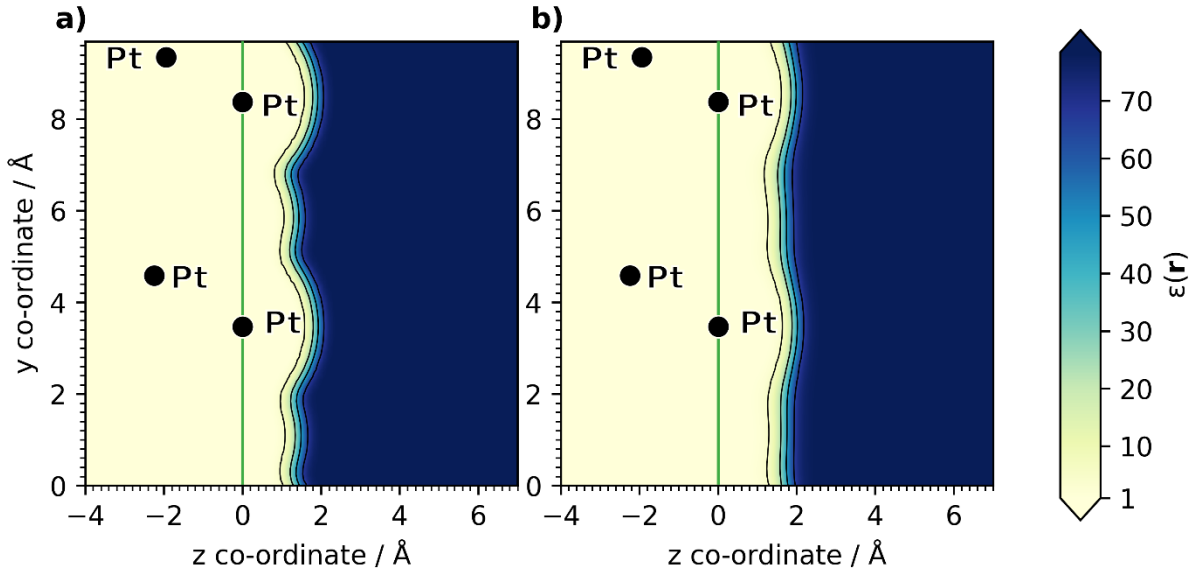

FIGURE 5.7: Dielectric permittivity across an xy-slice of the Pt(111) surface for: a) the soft sphere cavitation model and b) the isodensity model.

due to the absence of bonding interactions. These two modes of charge re-arrangement lead to dissimilar electronic structure changes for the surface states of Pt. This is evidenced by the density of states (DOS) plot of Figure 5.9, which shows the projected d-states of the first layer of Pt atoms for both solvent models, as well as their DOS differences compared to the Pt slab in vacuum. Both solvent models produce a similar degree of distortion, but these induced changes are inconsistent across the d-band. This is most prominent between -4 and -2 eV, where the molecular water induces an increase in the DOS at the -3.4 eV peak and a slight depletion of the -2.6 eV peak, whereas the continuum model reverses this observation. In addition, the implicit interface exhibits larger changes in d-states close to the Fermi level, especially for the trough at -1 eV, whereas the explicit interface has more pronounced distortions in the more negative regions of the d-band. Furthermore, as discussed by *Le et al.* [170], interaction with the Pt surface leads to significant broadening of the band states of water, while the valence band penetrating the Fermi level of Pt is evidence of charge transfer across the surface/solvent interface. Considering this and the charge density difference of Figure 5.9, the charge re-arrangement at the interface must arise through differing mechanisms for implicit and explicit solvent. Both soft sphere and isodensity cavitation models can replicate the overall charge displacement of the equilibrated AIMD simulation of the Pt/(H_2O) interface. The two cavitation models produce subtly different shapes of the dielectric function (Figure 5.7), where the soft sphere approach produces a less planar topology than the isodensity method.

The differences in the electronic structure induced by the soft sphere and isodensity models are relatively minor. Figure 5.8 shows the d-projected DOS and DOS difference plots of the two cavitation models. Here, we observed the two d-bands exhibit broadly

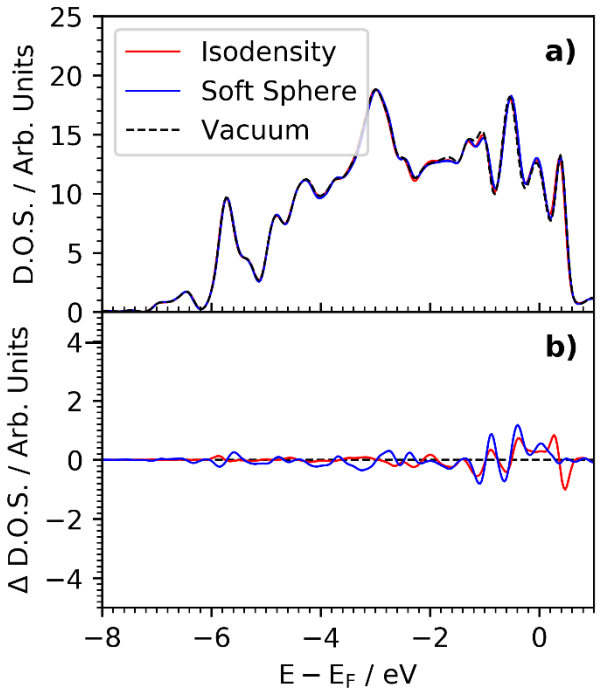

FIGURE 5.8: a) Density of States (DOS) of the surface Pt atoms, projected into the d-band, averaged across 90 AIMD snapshots for the Pt(4x4x4) in vacuum and Pt(4x4x4)/continuum interface with the soft sphere and isodensity model. Energies relative to the Fermi level, E_F. b) Difference of the PDOS d-band with respect to the Pt surface in vacuum.

similar changes between -2 to -6 eV, with slightly more pronounced changes near the Fermi level. These changes likely arise from the different topologies of the dielectric cavities, leading to differing electronic structure changes.

However, despite these differences in the electronic structure, the implicit solvent model replicates the overall displacement of charge towards the metallic surface that we observe with the explicit solvent model, correctly replicating electrochemical properties. This highlights the potential of the continuum solvent approach for further study in both catalytic and electrochemical studies.

5.4 Chapter Summary

This chapter presents a comparison of the Poisson-Boltzmann continuum solvent model and explicit solvent AIMD approaches for the Pt(111)/water interface.

We demonstrated a solute cavity parameterization method for the implicit solvent model to obtain equilibrium solvent work function changes due to the electronic structure changes of the AIMD simulations. Central to this method is the assumption that the

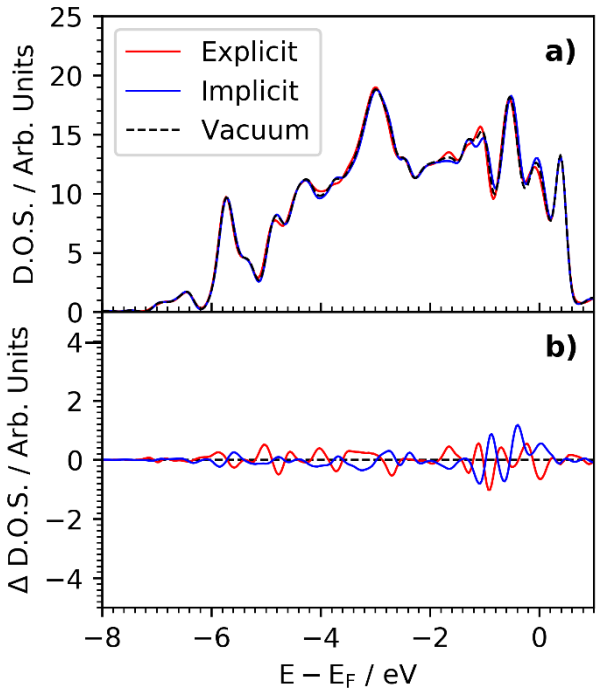

FIGURE 5.9: a) Projected Density of States (PDOS) of the surface Pt atoms, projected into the d-band, averaged across 90 AIMD snapshots for the explicit Pt(4x4x4)/H_2O interface, vacuum Pt(4x4x4) and Pt(4x4x4)/continuum interface with the soft sphere model. Energies relative to the Fermi level, E_F. b) Difference of the PDOS d-band with respect to the Pt surface in vacuum.

work function change arises mainly through electron re-arrangement at the Pt/H_2O interface ($\Delta\phi_{pol}$ = -1.2 eV), while the net dipole of water at the interface makes a relatively small contribution ($\Delta\phi_{orient}$= +0.3 eV). We also remarked on the use of dispersion including functionals in relation to the work function, where use of the rVV10 functional resulted in systematically higher values (0.2 eV), bringing computed values closer to experimental values in vacuum compared to non-dispersion including GGA functionals.

We further compared the ability of soft sphere and isodensity cavitation approaches to obtain accurate solvation enthalpies for the phenol/Pt(111) adsorption process. As the soft sphere model can parameterize the dielectric cavities of organic and metallic species separately, one can capture the solvation energetics of phenol and the metallic surface, as well as the work function of Pt in solvent. As a result, it was possible to corroborate the observations of bond additivity approaches, which predict that the introduction of solvent does not significantly alter the adsorption energy of phenol under low coverage.

Additionally, we showed that the implicit solvent model was able to reproduce the Pauli repulsion induced push-back of the explicit Pt/H_2O interface. However, the projected DOS of the Pt surface showed that the continuum model produced dissimilar

d-band distortions compared to molecular H_2O. This was a consequence of the various modes of water binding to the surface Pt, where overlap between the valence states of water and the d-band result in charge transfer as well as van der Waals' repulsion. Contrastingly, the continuum solvent resulted in a simpler homogenous pushback of charge towards the surface, resulting in less pronounced changes across the d-band, but producing larger changes closer to the Fermi level.

The work outlines a technique to obtain equilibrium properties of the electrochemical interface using a computationally inexpensive continuum solvent model. An implicit solvent model parameterized to obtain the work function correctly characterises the electron density pushback from the diffuse regions to the metallic surface. However, due to limitations of the implicit solvent model, the electronic structure changes induced by bonding are not represented. We also showed the importance of separately parameterizing the dielectric cavity for lighter elements and the heavier metallic atoms. We further note that this scheme does not apply to surfaces which introduce a significant degree of orientational order to the explicit water layer, as the implicit solvent cannot represent the intrinsic dipole of atomistic water. Further developments could use this parameterization approach to construct a more general scheme for a range of metallic surfaces, which will allow computationally efficient simulations of interfacial processes for applications in areas such as electrocatalysis, energy storage, and materials science.

Chapter 6

Adsorption of Organic Compounds to Pt(111)

This chapter provides a more detailed account of the implicit solvent model in its application to adsorption processes. It is shown how the implicit solvent model can be parameterized to obtain the free energy of solvation for the Pt(111) surface. Furthermore, simply computed statistical thermodynamic models are used to calculate the entropy of adsorption of various organic molecules. Finally, the parameterized implicit solvent model is used to obtain the aqueous free energy of adsorption for a range of coverages and orientations of the organic adsorbates. The methods presented provide a diagnostic toolset for evaluating adsorption free energies at low computational cost, which can support the growing body of experimental data in the field of heterogeneous catalysis. The material in this chapter was published in the Journal of Chemical Theory and Computation [205]:

G. A. Bramley, M.-T. Nguyen, V.-A. Glezakou, R. Rousseau and C.-K. Skylaris, *Journal of Chemical Theory and Computation*, 2022, DOI: 10.1021/acs.jctc.1c00894.

This work was performed in collaboration with Dr. Roger Rousseau, Dr. Vassiliki-Alexandra Glezakou, and Dr. Manh-Thuong Nguyen. Data pertaining to the MD simulations of Pt(111)/H_2O were performed by Dr. Manh-Thuong Nguyen. The above authors contributed to the manuscript with corrections and ideas, and the project was supervised by Prof. Chris-Kriton Skylaris.

6.1 Background

Adsorption free energies (ΔG_{ads}) play an important role in determining the overall thermodynamics of heterogeneous catalytic processes. Competitive adsorption between

different species dictates their relative coverages at the metallic surface, leading to reaction bottlenecks if the thermodynamic equilibrium largely favours the adsorption of one species over another [1, 60]. However, the enthalpic and entropic contributions to the free energy are altered by the presence of solvent (ΔG_{ads}^{solv}) [162]. For example, the effects of the solvent decrease ΔG_{ads}^{solv} of phenol on the surface of Pt(111) relative to ΔG_{ads}^{vac} from -220 [58] to 29 kJ mol^{-1} [49]. This is a result of the ΔH_{ads}^{solv} penalty of displacing the aqueous solvent layer, which is reduced by ΔS_{ads}^{solv} gain from the desorption of water into the liquid phase. Combined with further factors such as coverage [58, 206], the accurate calculation of ΔG_{ads}^{solv} requires understanding of the entropic and enthalpic contributions to the constituent steps of the adsorption process.

In atomistic simulations, implicit solvent methods provide a computationally inexpensive alternative to dynamic sampling techniques to calculate the free energy of solvation (ΔG_{solv}). In this method, the molecule is inserted into a cavity within a bulk dielectric medium, where ΔG_{solv} is calculated as a sum of: 1) polar contributions arising from the electronic response to the dielectric medium (ΔG_{pol}), and 2) non-polar contributions such as cavitation and dispersion-repulsion interactions, usually approximated from the shape and size of the cavity (ΔG_{npol}) [13]. Many studies have been conducted to assess the quantitative accuracy of implicit solvent techniques with respect to experiment and explicit solvent simulations - particularly for proteins [207, 208], small molecules [117, 123], and surfaces [209]. For large solute assays, the implicit solvent technique provides solvation free energies with errors of approximately 13.4 kJ mol^{-1} (Generalised Born/SASA) [117] to 2.3 kJ mol^{-1} (Poisson-Boltzmann Model) [123]. In protein studies, publications also assess the structural accuracy of implicit solvent techniques, where the continuum solvent approach can lead to structural errors [208], as well as spurious bottlenecks in the potential energy landscape for protein folding simulations [207]. In contrast, systematic quantitative studies for implicit solvation are rare for the adsorption process on metallic surfaces [49, 56], which means that thorough theoretical and experimental validation is required to perform predictive studies [210].

At the metallic surface, large errors of ΔG_{solv} occur with electron density isovalue based dielectric cavity functions [157], which mainly arise from the use of a single charge isocountour for species which require distinct cavity parameters (i.e., heavy metallic and light organic atoms such as C, H, and O) [55, 157]. The optimal parameters for the solvation of small hydrocarbons dramatically undersolvate Pt atoms at the (111) facet, which require much smaller cavities to give accurate values of ΔG_{solv}. This results in small ΔH^{solv} values compared to experimental binding enthalpies of aqueous solvent to Pt (-0.24 kJ mol^{-1} [157] vs. -9.61 kJ mol^{-1} [67] per atom). Consequently, anomalous values of ΔG_{ads}^{solv} arise, yielding either no improvements in the accuracy of adsorption free energies over standard vacuum calculations [211] or increased stabilisation of adsorption in solvent. In the case of phenol, this leads to ΔG_{ads}^{solv} errors of up to 200 kJ mol^{-1} for the implicit solvent model vs. Cyclic Voltammetry experiments

[49]. However, this inaccuracy is alleviated by separate parameterization of the metallic surface atoms and the organic adsorbates, leading to semi-qualitative agreement with experimental values of ΔG_{ads}^{solv} by decreasing the cavity size for the metallic atoms while maintaining the cavity parameters for the organic adsorbates.

In addition to constructing an implicit solvent model that accurately calculates ΔG_{ads}^{solv}, this study explores the ability of statistical mechanics-based models to calculate ΔS_{ads}^{solv} [212]. In the gas phase, the experiment has determined that the entropy of adsorbed species with strong lateral attractions ($S_{ads}^{gas,0}$) follows a linear relationship with its ideal 3D gas entropy ($S_{Mol}^{gas,0}$) [12],

$$S_{ads}^{gas,0} = 0.68(S_{Mol}^{gas,0} - S_{gas,1D-trans}^{0}), \tag{6.1}$$

where $S_{gas,1D-trans}^{0}$ is the standard molar entropy of the translational gas phase entropy in one dimension. As $S_{ads}^{gas,0}$ is approximately 2/3 of $S_{Mol}^{gas,0}$, it is assumed that translation on the z-axis is eliminated, but diffusion occurs freely on the xy-plane. This implies that entropy can be calculated using 2D ideal gas models to approximate $S_{ads}^{gas,0}$ (known as the free translator (FT) regime [213]. Compared to more sophisticated lateral translational entropy expressions such as the Hindered Translator (HT) model, FT forms an upper bound where all translational microstates in the xy-plane are equally accessible. However, for flat potential energy, which surfaces allow facile diffusion, comparisons of FT with Complete Potential Energetic Sampling (CPES) methods provide estimates of $S_{ads}^{gas,0}$ within 10 % error of full sampling techniques for a range of temperatures above 273 K [69]. To calculate the entropy of solvation, simple models using empirical properties based on SPT (Scaled Particle Theory) and the accentric factor model accurately calculate the ΔS_{solv} of small molecules [212]. Using comparisons to experimental data, this study will assess whether these simple geometric models based on statistical mechanics can provide approximate values for the entropy of solvation, and therefore the overall ΔS_{ads}^{solv}.

The role of coverage effects on the ΔG_{ads}^{solv} of small organics on metallic surfaces must also be understood to assess the accuracy of the implicit solvent model [206]. Although the decrease in adsorption enthalpy due to lateral repulsions between neighbouring adsorbates is observed in vacuum [58, 206, 214], less is known about the impact of coverage on a solvated surface. The experimental data fit to the Temkin isotherm shows that ΔG_{ads}^{solv} has a weak dependence on coverage, with only a small decrease in 0 to 5 kJ mol^{-1} from zero to full coverage for various phenolic compounds on the surface of Pt(111) [45, 49]. In contrast, ΔH_{ads}^{vac} decreases by 121 kJ mol^{-1} over the same coverage range for phenol [58]. It has been proposed that phenol forms islands of high coverage in solvent, even at low concentrations, which emulates the effect of saturation coverage in vacuum. This can rationalise the seemingly low binding energies of phenol at the Pt(111) surface in the aqueous phase, but this theory has not been confirmed [48]. As the implicit solvent simulates the energetics of solvation, this work of this chapter

uses this computationally affordable method to assess the coverage dependencies in the aqueous phase and to evaluate the local high coverage theory. In addition, different orientations of the adsorbates at the Pt(111) surface were tested to determine which geometries best match the experimental energetics of adsorption.

This study validates the approach of separately parameterizing the metallic surface atoms by comparing our results obtained with a commonly used electron density-based formulation of the implicit solvent model. Then, using the increasing quantity of experimental reference data for the aqueous adsorption free energies of organics on the surface of Pt [45, 49] paired with data from the older literature [47], a quantitative analysis is performed for the ΔG_{ads}^{solv} values obtained with the implicit solvent model for the Pt(111) surface. Validation of this new approach is achieved through comparisons with experimental data obtained from CV for the hydrogen underpotential deposition experiments [45, 49] and radiotracer experiments [47], which will allow for error estimations of ΔG_{ads}^{solv} in implicit solvent. Statistical mechanics approximations for the solvation entropy [13, 212] are also used to facilitate further comparison with experiment. Finally, the ability of the implicit solvent model to capture the coverage dependency of ΔG_{ads}^{solv} was investigated. This was achieved by performing simulations over a range of coverages with implicit solvent comparing their trends with experiment.

6.2 Methodology

6.2.1 *Ab initio* Simulations

Ab initio simulations were performed with the ONETEP [91] software package, which performs linear-scaling DFT simulations using localised non-orthogonal generalised Wannier functions (NGWFs) [215] with a basis of periodic sinc (psinc) functions. Simulations were performed using the Ensemble DFT (EDFT) method [136, 180], adapted to the localised NGWF framework of ONETEP by Ruiz-Serrano et al. [83], and further developed to include Pulay mixing of the Hamiltonians.

The optB88-vdW-DF1 functional is used throughout this work [216]. Core states are represented by the Projector Augmented Waves (PAWs) [98] of the GBRV pseudopotential library [181]. The valence shell of Pt is represented with 12 NGWFs with cutoff radii of 12.0 a_0 and an electronic configuration of $5p^6 6s^1 5d^9 6p^0$. The psinc kinetic energy cut-off was set to 850 eV throughout. Geometry optimisations were performed with a LBFGS algorithm [217], using the universal sparse preconditioner [106] with a convergence threshold of $|\mathrm{F}|\mathrm{max} < 3\mathrm{x}10^{-3}$ $\mathrm{E_h}$ $\mathrm{a_0^{-1}}$.

Using the example of a commonly used, conventional plane wave code of a charge based implicit solvent model, comparison calculations for the electron density based

dielectric cavity method were performed using the VASP Plane Wave DFT code package [81, 143, 218] with the VASPsol implicit solvent application [219, 220]. Single-point energy calculations in VASP used the optimised geometries from ONETEP calculations. PAW pseudopotentials [98, 101] were used to represent the core electrons. A kinetic energy cut-off of 450 eV was used. Energy minimisation was performed with the RMM-DIIS procedure using Kerker preconditioning.

In light of the two different kinetic energy cut-offs for VASP and ONETEP, it should be noted that the cut-offs in each code are not completely analogous to each other. Firstly, VASP considers the plane wave cut-off in a sphere in reciprocal space, whereas ONETEP includes the plane waves in a cube. Secondly, ONETEP requires higher kinetic energy cut-offs than plane wave codes to achieve sufficient localisation of the NGWFs [221]. In both cases, the kinetic energy was incrementally increased on a test slab system until meV accuracy in total energy was achieved.

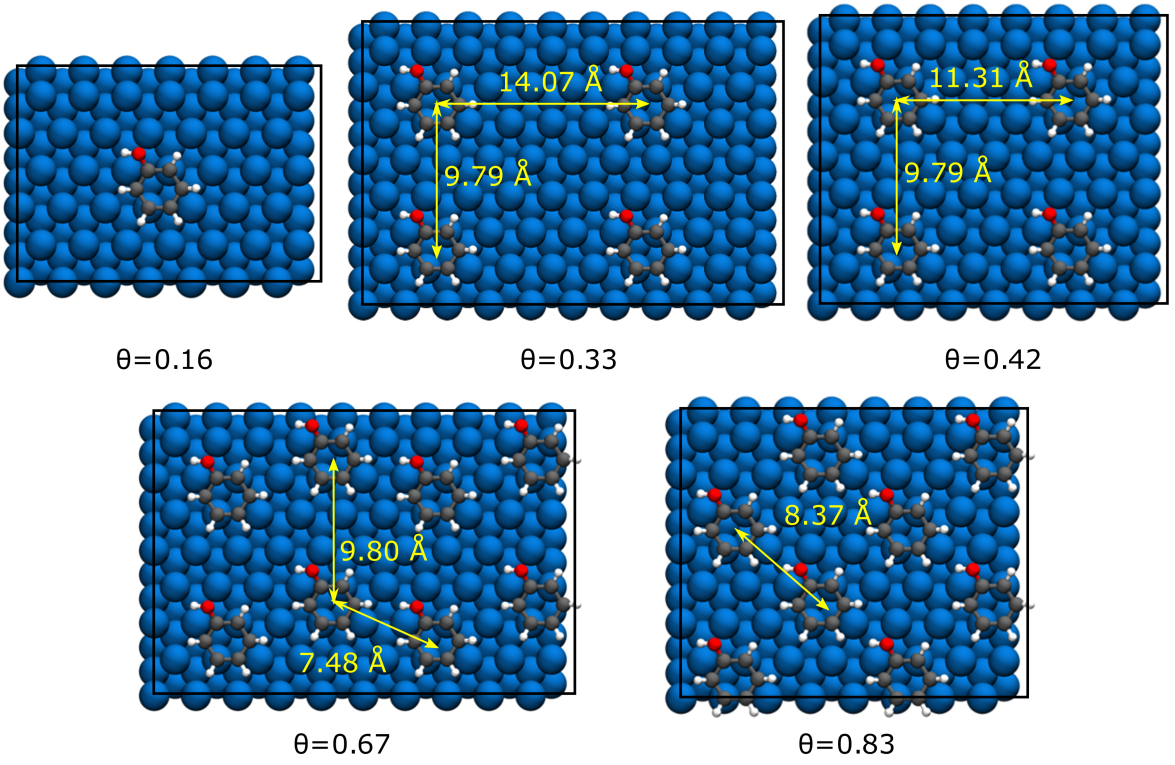

FIGURE 6.1: The coverages of phenol on the Pt(111) facet. All adsorbates horizontal with respect to the surface.

The slab geometries used throughout are based on a four-layer orthogonal Pt(111) facet with dimensions of (7×6), (8×8) or (10×8) where appropriate for the desired coverage. Coverage was also varied by increasing the number of adsorbates on the surface. Figures 6.1 and 6.2 demonstrate the variation in coverage for the surface slabs under study for phenol and cyclohexanol. Appendix D shows a similar scheme for furfural and hydrogen geometries. Large simulations cells were necessary as ONETEP uses the

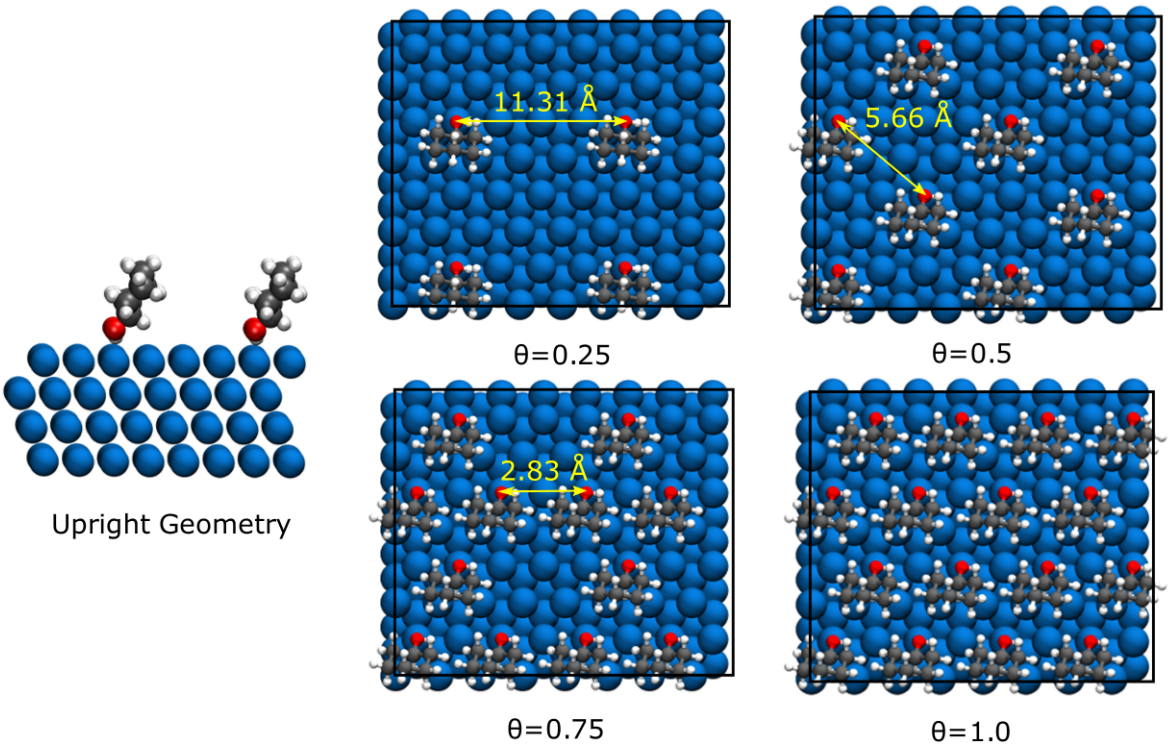

FIGURE 6.2: The coverages of cyclohexanol on the Pt(111) facet. All adsorbates horizontal with respect to the surface.

Γ-point only for total energy calculations, meaning the unit cell in real space must be large enough to sufficiently sample reciprocal space for a single **k**-point.

6.2.2 Implicit Solvation

Implicit solvent calculations were performed using ONETEP's Poisson-Boltzmann (PB) implicit solvent module [114], where solutions to the PB equation are calculated using the bespoke DL_MG multigrid parallel solver [110]. The free energy of solvation (ΔG_{solv}) is expressed as:

$$\Delta G_{solv} = \Delta G_{pol} + \Delta G_{npol} \tag{6.2}$$

where ΔG_{pol} is the electronic response of the molecule to the bulk dielectric and ΔG_{npol} is the nonpolar contribution composed of the cavitation free energy and dispersion-repulsion contributions. ΔG_{npol} is calculated using a linear relationship between the Surface Accessible Surface Area (SASA) and the Surface Accessible Volume (SAV):

$$\Delta G_{npol} = (\alpha * \gamma)S + pV \tag{6.3}$$

where γ is the surface tension of the solvent (0.07415 N m^{-1}), p is the solvent pressure (-0.00035 GPa) and α is an additional scaling factor used to improve the fit of ΔG_{npol} to the experimental values of ΔG_{solv}.

The dielectric cavity is defined using the Soft Sphere atomic radii based cavity functions of *Fisicaro et al.* [123]. In this model, the dielectric cavity is defined as:

$$\epsilon(\mathbf{r}_i, \{\mathbf{R_i}\}) = (\epsilon_0 - 1)\left\{ \prod_i h(\{r_i^{vdW}, \Delta, f\}; \|\mathbf{r} - \mathbf{R}_i\|) \right\} + 1 \tag{6.4}$$

where ϵ_0 is the bulk dielectric constant (78.65), and h is a set of continuous, atom centered distance-based functions defined between 0 and 1, expressed as:

$$h(\{r_{vdW}^i, \Delta, f\}; \|\mathbf{r} - \mathbf{R}_i\|) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\|\mathbf{r} - \mathbf{R}_i\| - f r_i^{vdW}}{\Delta}\right)\right] \tag{6.5}$$

which is defined by parameters Δ (the smearing width of the dielectric cavity function), r_{vdW}^i (the radius of the cavity of atom i taken from the vdW radius set of Alvarez [129]), and f (the linear scaling factor), while taking $\|\mathbf{r} - \mathbf{R}_i\|$ as an argument (the distance between the position of point $\mathbf{r}$ and atom i ($\mathbf{R_i}$)). The scaling factor of f = 1.20 is used to minimise the MAE of a test set of 20 neutral molecules with respect to experimental values [125] (details about the ONETEP implementation can be found in the method section of ref[157]). The scaling factor α (Equation 6.3) was further parameterized to minimise MAE with respect to the experimental ΔG_{solv} of 271 neutral molecules ($\alpha = 0.86$). Pt uses a van der Waals radius of (2.29 Å), and f is varied to match the experimental values of ΔG_{solv} of a single Pt atom of the (111) facet. We outline this process in Section 6.2.5. We modelled this approach on our previous publication, where f is varied to obtain the correct work function of Pt(111) in solvent, which provides initial guidance in obtaining optimal values of ΔG_{solv} [157]. To prevent the erroneous inclusion of dielectric within the metallic surface, a region within the simulation cell is defined where $\epsilon(\mathbf{r}) = 1$ from the top to bottom layer of the Pt slab. For the two high coverage geometries ($\theta = 0.75$ and 1.0) of cyclohexanol in vertical orientation, the exclusion region is extended through the adsorbate to prevent the creation of dielectric regions between the metallic slab and cyclohexanol. Total energy calculations in solvent are performed using the relaxed geometries of the vacuum phase, as force calculations for the soft sphere cavity model within our implicit solvent scheme are yet to be implemented. However, for the organic adsorbates under study, the implicit solvent leads only to minor geometric changes [222].

The charge-based implicit solvent model of VASPsol was used to compare against our implementation of the Fisicaro method. Default settings for the H_2O solvent were used. In contrast to the soft sphere method, the charge based cavity in VASPsol method is defined using the charge density, $n(\mathbf{r})$:

$$\epsilon(n(\mathbf{r})) = 1 + \frac{1}{2}\mathrm{erfc}\left(\frac{\ln\left(n(\mathbf{r})/n_c\right)}{\sigma\sqrt{2}}\right)\Big(\epsilon_0 - 1\Big), \tag{6.6}$$

where similar to the soft sphere model, the complimentary error function defines a

smoothly switching function between 0 and 1. n_c defines the charge isocontour of the function where the complimentary error function is 0.5, and σ defines the width of the switching function.

6.2.3 Calculation of ΔG_{ads}^{solv}

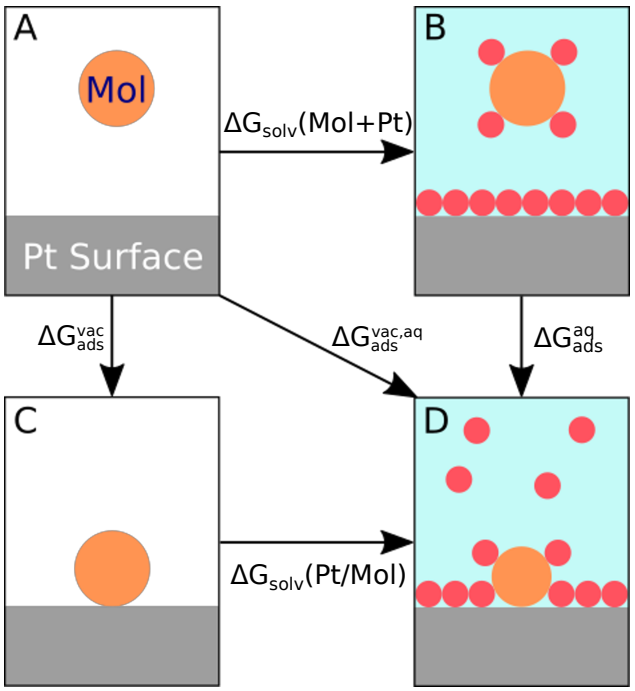

FIGURE 6.3: The process of adsorption according to four difference reference states is shown where A) The molecule and Pt(111) are isolated from one another in vacuum. B) Both the molecule and Pt(111) are isolated from one another in the solvent phase. C) The molecule is adsorbed onto the Pt(111) surface. D) The molecule is adsorbed onto the Pt(111) surface in the aqueous phase, which releases a number of water molecules into the aqueous phase relative to state B.

The free energy of adsorption in the vacuum phase (ΔG_{ads}^{vac}) is calculated as the difference in free energies between the adsorbed system in vacuum (M/Mol) and the isolated metallic slab and molecule (M+Mol),

$$\Delta G_{ads}^{vac} = G(M/Mol_{(vac)}) - (G(M_{(vac)}) + G(Mol_{(vac)}), \tag{6.7}$$

which corresponds to the change in free energy between states A and C in the Hess cycle of Figure 6.3. Assuming that the overall entropy change for the $M_{(vac)}$ component of the system is negligibly small, ΔS_{ads}^{vac} is approximated solely through entropy changes experienced by the molecule, and ΔH_{ads}^{vac} through the total energy differences 0 K from the DFT calculations (equation 6.7). The entropy of the isolated molecule, $S(Mol_{(vac)})$ can be decomposed into translational, rotational, and vibrational components [212],

$$S(Mol_{(vac)}) = S_{Mol}^{vac,3D} = S_t^{vac,3D} + S_r^{vac,3D} + S_v^{vac,3D}. \tag{6.8}$$

$S_v^{vac,3D}$ is small compared to the other components, so it is assumed $\Delta S_{v,ads}^{vac,3D} = 0$. The method for calculating the translational and rotational terms is shown in Appendix A.1. Alternatively, $S(Mol_{(vac)})$ could be taken as the empirical standard state gas phase entropy, but as solvation entropies in this study are calculated with analogous statistical mechanical approximations, the same calculation of the entropy as that of the gas phase is used for the sake of consistency.

The entropy of the molecule adsorbed on the metallic surface is approximated by using the 2D ideal gas approximation such that,

$$S(M/Mol_{(vac)}) = S_{Mol}^{vac,2D} = S_t^{vac,2D} + S_r^{vac,2D} + S_v^{vac,2D}, \tag{6.9}$$

where $S_v^{vac,2D}$ is 0, and the entropy components are derived in Appendix A. In this approximation, it is assumed that the molecule can freely translate across the metallic surface, but is constrained to the xy-plane. We further assume this holds at high coverages, as noted by *Campbell et al.* [12]. We note that the lost degrees of rotational freedom are potentially dependent on the binding mode of the molecule, especially in cases where non-yaw rotations would result in desorption (eg. phenol on Pt with four binding centres). This possibility will be explored in the results section.

Overall, ΔG_{ads}^{vac} is then calculated as,

$$\Delta G_{ads}^{vac} = (H(M/Mol_{(vac)}) - (H(M_{(vac)}) + H(Mol_{(vac)}))) - T((S_{Mol}^{vac,2D} - S_{Mol}^{vac,3D}), \tag{6.10}$$

where $S(M_{(vac)})$ is assumed to be negligible.

Calculating the free energy of adsorption under aqueous conditions (ΔG_{ads}^{solv}) requires further considerations. The implicit solvent calculates entropy components related to solvation within ΔG_{solv}, but the entropy of adsorption must be accounted for in a similar fashion to eq. 6.10. First, starting with the change of free energy from state B to D in Figure 6.3,

$$\begin{aligned} \Delta G_{ads}^{solv} = &(H(M/Mol_{(vac)}) - (H(M_{(vac)}) + H(Mol_{(vac)}))) \\ &+(\Delta G_{solv}(M/Mol) - (\Delta G_{solv}(M) + \Delta G_{solv}(Mol))) \\ &-T(S_{Mol}^{vac,2D} - S_{Mol}^{vac,3D}), \end{aligned} \tag{6.11}$$

where the first three terms represent the enthalpy of adsorption for a single point energy calculation in the gas phase, and the following three terms represent the free energy changes of solvation calculated by the implicit solvent model, as expressed in Equation 6.2. We assume the change in entropy due to solvation is included in ΔG_{solv}, therefore the change in entropy due to adsorption in the last two terms are equivalent to the gas phase.

6.2.4 Upper and Lower Bound Entropy of Adsorption (ΔS_{ads}^{solv})

Furthermore, the change in adsorption entropy in solvent (ΔS_{ads}^{solv}) is approximated using the solvation entropy methods of *Garza* and the 2D ideal gas model. Using different approximations for ΔS_{solv} for the Pt(111) surface, an upper and lower bound for ΔS_{ads}^{solv} is established.

The entropy for the molecule in the solvent phase is calculated using the approach of *Garza* [212], which adopts the ideal gas 3D equation to the entropy of solvation as,

$$S_{Mol}^{solv,3D} = S_{Mol,t}^{solv,3D} + S_{Mol,r}^{solv,3D} + S_{Mol,c}^{solv,3D}, \tag{6.12}$$

where $S_c^{solv,3D}$ is the cavitation entropy. A full derivation of terms is discussed in the Appendix A.2.

The overall entropy of adsorption in the aqueous phase is expressed as,

$$\Delta S_{ads}^{solv} = S_{M/Mol}^{solv,2D} - (S_M^{solv,3D} + S_{Mol}^{solv,3D}), \tag{6.13}$$

where $S_{M/Mol}^{solv,2D}$ consists of the translational and rotational terms of $S_{Mol}^{vac,2D}$, the ΔS_{solv} for the metallic surface and adsorbate.

$$\Delta S_{ads}^{solv} = S_{Mol}^{vac,2D} - (S_{Mol,r}^{solv,3D} + S_{Mol,t}^{solv,3D}) + \Delta S_{Mol,c}^{solv} + \Delta S_M^{solv}, \tag{6.14}$$

where $S_{Mol}^{solv,3D}$ is separated out into translational and rotational components, $\Delta S_{Mol,c}^{solv}$ is the change in entropy due to cavitation for the molecule, and ΔS_M^{solv} is the change in solvation entropy for the metallic surface.

As there is no experimental measurement of ΔS_M^{solv} (i.e., $S_M^{solv,3D}$), reasonable approximations must be derived. Here, an upper and lower bound is proposed from simple approximations based on the mobility of water at the Pt(111)/H_2O interface. *Akinola et al.* [49] propose a scheme using the ice-like bi-layer model for the aqueous metallic interface [223, 224]. Therefore, the entropy of solvation is approximated as the entropy of transition from the liquid to the solid phase, that is, the entropy of fusion of water [74],

$$\Delta S_M^{solv} = \Delta S_{fus} = 0.72 * 25.1 \text{ J K}^{-1} \text{ mol}^{-1} = 18.1 \text{ J K}^{-1} \text{ mol}^{-1} \text{ (per Pt atom)}, \tag{6.15}$$

where approximate the number of water molecules released upon adsorption by multiplying the cross-sectional area of the molecular binding interface (A_{ov}) by the number of H_2O molecules per Pt atom (0.72 [225]). However, water retains a significant amount of mobility at the Pt/H_2O interface at 298 K on a ps time scale [226], meaning surface waters are expected to be more disordered than $H_2O_{(s)}$. Therefore, the entropy of fusion approximation ΔS_{solv} is interpreted as an upper bound.

In contrast, ΔS_M^{solv} can be approximated as the cavitation entropy ($\Delta S_M^{solv} = \Delta S_{M,c}^{solv}$) through the accentric factor method (Appendix A.2. In this method, it is assumed adsorbed water retains the mobility of a weakly bound solvation shell (Equation A.20) [212, 227]. The metal cavity volume dependent term (V_M) varies with the number of slab layers, which leads to a nonphysical dependence of $S_{M,c}^{solv}$ on N. Therefore, we consider the V_M for a single layer of Pt(111) only and normalise by a factor of two so that only one side of the slab is considered. However, the water on the surface of Pt (111) is less mobile than surfaces with weaker adhesion enthalpies (eg graphene), indicated by lower translational and rotational diffusion coefficients [226]. As a result, we infer that the cavitation entropy calculated through the accentric cavity method is a lower bound for the entropy of solvation of the Pt atom. Establishing the upper and lower bounds establishes an expected range of ΔS_M^{solv} values for different coverages and adsorbate orientations.

The packing factor of water ($\mathcal{G}(R_M, R_S)$, Equation A.18) is different for metallic and organic solutes. Consequently, $S_{M/Mol,c}^{solv,2D}$ cannot be calculated directly and $\Delta S_{ads,c}^{solv}$ is taken as the approximate volume of the cavity lost for the surface and molecule after adsorption (Equation 6.12). For the metallic cavity, we multiply $S_{M,c}^{solv}$ per unit area by the cross sectional area of the M/Mol binding region. We approximate the area as the cross-sectional van der Waals' area of the molecule at the binding interface, A_{ov}. We calculate the loss of $S_{Mol,c}^{solv}$ upon adsorption as a proportion of the surface area lost by the molecule upon adsorption of the metallic slab. Overall, we express gain of cavitation entropy as,

$$\Delta S_{Mol,c}^{solv} = \frac{A_{loss}}{A_{Mol}} S_{M,c}^{solv,3D}, \tag{6.16}$$

and

$$\Delta S_{M,c}^{solv} = \frac{A_{ov}}{A_M} S_{Mol,c}^{solv,3D}, \tag{6.17}$$

where A_{ov} is the cross-sectional area of the adsorbate on the metallic surface, A_M the cross-sectional area of the metallic slab, A_{loss} is the van der Waals' surface area lost by the molecule upon adsorption, and A_{Mol} is the total van der Waals' surface area of the molecule. All volumes and surface areas are calculated using the van der Waals' radii of Alvarez [129].

6.2.5 Approximating ΔG_{solv} for Pt(111)

Parameterization of the implicit solvent model follows the method of our previous publication [157], where the cavity radius of Pt atoms for a four-layer (7x6) Pt(111) facet are varied through a range of scale factors (f) from the default cavity radius of 2.29 Å [129]. By performing a third order polynomial fit, we obtain a value matching the experimental ΔG_{solv} of a single Pt atom at the surface. However, as there is no direct measurement

of ΔG_{solv}, for the Pt(111) surface we must approximate this value as a sum of measurements of ΔH_{solv} and $T\Delta S_{solv}$,

$$\Delta G_{solv} = \Delta H_{solv} - T\Delta S_{solv}. \tag{6.18}$$

ΔH_{solv} was calculated as the difference between the bond energy of the solid Pt(111)/water interface ($E_{adh,Pt/H_2O_{(s)}}$) and the surface energy of solid water ($\gamma_{H_2O_{(s)}}$, renormalised to the surface area of a single Pt site), which accounts for the heat released by water upon reentry into the bulk,

$$\Delta H_{solv,Pt} = E_{adh,Pt/H_2O_{(s)}} - \gamma_{H_2O_{(s)}}. \tag{6.19}$$

As shown by *Singh et al.* [48], $E_{adh,Pt/H_2O}$ can be taken as the Pt-H_2O bond energy, where it is assumed $E_{adh,Pt/H_2O_{(s)}} = E_{adh,Pt/H_2O_{(l)}}$. $E_{adh,Pt/H_2O}$ is taken as 0.32 J m^{-2} [67] and $\gamma_{H_2O_{(s)}}$ as 0.102 J m^{-2}, from which we obtain the change of bonding energy per unit area. Assuming a single Pt atom occupies 7.68 Å^2 (from a lattice constant of 3.92 Å), it can be approximated that $E_{adh,Pt/H_2O_{(s)}} = 14.81$ kJ mol^{-1} and the equivalent $\gamma_{H_2O_{(s)}}$ for Pt is 5.04 kJ mol^{-1}. Overall, this yields $\Delta H_{solv,Pt} = -9.76$ kJ mol^{-1} (per mol of Pt as opposed to H_2O released).

The contact angle of the solvent drop with respect to the surface plane (θ_C) also allows access to ΔG_{solv} for the Pt surface. This quantity can be measured experimentally and theoretically via. molecular dynamics simulations [226]. The Young-Dupres equation ($W_{adh} = \gamma_{H_2O}(\cos(\theta_C) + 1)$) allows access to ΔG_{solv} of the Pt(111) facet through the relation proposed by *Fisicaro et al.* [123],

$$\cos(\theta_C) = \frac{\Delta G_{solv}}{\gamma_{H_2O}}. \tag{6.20}$$

However, previous simulations have shown that a range of noble metal surfaces with adhesion energies between 179-267 mJ m^{-2} share the contact angle of $\theta_C = 0$ [226]. For highly hydrophilic surfaces, where $\Delta G_{solv} > \gamma_{H_2O}$ in Equation 6.20, θ_C no longer differentiates between their solvation free energies. We therefore do not believe this method is extendable to other noble metal surfaces, and limit our discussion to the methods proposed earlier in this discussion.

ΔS_{solv} has not yet been measured experimentally for the Pt(111) surface. In Section 6.2.4, we propose two approximations for the entropy of water at the metallic aqueous interface where: the upper bound $\Delta S_{fus} = \Delta S_{solv}$ (where ΔS_{fus} is the fusion entropy) or a lower bound where $\Delta S_{solv} = \Delta S_{M,c}^{solv,3D}$ (Equations 6.15 and 6.17). The entropy value for the upper bound is $\Delta S_{solv} = -5.41$ kJ mol^{-1} and the lower bound is $\Delta S_{solv} = -2.58$ kJ mol^{-1} per surface Pt(111) atom.

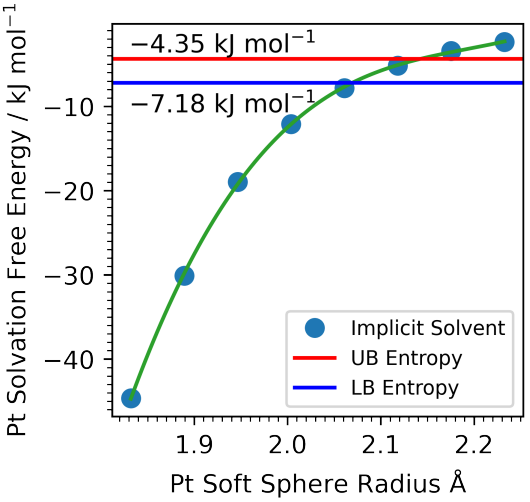

FIGURE 6.4: The change of ΔG_{solv} per Pt atom with respect to the vdW radius of $r_{vdw} = 2.29$ Å scaled between $f = 0.8$ to 0.975 of the (111) facet for a 4 layer (7×6) supercell. Upper bound (UB) and lower bound (LB) entropy lines are shown.

Overall, this yields ΔG_{solv} values of -4.35 kJ mol^{-1} and -7.18 kJ mol^{-1} for upper and lower entropy bounds respectively, which we then used to parameterize the implicit solvent model (Figure 6.4). Performing a third-order polynomial fit for the values of ΔG_{solv} gives optimal values of $f = 0.904$ (lower bound entropy) and $f = 0.936$ (upper bound entropy) for the r_i^{vdW} of Pt. The scaling factor for organic van der Waals' radii are maintained at their optimal value of $f = 1.201$.

6.2.6 Calculation of $\Delta G^{diff}_{ads,\theta}$ and $\Delta G^{int}_{ads,\theta}$ with Temkin Isotherm

The integral heat of adsorption measures the total heat released per molecule over a range of coverages, where we define coverage θ throughout this paper as,

$$\theta = \frac{N_{Mol}}{N_M} \frac{N_M}{N^{sat}_{Mol}}, \tag{6.21}$$

where N_{Mol} is the number of adsorbed molecules, N_M is the number of surface metal atoms and N^{sat}_{Mol} is the number of molecules required to achieve saturation coverage, such that $\theta = 1$ at full coverage. Calorimetry and hydrogen UPD experiments measure the differential forms of $\Delta H_{ads,\theta}$ and $\Delta G_{ads,\theta}$ [58], which take into account the varying heat released with an infinitesimal change of coverage at θ. This extends to measurements of thermodynamic properties calculated in solvent through the Temkin isotherm, where the free energy is measured as a function of coverage via. varying concentrations [47]. The following is a brief account of the method used by *Chaudhary et al.* [206] to convert integral to differential heats of adsorption. The desorption energy represents

the integral enthalpy of adsorption coverage θ,

$$\Delta H^{int}_{ads,\theta} = \frac{E(M/Mol) - (E(M) + NE(Mol))}{N}, \tag{6.22}$$

where coverage is controlled by the size of the periodic supercell in the xy-plane and and the number of adsorbates adsorbed to the facet. The differential heat/free energies ($\Delta H^{diff}_{ads,\theta}$) is represented by the change of heat/free energy with respect to the number of molecules:

$$\Delta H^{diff}_{ads,\theta} = \left(\frac{\partial Q}{\partial N}\right)_{\mathfrak{A},\Gamma}, \tag{6.23}$$

where Γ relates the surface site density and the number of surface sites occupied by the adsorbate, and $\mathfrak{A}$ is the surface area of the absorbate. *Chaudhary et al.* propose conversions of integral to differential heats/free energies of adsorption [206], by expressing the integral heat of adsorption (for a single molecule) in terms of the differential heat of adsorption:

$$\Delta H^{int}_{ads,\theta} = \frac{1}{\theta}\int_0^{\theta} \Delta H^{diff}_{ads,\theta} d\theta, \tag{6.24}$$

which when rearranged yields,

$$\Delta H^{diff}_{ads,\theta} = \frac{d\theta \Delta H^{int}_{ads,\theta}}{d\theta} = \Delta H^{int}_{ads,\theta} + \theta \frac{d\Delta H^{int}_{ads,\theta}}{d\theta}. \tag{6.25}$$

By measuring $\Delta H^{int}_{ads,\theta}$ over varying coverages, we can obtain a linear relationship between the integral heat of adsorption and the coverage by means of Equation 6.22.

$$-\Delta H^{int}_{ads,\theta} = -\Delta H^{int}_{ads,\theta=0} + b\theta, \tag{6.26}$$

where b is a constant which describes a linear dependence of $\Delta H^{int}_{ads,\theta}$ on coverage. This is more readily recognised as the Temkin isotherm. This expression then allows the evaluation of the differential heat of adsorption in Eq. 6.25,

$$-\Delta H^{diff}_{ads,\theta} = (-\Delta H^{int}_{ads,\theta=0} + b\theta) + \theta\left(\frac{d}{d\theta}(-\Delta H^{int}_{ads,\theta=0} + b\theta)\right) = -\Delta H^{int}_{ads,\theta=0} + 2b\theta, \tag{6.27}$$

where $db/d\theta = 0$. As a result, the values of the adsorption enthalpies obtained through DFT calculations can be compared to experimental data obtained with UPD or calorimetry methods. Unless otherwise indicated, the heats and free energies calculated with DFT are described in terms of the differential heat of adsorption.

6.3 Results and Discussion

6.3.1 Entropy of Adsorption from First Principles

Campbell's Bond Additivity model treats the heats of adsorption of phenol to Pt(111) in the aqueous environment with a simple bond breaking/bond formation approach [48]. These calculations were enabled by recent experimental data, describing the bond energy of the Pt/H_2O interface, and enthalpies of adsorption and solvation for the molecule.

Figure 6.5 shows the entropy of adsorption in the liquid phase with upper and lower bounds compared to experimental values [47]. We include the effect of orientation, placing compounds either in the horizontal orientations (with aromatic groups/aliphatic chains parallel to the metallic surface), or vertical orientations (where co-ordination takes place through the oxygen groups, with the aromatics/aliphatic groups perpendicular to the metallic surface).

ΔS_{ads}^{solv} are larger in their horizontal than vertical geometries. The surface area footprint of the horizontal geometries tends to be larger than that of the vertical geometries, resulting in a greater amount of surface area lost upon adsorption. This corresponds to the desorption of a greater number of H_2O molecules, making horizontal adsorption more entropically favourable.

First addressing the aromatic compounds, compared to experiment, we note that the vertical geometries drastically underestimate ΔS_{ads}^{solv}. This applies to the upper bound entropy values, where errors of $\Delta S_{ads,solv}$ are between 117-155 J K^{-1} mol^{-1} with respect to experiment. In contrast, far better agreement is observed in the horizontal geometry, where the experimental values lie close to the calculated upper entropy bound (between 10-37 J K^{-1} mol^{-1} error).

In contrast, the experimental values of ΔS_{ads}^{solv} for butanol and valeric acid show the greatest agreement with the vertical geometry model at the low coverage limit. This contrasts to the binding mode of alcohol in the vacuum phase at low coverages, where weak physisorption between the Pt surface and aliphatic chain leads to a horizontal adsorption [228]. The values of ΔS_{ads}^{solv} show a strong correlation with the size of the molecule, varying between +93 J $^{-1}mol^{-1}$ for benzene and +198 J K^{-1} mol^{-1} for naphtholic acid. Entropy gain increases because of the greater number of desorbed waters from the metallic surface and loss of co-ordinating waters from the cavity of the molecule. This trend is also reflected in the experimental data. However, *Akinola et al.* suggest that there are large errors for the experimental data presented, which we discuss in detail in Appendix D. With further experimental data, we believe that the model presented in this study has the potential to provide geometric insights into the binding of small organic molecules to the metallic surface with minimal computational cost.

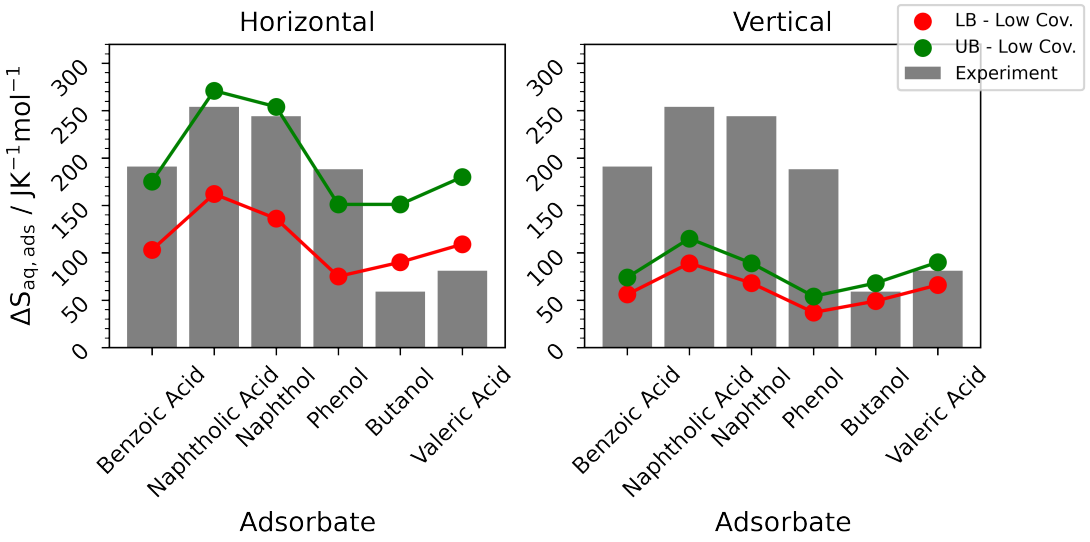

FIGURE 6.5: ΔS_{ads}^{solv} for adsorbates to the Pt(111) facet. Shown for the low coverages (Single adsorbate to the Pt(7x6x5) surface). Left plots shows adsorbates where the aliphatic chain/aromatic group are parallel to the metallic surface, and the right where they are perpendicular and bound through the oxygen groups. Comparisons performed between upper and lower approximate entropy bounds to experimental values of FTIR experiments [47].

Similarly to the bond additivity model [48, 229], which predicts the values of ΔH_{ads}^{solv} for the Pt(111) facet, the entropy models presented provide a low-cost qualitative approach to represent trends in adsorption entropy in the aqueous phase. By providing a lower entropy bound, we show structural trends can be represented. Furthermore, by combining this approach with bond additivity models for ΔH_{ads}^{solv}, one can obtain approximate upper and lower bounds for the overall ΔG_{ads}^{solv}. As more experimental data are gathered and the aqueous phase adsorption process is better understood, these early models can be refined further to reduce their overall errors.

6.3.2 Comparison of Electron Density and Soft Sphere Cavity Models at Low Coverages

Figure 6.6 shows the values of ΔG_{ads}^{solv} obtained with the upper and lower entropy bounds of ΔG_{solv} compared to experimental data. To show the improvements yielded by this approach, we also present results from the charge density dielectric cavity model of VASP. The aromatic adsorbates were measured in their strongest binding geometries in vacuum, that is, with their hydrocarbon rings parallel with the Pt surface. Cyclohexanol is measured in the horizontal orientation, and hydrogen measures the dissociation process of ($\frac{1}{2}H_{2(g)} \rightarrow 0.5H*$). As previously discussed, the charge based model parameterized to capture the ΔG_{solv} of molecules composed of light elements is ill-suited to calculate the ΔG_{solv} of heavy/metallic atoms [157]. This is demonstrated by the relatively higher errors observed for all molecules with respect to experiment.

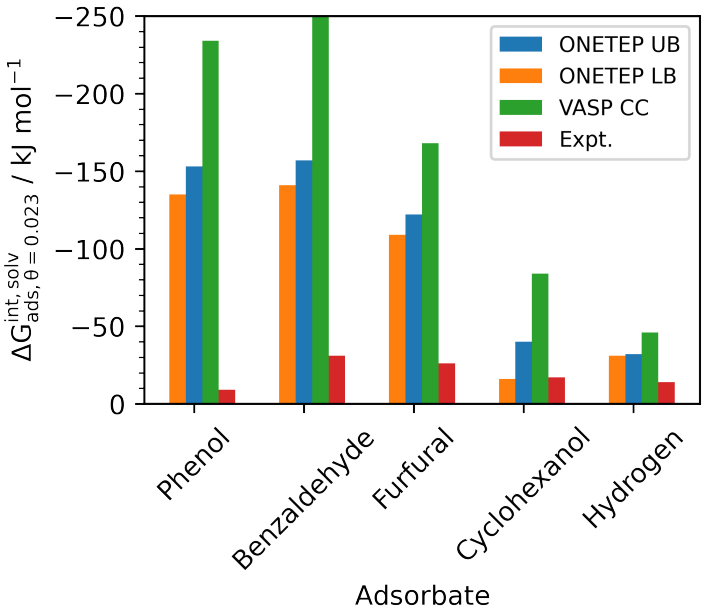

FIGURE 6.6: Comparison of $\Delta G^{int,solv}_{ads,\theta=0.023}$ obtained using the soft sphere cavity model (upper and lower entropy bounds) and the charge based dielectric cavity model of VASP.

At the selected coverage ($\theta \approx 0.15$ for all aromatic species, $\theta = 0.05$ for hydrogen and $\theta = 0.09$ furfural and cyclohexanol) the correct qualitative trend of decreasing ΔG^{solv}_{ads} compared to ΔG^{vac}_{ads} is observed using the reparameterized soft sphere model. The upper and lower bound parameterization yields qualitatively accurate ΔG^{solv}_{ads} values compared to experiment for cyclohexanol and hydrogen dissociation. However, for aromatic compounds in their horizontal geometries, we observe errors on a scale of 100 kJ mol^{-1} compared to experiment.

However, as shown in Figure 6.7 for phenol, QM/MM simulations of Pt(111) adsorption with explicit water yield $\Delta G^{int,solv}_{ads,\theta=0.18} = -78$ kJ mol^{-1} compared to $\Delta G^{int,vac}_{ads,\theta=0.18} = -195$ kJ mol^{-1} in the gas phase [10]. This relatively large decrease in adsorption free energy is better aligned with the experimentally observed decreases in $\Delta G^{int,solv}_{ads,\theta}$. This is driven by driven by larger predicted adhesion free energies ($G_{adh,Pt/H_2O(l)} = -21.8$ kJ mol^{-1} per Pt atom [10]), compared to experimental values of $E_{adh,Pt/H_2O(s)} = -15.1$ kJ mol^{-1} [67] used in this work to calculate ΔG_{solv} for Pt(111). As a result, the loss of water of is further energetically disfavoured, reducing $\Delta G^{int,solv}_{ads,\theta}$. The magnitude of the observed errors implies that there may be further effects not fully accounted for using the vacuum phase geometries at low coverages.

As coverage and orientation are known to have a dramatic effect on adsorption enthalpy in vacuum, we investigate the role of these effects in Section 6.3.3. Across each simulation, the upper entropy bound provides the greatest decrease in ΔG^{solv}_{ads} compared to experiment. However, the overall difference between the upper and lower bound is relatively small (10 kJ mol^{-1} for phenol). This error would increase at higher

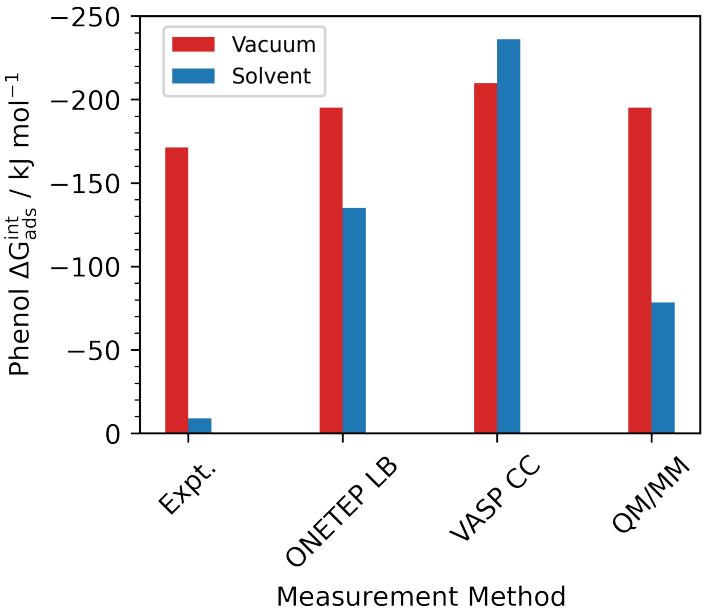

FIGURE 6.7: Comparison between methods of calculating $\Delta G^{int,vac}_{ads,\theta}$ for phenol (horizontal, $\theta = 0.15$) on Pt(111). Obtained using the soft sphere cavity model in ONETEP (lower entropy bound) and the charge based dielectric cavity model of VASP. Also compares experimental values of $\Delta G^{int,vac}_{ads}$ the gaseous and aqueous phase at 298 K [45], and QM/MM simulations of phenol adsorption at $\theta = 0.19$ [10].

temperatures and larger adsorbates, but remains small for all the systems under study. We will briefly discuss other possible factors contributing to errors. Firstly, each molecule is adsorbed to their most stable adsorption site, when the overall adsorption energy is closer to an average of several weaker adsorption sites [206]. This leads to an overestimate for ΔG^{vac}_{ads} of approximately 12 kJ mol^{-1} for phenol [206]. Secondly, the optB88-vdW functional is known to overbind phenol molecules to the Pt(111) by approximately 11 kJ mol^{-1} at the zero coverage limit [214, 230]. Assuming that both effects are at play in the implicit solvent calculation, this leads to an approximate overestimation of $\sim$ 20 kJ mol^{-1}. Furthermore, we note that the definition of entropy is an approximation based on the 2D ideal gas model. This assumes that the adsorbate is at temperatures sufficiently high enough to achieve barrierless diffusion across the metallic surface. Although the 2D ideal gas formulation appears to hold at temperatures of 150 K for simple alcohols and alkanes, the strong binding between Pt(111) and aromatic molecules may further impede translational motion. Additionally, we approximate that the entropy of adsorption is equivalent in vacuum and solvent, where the changes entropy due to restrictions of movement in solvent are accounted for solely through the implicit solvent model. In theory, a portion of this entropy contributed should be accounted for by the use of well parameterized implicit solvation model, but as an untested approximation, this must still be considered as a source of error.

Under the definition of integral ΔG^{solv}_{ads} at low coverages, discrepancies of up to 100 kJ mol^{-1}

occur for the molecules under study. However, in all experimental data, ΔG_{ads}^{solv} becomes more exothermic relative to ΔG_{ads}^{vac}. We have shown that by reparameterizing the cavity of the Pt atom while maintaining the optimised cavity parameters of the organic adsorbate, the implicit solvent model correctly replicates this trend, but other factors such as coverage, orientation and the definition of ΔG_{ads}^{solv} must be considered to achieve meaningful experimental comparisons.

6.3.3 Temkin Isotherms of Implicit Solvent for Horizontal Geometries

Using increasing coverages by varying the cell size and number of adsorbates as shown in the methodology, we investigated the change in adsorption free energy in solvent.

TABLE 6.1: Linear fits to the Temkin isotherm of organic adsorbates on the Pt(111) surface according to $\Delta G_{ads,\theta}^{int} = -\Delta G_{ads,\theta=0}^{int} + b\theta$. Implicit solvent values shown for the lower bound entropy estimate. All values in kJ mol^{-1}

Adsorbate	$\Delta G_{ads,\theta=0}^{int,vac}$	b (Vacuum)	$\Delta G_{ads,\theta=1}^{int,vac}$	$\Delta G_{ads,\theta=0}^{int,solv}$	b (Solvent)	$\Delta G_{ads,\theta=1}^{int,solv}$
Hydrogen Atom	−26	16	−20	−30	12	−18
Phenol	−167	34	−133	−112	15	−98
Furfural (Flat)	−159	36	−123	−99	17	−116
Cyclohexanol (Vertical)	−26	−14	−40	9	−18	−9

TABLE 6.2: The predicted $\Delta G_{ads,\theta}^{diff}$ in vacuum in solvent compared to existing experimental data. Calculated for $\theta = 1$. Implicit solvent values shown for the lower bound entropy estimate. All values in kJ mol^{-1}.

Adsorbate	$\Delta G_{ads,\theta=1}^{diff,vac}$	$\Delta G_{ads,\theta}^{diff,vac}$ (Expt.)	$\Delta G_{ads,\theta=1}^{diff,solv}$	$\Delta G_{ads,\theta}^{diff,solv}$ (Expt.)
Hydrogen Atom	−5	−24[a] / −17[b]	−6	−20[a] / −15[b]
Phenol	−98	-	−82	−9[45]
Furfural	−88	-	−133	−26[49]
Cyclohexanol	−54	-	−27	−17[45]

[a] *Ab initio* MD[1]. [b] Van't Hoff [1].

In the vacuum phase, the Temkin isotherm (Equations 6.26 and 6.27) provides an accurate model for the change in adsorption free energies with increasing coverage. Maximising the nearest-neighbour distance between each adsorbate yields good correlation with the experimental heats of adsorption for phenol [206]. However, the change in adsorption energies with respect to coverage is dramatically weaker in solvent than in vacuum. This is demonstrated by the lower value of b (Equation 6.26). However, the source of this change is unclear. Organic adsorbates are proposed to cluster in areas of high local coverage, which means that in the local environment, $\Delta G_{ads,\theta}^{diff,vac}$ is measured at approximately saturation ($\theta = 1$) regardless of the adsorbate concentration [45]. The second possibility is that the solvent screens the repulsive interactions between adsorbates. In this case, adsorbates still maximize their nearest neighbour distance to avoid lateral repulsions, but the effect is significantly weakened by the presence of solvent.

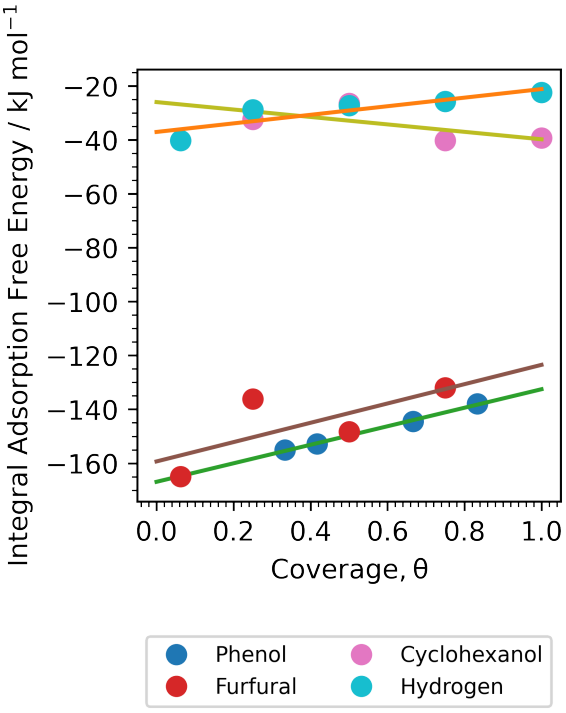

FIGURE 6.8: Integrated free energy change of adsorption in vacuum across a range of coverages on Pt(111). All coverages normalised to the respective saturation coverage of each molecule ($\theta_{sat} = 1, 0.25, 0.25, 0.15$ for hydrogen, furfural, cyclohexanol and phenol respectively). Linear fit to the Temkin isotherm $\Delta G^{int}_{ads,\theta} = \Delta G^{int}_{ads,\theta=0} + b\theta$. Horizontal geometries used for all adsorbates.

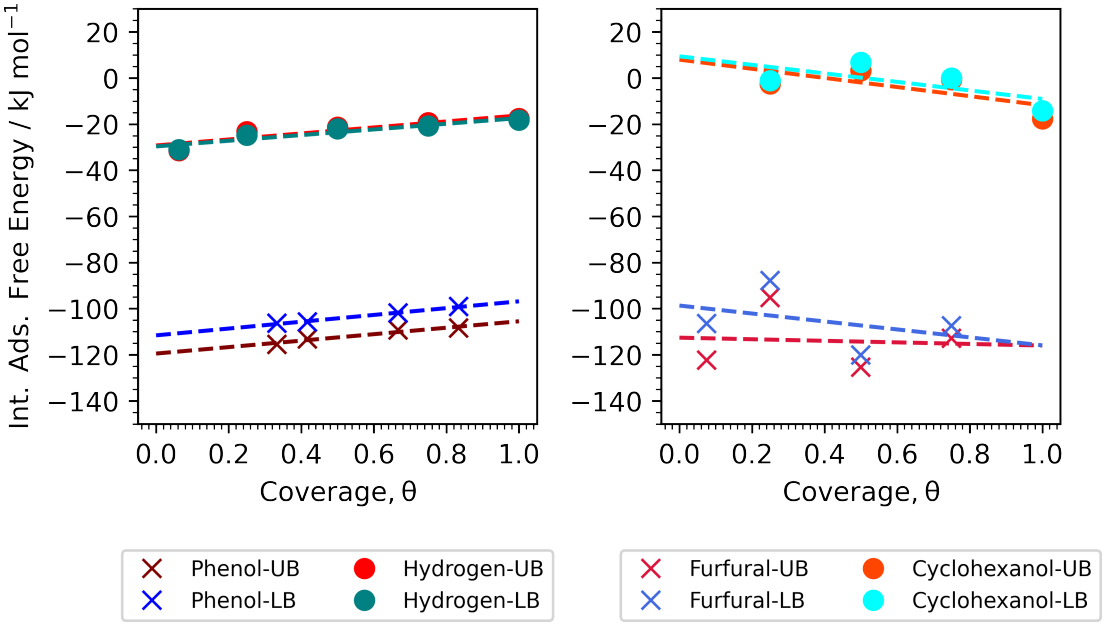

FIGURE 6.9: Integrated free energy change of adsorption in solvent across a range of coverages on Pt(111). All coverages normalised to the respective saturation coverage of each molecule ($\theta_{sat} = 1, 0.15$ for hydrogen and phenol respectively). Linear fit to the Temkin isotherm $\Delta G^{int}_{ads,\theta} = \Delta G^{int}_{ads,\theta=0} + b\theta$. Horizontal geometries used for all adsorbates apart from cyclohexanol.

To investigate the effect of coverage, we utilise the parameterized solvent model and fit $\Delta G^{int,vac}_{ads,\theta}$ and $\Delta G^{int,diff}_{ads,\theta}$ to the Temkin isotherm (Figures 6.8 and 6.9). This will either support or refute the assumption that adsorption proceeds as predicted in the vacuum phase, and this will present better guidance for future study. The tabulated values of the Temkin isotherm are shown in Tables 6.2 and 6.1.

The dissociation of H_2 in the vacuum phase ($\Delta G^{diff,vac}_{ads,\theta=1}$) compares favourably with the results obtained from experiment, but underestimates the overall change in free energy by 10 kJ mol^{-1} (Table 6.2). To meaningfully compare our results with experiment, we note that the values of *Yang et al.* [1] were measured relative to the gaseous reference state of H_2, while our results measure adsorption according to the reaction from state B to D of Scheme 6.3. To align the experimental values to the solvated H_2 reference state, we subtract the entropy of solvation for molecular hydrogen from the experimental and *ab initio* MD values ($\Delta S_{solv}(\frac{1}{2}H_2) = -55$ J K^{-1} mol^{-1} [231]).

In the experiment, the overall free energy change of adsorption in the aqueous phase is 6 kJ mol^{-1} lower than that of the vacuum. This is reflected in the implicit solvent results, where $\Delta G^{int,solv}_{ads,\theta}$ relative to $\Delta G^{int,vac}_{ads,\theta}$ decreases by 5 kJ mol^{-1} for the upper bound of Pt(111) solvation entropy and 4 kJ mol^{-1} for the lower bound. Because b is larger in vacuum than in solvent (Table 6.1), $\Delta G^{diff,solv}_{ads,\theta=1}$ is very close in value despite $\Delta G^{int,solv}_{ads,\theta}$ values consistently being lower than $\Delta G^{int,vac}_{ads,\theta}$. Given these small observed changes in $\Delta G^{int,solv}_{ads,\theta}$, we propose the unfavourable dissociation of hydrogen on Pt(111) compared to vacuum is largely driven by the insolubility of H_2 in water. AIMD techniques further support this, showing that the presence of solvent had only a minor influence on the energetics of the Pt-H bond (~ -2 kJ mol^{-1} [232]). However, we note the diffusion coefficients of adsorbed H on the Pt(111) surface are lower in the aqueous phase ($D_{xy} = 7 \times 10^{-5}$ cm^2 s^{-1} in solvent and $D_{xy} = 33 \times 10^{-5}$ cm^2 s^{-1} in vacuum), leading to a further reduction of entropy upon adsorption [1]. This is also accompanied by a larger decrease in the diffusion coefficient for H_2 in the first aqueous layer compared to the gas phase ($D_{xy} = 361 \times 10^{-5}$ cm^2 s^{-1} in vacuum and $D_{xy} = 11 \times 10^{-5}$ cm^2 s^{-1} in solvent), suggesting that most of the entropy loss for H_2 results from solvation. This rationalises the large pressures required to obtain partial coverages (10 atm yields $\theta \approx 0.2$) in contrast to adsorption in the gas phase, where saturation coverage of H on Pt(111) is achieved with 0.1 atm. However, without knowing the entropy of solvation of H_2 within the experimental conditions of *Yang et al.* [1], quantitative analysis is limited.

However, for aromatic compounds in their horizontal geometry, we observe overestimations of $\Delta G^{diff,solv}_{ads,\theta}$ and $\Delta G^{int,solv}_{ads,\theta}$ observed at low coverages. Even at high coverages, the $\Delta G^{diff,solv}_{ads,\theta}$ for phenol and furfural is dramatically overestimated compared to the experimental values ($\Delta G^{diff,solv}_{ads,\theta=1} = -82$ and -133 kJ mol^{-1} vs. -9 and -26 kJ mol^{-1}). However, we observe the orderings of adsorption energies are correct for the two adsorbates. The observed overbinding for phenol relative to experiment is also observed in the ΔH^{solv}_{ads} of the bond additivity model (approximately -70 vs. -19 kJ mol^{-1} [229]).

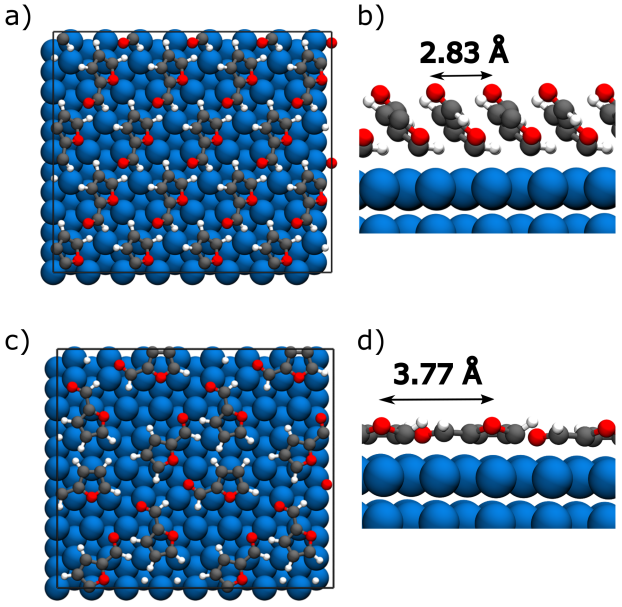

FIGURE 6.10: Geometries of furfural on the Pt(111) facet. a) and c) represent the bird's eye view of $\theta = 1.0$ vertical orientation and $\theta = 0.75$ horizontal orientation coverages respectively. b) and c) show the same geometries viewed from the side.

However, we have shown that, according to the Temkin model presented, increasing coverages do not rationalise the overbinding observed in computationally derived $\Delta G^{solv}_{ads,\theta}$ compared to experiment for the horizontal orientation. We will explore the effects of orientation in the following section as another source of error.

However, the implicit solvent correctly captures the weaker correlation of adsorption free energies with respect to θ (b = 35 in solvent, b = 72 in vacuum). This is reflected in the experiment where $b = 5$ [45] and $b = 72$ [58] for solvent and vacuum respectively. In implicit solvent, this is rationalised by screening for repulsive interactions between the adsorbates [233]. However, the value of b obtained with the implicit solvent method is significantly larger than experimental values for phenol. For furfural, similar to phenol, we observe a significantly weaker correlation of $\Delta G^{diff,solv}_{ads,\theta}$ in solvent ($b = -17$ and $b = -5$ for the lower bound and upper bound respectively) relative to vacuum ($b = 36$) caused by dielectric screening.

6.4 Vertical Orientations

Implicit solvent models parameterized to obtain ΔG_{solv} for the Pt(111) surface yields experimentally observed decrease of ΔG^{solv}_{ads} relative to ΔG^{vac}_{ads}. However, the continuum solvation model also yields large errors with respect to experiment. The implicit solvent Temkin isotherm plots demonstrate coverage effects alone do not rationalise the high ΔG^{solv}_{ads} compared to experiment. It is known the orientation of adsorbates are

sensitive to a variety of conditions, such as adsorbate concentration [52], voltage [234] and the species of electrolyte [53]. Therefore, we will also explore orientation effects (i.e., vertical and horizontal geometries) which may correctly align experimental and implicit solvent ΔG_{ads}^{solv} values.

$\Delta G_{ads,\theta}^{int,vac}$ for cyclohexanol is measured in the vertical orientation across a range of coverages. $\Delta G_{ads,\theta}^{int,vac}$ for cyclohexanol is significantly smaller than the aqueous free energy of adsorption for phenol and furfural adsorbed horizontally. Especially in the lower bound entropy, adsorption in solvent is predicted to be endothermic. These effects result from the relatively weak binding of cyclohexanol to Pt (which is observed in vacuum), combined with the energetically unfavourable desolvation of the Pt surface and -OH fragments. These observations are consistent with experimental findings, where cyclohexanol in the measured concentration range ($10^{-6} - 10^{-3}$ M) fails to achieve saturation coverage, in contrast to other measured organic molecules [45]. The higher value of $\Delta G_{ads,\theta=1}^{int,solv}$ suggests that cyclohexanol is stable on the Pt surface at high local coverages, as is theorised for phenol adsorption in the aqueous phase [48]. This is rationalised by the apparent attractive interactions between the cyclohexanol fragments and the exclusion of the surface area of the dielectric cavity, which models the favourable desolvation of the hydrophobic alkane fragment.

The negative value of $b = -18$ for the adsorption of cyclohexanol in solvent compared to vacuum $b = -14$ is driven by the exclusion of the cavity surface area of the adsorbate at higher coverages. In the vertical binding mode, desolvation of the fragment exposed to the dielectric only occurs when the adsorbates are within proximity of each other. At small distances, their dielectric cavities merge and the surface accessible surface area decreases, meaning $\Delta G_{ads,\theta}^{int,solv}$ becomes more exothermic.

We note that in the limit of high local coverages, the Temkin isotherm model measured across the range of coverages shown here becomes a less useful comparison to experiment. This is because, despite increasing the number of sites occupied with higher concentrations in the experiment, the measured adsorption value of $\Delta G_{ads,\theta}^{int,solv}$ reflects the higher coverage values. As a result, the energetics of adsorption lead to $\Delta G_{ads,\theta}^{diff,solv} \approx \Delta G_{ads,\theta}^{int,solv}$. Following this rationale, we acquire semiquantitative agreement with the experiment, where the calculated value of $\Delta G_{ads,\theta}^{int,solv} = -14$ kJ mol^{-1} compares favourably with the experimental value of $\Delta G_{ads,\theta}^{diff,solv} - 17$ kJ mol^{-1} for cyclohexanol.

At saturation coverages of furfural (Table 6.3), $\Delta G_{ads,\theta=1}^{int,solv} = -38$ kJ mol^{-1} compared favourably to experimental measurements of $\Delta G_{ads,\theta=1}^{int,solv} = -17$ kJ mol^{-1}. Relative to the horizontal orientation, $\Delta G_{ads,\theta=1}^{int,solv}$ is 71 kJ mol^{-1} weaker. Overall, the better agreement with the experiment is calculated using vertical geometries of cyclohexanol and furfural. However, this contradicts the predictions of the entropy model of Section 6.3.1,

TABLE 6.3: Impact of orientation on values of $\Delta G_{ads}^{int,solv}$ to Pt(111) at saturation coverage compared to experiment. Implicit solvent values shown for the lower bound entropy estimate.

Molecule	Vertical $\Delta G_{ads,\theta=1}^{int,solv}$ kJ mol^{-1}	Horizontal $\Delta G_{ads,\theta=1}^{int,solv}$ kJ mol^{-1}	Expt. $\Delta G_{ads,\theta=1}^{int,solv}$ kJ mol^{-1}
Furfural	−38	−107	−26 [45]
Cyclohexanol	−14	-	−17 [45]
Phenol	-	−102	−9 [45]

where for aromatic compounds, the experimental entropies of adsorption lie approximately between the upper and lower bound of the horizontal orientation. In contrast, the upper bound of ΔS_{ads}^{solv} calculated using the vertical geometry underestimates ΔS_{ads}^{solv} by approximately 50 kJ mol^{-1}.

Further studies using the Bond Additivity model measure similar overbinding of horizontal phenol to the plane of Pt(111) ($\Delta H_{ads}^{int,solv}$=−72 kJ mol^{-1} [229] vs. +38 kJ mol^{-1} [49]). In contrast, as discussed in Section 6.3.2, QM/MM studies achieve better agreement with experimental values of $\Delta G_{ads}^{int,solv}$ [10]. However, the weaker $\Delta G_{ads}^{int,solv}$ is likely a consequence of less favourable desolvation of the metallic surface, driven by larger values of ΔG_{solv} for Pt(111) relative to experimental values. Therefore, we propose three potential sources of error associated with the ΔG_{ads}^{solv} calculations:

1. Inaccurate estimates of ΔG_{solv} for the Pt(111) surface.
2. Improperly accounting for the orientation of the target adsorbate.
3. Not accounting for electrochemical effects induced in experiment.

For the first point, we emphasise the importance of accurate measurements of ΔG_{solv} for the metallic surface. Alternatively, we must consider that computational simulations and bond additivity approaches take place in the absence of applied potential and electrolyte. As a result, the binding strength of aromatics could be weakened by effects that are not reflected in implicit solvent approaches, AIMD simulations [3] or bond additivity approaches [229] for uncharged systems. Simulations that include potential and electrolyte effects are possible within the implicit solvent model of ONETEP [127, 128], but are beyond the scope of this study. Owing to the discrepancies observed in our calculations and in other bond additivity approaches across a range of coverages and orientations, a wider breadth of studies are required to fully understand the thermodynamics of adsorption in the electrochemical environment. However, the method presented shows that implicit solvents can obtain the correct qualitative trends of adsorption in the aqueous phase.

6.5 Conclusions

In this work, we have shown that inexpensive implicit solvent models can reproduce experimental adsorption free energies for small organic molecules in solvent. Parameterization requires separate dielectric cavity definitions for the metallic surface, achieved through the soft sphere model [123]. In contrast, isocontour charge density based cavity models are unsuitable for this purpose, where calculations performed in VASPsol drastically underestimate the free energy of adsorption of organics to Pt(111) in the aqueous phase. Furthermore, we have shown that in the absence of direct experimental measurements of ΔG_{solv} for the surface, approximations of ΔH_{solv} and ΔS_{solv} yield semiquantitative results for the overall ΔG_{solv}. We find that simple approximations based on the ice-like bilayer model [49] or the accentric cavity model [212] provide a reasonable upper and bound for ΔS_{solv}. However, a greater body of experimental data is required to validate which bound better describes the entropy of solvation for the surface. In this case, the difference between the upper and lower bound in the calculation of ΔG_{ads}^{solv} is small ($\sim$ 10 kJ mol^{-1}), but this error necessarily increases with temperature and the surface area of the adsorbate. In the absence of experimental data, we predict ΔH_{solv} can be reasonably calculated through *abinitio* values of ΔH_{ads} of water to the chosen interface and the water surface energy [48, 67]. If this is not available, we have shown that the parameterizing against the change of work function in solvent may act as a reasonable guess for ΔG_{solv} [157]. We predict that, provided the above quantities are known, these methods can be extended to other metallic surfaces.

In addition, we have emphasised the importance of correctly accounting for coverage effects when comparing to experimental data. We also present the utility of the entropy approximation and overall values of ΔG_{ads}^{solv} in predicting structural information for the Pt(111)/Adsorbate interface in water. However, we note that a lack of atomistic understanding of adsorption in the aqueous phase makes a facile comparison with experiment difficult. Commonly in the literature, it is assumed that oxygenated aromatics adsorb horizontally to the Pt(111) surface in the aqueous phase much like in the vacuum phase. However, we observe overbinding of approximately 100kJ mol^{-1} relative to experimental values of ΔG_{ads}^{solv}. We propose that significant weakening in the strength of binding in experiment for aromatic compounds arise through either a change from a horizontal to a vertical orientation, or applied potential/electrolyte effects not included in bond additivity based approaches. Our findings also support the presence of high local coverages of adsorbates at the aqueous Pt(111) interface [48], as shown by the relatively high values of ΔG_{ads}^{solv} predicted at saturation coverage for cyclohexanol. In an area where experimental data is limited and difficult to measure, the proposed techniques provide an inexpensive toolset for the validation and prediction of ΔG_{ads}^{solv} and ΔS_{ads}^{solv} on solvated metallic surfaces. However, more statistically significant validation requires more empirical measurements for the aqueous entropy/free energy of adsorption.

Chapter 7

Conclusions

7.1 Conclusions and Thesis Summary

The work in this thesis has focused primarily on the application of the implicit solvent model to platinum surfaces, allowing us to obtain electrochemical properties and model adsorption processes in the aqueous phase. In support of these studies, we have also described the improvements made to the ONETEP code, where we have accelerated the convergence of single-point energy calculations for metallic systems and increased the flexibility of the dielectric cavity in the implicit solvent framework through the soft sphere cavity model. Where possible, we have made detailed comparisons of our calculated values to experiment, rooting our discussion to the frontiers of applied heterogeneous catalysis.

We presented our implementation of the soft sphere cavity model, in conjunction with a parameterization of the non-polar solvation free energy change in terms of the solute surface area and volume (SASA+SAV). Through comparisons with a large assay of ΔG_{solv} values obtained with through experiment and the original charge based cavity model, we find that the soft sphere model provides accurate free energies of solvation. Overall, we obtain a MAE $\sim$ 4.0 kJ mol^{-1} with respect to the experimental test set, which is comparable to the original implementation of the soft sphere solvent model in BigDFT (4.68 kJ mol^{-1}) [123]. Furthermore, we find that the fully parameterized SASA+SAV model leads to a modest improvement in the accuracy of 2 kJ mol^{-1} in the MAE compared to the original SASA model with a scaling factor designed to account for the dispersion-repulsion for a set of long-chain alkanes. These implementations were critical in providing a flexibly parameterizable solvent model, which can accurately account for the different optimal cavity radii required to calculate accurate thermodynamic properties for heavy metallic species and light organic molecules.

Furthermore, we have validated the analytical forces obtained with the electrolyte model of ONETEP by comparison with exact forces obtained through finite differences. We

found that overall, there are negligible force errors associated with the implicit representation of the electrolyte for a range of charged and uncharged systems. These findings were strong even at very high electrolyte concentrations. We conclude that no additional analytical force terms are required for geometry relaxation calculations using the electrolyte model. However, we found that the fixed cavity approach did lead to significant force errors, due to the lack of convergence of the total energy with respect to the dielectric cavity function. Fortunately, this error can be eliminated by updating the dielectric cavity self-consistently, thus reducing the need for additional analytical force terms to describe the implicit dependence of the cavity on changes in atomic positions. By ensuring that a method exists to accurately calculate forces, this validation will enable future studies of charged surfaces in the electrochemical environment with the electrolyte model.

Furthermore, we gave an account of our implementation of Pulay Hamiltonian mixing for ONETEP in order to accelerate single-point energy calculations for large metallic systems. Grounding our discussion in the language of numerical methods, we find that the Helmholtz free energy converges significantly more quickly in the Pulay scheme than the original damped fixed point approach. With the linear order of convergence of the latter scheme, a large number of iterations are required to attain self-consistent energies. In contrast, the Pulay mixing scheme generally performs much larger steps towards converged energy, leading to convergence of the inner loop to within 10-25 inner loop iterations for metallic systems. Furthermore, by defining the step size λ as a fixed, user-specified constant instead of searching an optimal λ value through the original line search routine, we reduce the number of computationally intensive energy evaluations to one per each inner loop step. For large Pt(111) metallic slabs of 450 atoms, we observed up to three times speed-ups for the combined fixed λ and Pulay mixing compared to the original implementation of line search optimised λ with linear mixing. Overall, the greater efficiency conferred by the Pulay Mixing method expands the scope of studies concerning metallic systems in ONETEP, both increasing the number and the size of systems that can be evaluated.

Using the implementation of the soft sphere cavity model, we demonstrated that the implicit solvent approach can obtain the work function changes induced by water for the Pt(111) surface. By parameterizing the dielectric cavity to match the induced dipole changes of the metallic surface obtained from AIMD studies, we were able to replicate the experimental potential of zero charge offset by the intrinsic dipole of water at the metal/liquid interface. Closer analysis of the electronic structure reveals that the implicit solvent approach replicates the charge displacement towards the metallic surface observed in the explicit solvent approach but does not necessarily describe more complex features of the electron density induced by Pt-OH_2 bonding. We also discussed the role of nonlocal dispersion including functionals in calculating the work function for metallic surfaces, and find that the improved description of the electron density with

VV10 in the diffuse electron density region leads to better agreement with experimental values. We then evaluate the ability of the work function parameterization to obtain adsorption enthalpies for phenol in the aqueous phase. Although this method obtains a qualitative decrease of ΔG_{ads}^{solv} compared to ΔG_{ads}^{vac} observed in experiment [45, 47], the work function based parameterization overestimates the ΔH_{solv} for the Pt(111) facet, leading to overestimated changes of ΔH_{ads}^{solv} with respect to ΔH_{ads}^{vac}.

We then reconcile our implicit solvent approach with experimental studies of adsorption processes by performing an alternative parameterization of the dielectric cavity in terms of the ΔG_{solv} of the Pt(111) surface. To support this work, we provide a simple method for evaluating the entropy of adsorption in the aqueous phase by deriving a lower bound to the ΔS_{solv} of the Pt(111) surface based on the cavitation entropy of solvation [212], paired with the upper bound based on the assumption that H_2O takes on the entropy of the solid phase upon adsorption [49]. Although the existing body of experimental work best matches the upper bound, questions surround the accuracy of much of this experimental data because of the effects of strong-binding electrolytes and the errors of the Van't Hoff plots resulting from the small range of equilibrium constants over the temperature range measured [49]. Evaluating the ΔG_{ads}^{solv} of the implicit solvent model for a range of adsorbates, coverages, and orientations, we find that the commonly assumed structure of aromatic adsorbates (parallel to the metallic surface) leads to large overestimations of the free energy of aqueous adsorption (by approximately 50 kJ mol^{-1}). This finding is supported by AIMD studies, where the ΔG_{ads}^{solv} for phenol is similarly more exothermic relative to experiment (-151 vs. -21 kJ mol^{-1} [3, 45]). In contrast, for furfural and cyclohexanol, the vertical orientations with respect to the metallic surface yielded values of ΔG_{ads}^{solv} within 20 kJ mol^{-1} of experiment at high coverages. Taking into account the various assumptions used in the construction of this model, these values are well within the error tolerances. Furthermore, we show that this model qualitatively reproduces the experimental findings of ΔG_{ads}^{solv} for the dissociative adsorption of H_2 on the Pt(111) surface in the aqueous phase. In our model, the energetically disfavoured displacement of the dielectric by H_2 approximately captures the energetic penalty of breaking the Pt-OH_2 bonding interaction. Combining these techniques provides a useful set of diagnostic tools for materials discovery and experimental work in heterogeneous catalysis studies, allowing access to structural information at relatively low computational costs compared to explicit solvent approaches. We further predict that this framework is readily extendable to larger ranges of adsorbates and that, provided that approximations of ΔH_{ads} of H_2O are available, it can be applied to other metallic surfaces.

Overall, our work demonstrates the great utility of the implicit solvent model for both electrochemical and heterogeneous catalytic applications. As long as accurate parameterizations are performed for the desired property, the continuum solvent approach

can provide a range of properties pertaining to the metallic/liquid interface and adsorption processes. We have furthermore implemented improvements to the existing ONETEP code for both the solvent model and ensemble DFT, enabling more extensive studies through the reduced computation costs introduced by Pulay mixing EDFT. These developments have allowed the study of the implicit solvent model applied to a range of adsorbates at the Pt(111) surface across a range of coverages. We show that in conjunction with easily calculated statistical thermodynamic descriptions of the entropy changes, the adsorption free energy at metallic surfaces in the aqueous phase can be obtained with reasonable accuracy, while avoiding arduous dynamical simulations associated with explicit water simulations. These studies allow access to structural information for the adsorption process, suggesting the preferred vertical orientation of organic molecules at the Pt(111) surface in the aqueous phase. We have therefore shown that the multifaceted effects of water at the metallic interface can be represented with a well-parameterized, computationally inexpensive continuum solvation model. In the evolving field of heterogeneous catalysis, the readily extendable methods described in this work will provide future insights into the processes occurring at the aqueous metallic interface in addition to those presented. In the future, the insights provided by this work has potential to guide the rational design of heterogeneous catalysts.

7.2 Future Work

Although the experimental work precisely describing the thermodynamics of adsorption in the aqueous phase is quickly evolving, the difficulty in obtaining structural information of adsorbates at the metallic interface in the aqueous phase obfuscates the precise binding mechanisms between the surface and adsorbate. We have based comparisons of our models on existing experimental data, but acknowledge there are a range of effects such as voltage and electrolyte concentration which may influence the structure and binding energies in experiment. More complete validation of the methods described in this work can be made when the body of experimental data grows, providing deeper insight into the predictions made by our model of adsorption in the aqueous phase.

Furthermore, the work presented can act as a framework for simulations of charged surfaces. With recent developments of the Poisson Boltzmann model in ONETEP [127] and the implementation of grand canonical ensemble DFT [128], studies of adsorption processes for charged interfaces in the presence of electrolyte can be performed. Such studies will better match the electrochemical conditions used in many experimental measurements [45] and the expanding field of elecrocatalysis, which allows for the conversion of biomass derivatives at low temperatures [235]. Further developments made with Electronic Decomposition Analysis (EDA) for metallic surfaces can furthermore provide qualitative insight to the various bonding interactions between the adsorbates

and the metallic surface for a range of coverages. Such a technique could elucidate the specific role of charge transfer, adsorbate-adsorbate interactions and dipole changes in observed changes in ΔH_{ads} experienced at different adsorbate concentrations [236].

However, the model as presented can be used to calculate the energetics of adsorption for a wider range of adsorbates. Furthermore, by calculating the adsorption free energies of water on other noble metal surfaces, the ΔG_{solv} of alternative catalytic systems can be measured. This increased scope will indicate the comparative adsorption free energies of different species, aiding the mechanistic understanding of heterogeneous catalytic processes. Further in the future, these techniques could be applied to nanoparticles and their supports. This will give a more realistic model of the catalytic environment by including the influence of supports on the binding energies of adsorbates [75, 182]. Our model will further improve this representation by accounting for the energetics of solvation.

Appendix A

Entropy Derivations

A.1 Entropy of Isolated Molecule

The overall entropy of an ideal gas is calculated as.

$$S^{vac,3D} = S_t^{vac,3D} + S_r^{vac,3D}. \tag{A.1}$$

S_t^{vac} is calculated in terms of the translational partition function,

$$S_t^{vac,3D} = k \ln (q_t^{vac,3D}) + k + kT \left(\frac{\partial \ln (q_t^{vac,3D})}{\partial T} \right)_V, \tag{A.2}$$

where k is the Boltzmann constant and q_t is calculated through the simple approximation for a molecule of mass m in a box of volume V:

$$q_t^{vac,3D} = \left(\frac{2\pi mkT}{h^2} \right)^{3/2} V. \tag{A.3}$$

S_r^{vac} is calculated through the rotational partition function:

$$S_r^{vac,3D} = k \ln (q_r^{vac,3D}) + kT \left(\frac{\partial \ln (q_r^{vac,3D})}{\partial T} \right)_V, \tag{A.4}$$

where $q_r^{vac,3D}$ is the rotational partition function for a non-linear molecule using the rigid rotor approximation:

$$q_r^{vac,3D} = \frac{\pi^{1/2}}{\sigma_r} \left(\frac{8\pi^2 IkT}{h^2} \right)^{3/2}. \tag{A.5}$$

where σ_r is the rotational symmetry factor of the molecule and the inertia I is calculated as the product of the diagonal elements of moment the intertia tensor $I = (I_{xx} I_{yy} I_{zz})^{1/3}$.

A.2 Entropy of Adsorption

The standard entropy of an adsorbate is commonly described by the 2D lattice gas or 2D ideal gas approximations. The latter applies for molecules with sufficient thermal energy to diffuse freely across the metallic surface (i.e., high temperatures), in addition to having weak adsorbate-adsorbate interactions [12]. The lattice gas approximation is used where the molecules are restricted to their respective sites on the metallic surface [68].

In the ideal 2D gas approximation, we assume the molecule loses the translational degree of freedom orthogonal to the metallic surface upon adsorption, where it is free to diffuse across area, A:

$$S_t^{vac,2D} = k \ln (q_t) + k + kT \left(\frac{\partial \ln (q_t^{vac,2D})}{\partial T} \right)_A, \tag{A.6}$$

where q_t is modified to:

$$q_t^{vac,2D} = \left(\frac{2\pi mkT}{h^2} \right) A. \tag{A.7}$$

The modification of $S_r^{vac,2D}$ depends on the nature of the adsorbate and its environment [12] [73]. *Campbell* approximates that approximately 1/3 of the rotational entropy is lost overall for smaller organic molecules. As this reduction in entropy does not simply correspond to a specific principal axis of rotation, we shall simply adopt:

$$S_r^{vac,2D} = 2/3 * S_r^{vac}. \tag{A.8}$$

However, we will explore possible cases where further degrees of rotational freedom are removed.

A.3 Entropy of Solvation

Garza [212] showed that the entropy of a solvated molecule ($S^{solv,3D}$) can be calculated with only geometric shape parameters (cavity surface area, cavity volumes and van der Waals' radii) and simple physical constants. Similarly to the gas phase, $S^{solv,3D}$ is partitioned into four components:

$$S^{solv,3D} = S_v^{solv,3D} + S_t^{solv,3D} + S_r^{solv,3D} + S_c^{solv,3D}, \tag{A.9}$$

where $S_v^{solv,3D}$ is taken as 0, and $S_c^{solv,3D}$ is the entropy of formation for the solvent cavity.

$S_t^{solv,3D}$ is calculated using the same Sakur-Tetrode expression (Equation A.2) and partition functions (Equation A.3 as the 3D gas phase. However, the volume the molecule

can move in is restricted to a cavity of V_c by the surrounding aqueous medium. V_c is defined as a sum of the van der Waals' volume of the molecule (V_M) and the volume of free space per solvent molecule (V_{free}).

$$V_c = (V_M^{1/3} + V_{free}^{1/3})^3, \tag{A.10}$$

where V_{free} is defined as,

$$V_{free} = \frac{M_w^S}{N_A \rho} - V_S, \tag{A.11}$$

where V_S is the van der Waals' volume of the solvent, ρ is the mass density of water, N_A is Avogadro's Number, and M_w^S is the molecular weight of water. Additional terms in the original work correct for the entropy contribution of the molecule 'hopping' from one cavity to another, which is expressed as a scaling factor applied to the volume. However, this value assumes unity for all molecules studied in this work and is excluded for brevity.

$S_r^{solv,3D}$ is calculated using the same rigid rotor, non-linear assumptions of the 3D gas phase. However, additional terms are used to co1rrect the translational entropy loss caused by confining the radius of gyration to V_{free}. In this model, the volume the rapidly rotating linear/spherically symmetric rotor can move within V_{eff} (defined as a sphere of radius r_{eff}) is reduced by a volume defined by its volume of gyration (r_g):

$$r_g^2 = \frac{1}{N_{at}} = \sum_{k=1}^{N_{at}} (\mathbf{r}_k - \mathbf{r}_{mean})^2. \tag{A.12}$$

$S_r^{solv,3D}$ is then expressed as,

$$S_r^{solv,3D} = k \ln (q_r^{vac,3D}) + kT \left(\frac{\partial \ln (q_r^{vac,3D})}{\partial T} \right)_V + S_t(T, V_{r_c - r_g}) - S_t(T, V_{r_c}), \tag{A.13}$$

where $V_r = 4\pi r^3/3$ is the volume of a sphere with a given r. The rotational partition function is unmodified from the 3D gas phase case. Note that the partition function is unmodified from the 3D gas phase case (eq. A.5).

$S_c^{solv,3D}$ can be expressed using arguments based on either Scaled Particle Theory or an Acentric Factor Approximation. The former was found to be very sensitive to the choice of scaled solvent radii R_S, where small differences in experimental values of R_S resulted in significantly different values of $S_c^{solv,3D}$. Therefore, the acentric factor approximation is used in this work.

The acentric factor approximation calculates entropy of cavity formation by taking the difference between the entropy of vaporization between a simple liquid (spherical and nonpolar) and the real liquid [227]. The deviation of the real from the simple liquid (i.e., deviations from the Principle of Corresponding States) is represented by the Pitzer

accentric factor (ω), which is measured as 0.344 for water.

$$\Delta S_{vap}^{ideal} - \Delta S_{vap}^{real} = -\frac{\Delta H_{vap}^{ideal}}{T_c}\omega = -5.365\omega k, \tag{A.14}$$

where 5.365 is derived from the enthalpy of vaporization of an ideal liquid under the corresponding state theorem, $\Delta H_{vap}/T_c$.

ΔS_{vap}^{real} contains contributions from rotational degrees of freedom which are absent from the ideal liquid, which only represents changes in entropy arising from transitional degree of freedom. Therefore, the entropy change of solvation due to $S_r^{solv,3D}$ must be removed to recover the entropy of cavitation of a pure liquid,

$$S_c^{solv,3D,pure} = -5.365\omega k - S_t(T, V_{r_c - r_g}) + S_t(T, V_{r_c}), \tag{A.15}$$

where the latter two terms represent the change in rotational entropy of the solvent relative to the gas phase ($\Delta S_{r,S}^{solv}$).

For a mixture of compounds, a further factor is introduced to account for the greater or smaller number of coordinated solvent molecules due to the different shape and size of the solute. This is approximated by $\mathcal{G}(\mathrm{R}_M, \mathrm{R}_S)$:

$$S_c^{solv,3D,mix} = -(5.365\omega k + \Delta S_{r,S}^{solv})\mathcal{G}(\mathrm{R}_M, \mathrm{R}_S), \tag{A.16}$$

where $\mathcal{G}(\mathrm{R}_M, \mathrm{R}_S)$ approximates the coordination of solvent molecules around the solute, with $\mathcal{G}(\mathrm{R}_X, \mathrm{R}_X) = 1$. This is approximated by assuming packing occurs for a box shape, as opposed to a spherical set of cavities, i.e., $\mathcal{G}(\mathrm{R}_M, \mathrm{R}_S) = A_M/A_S$, where A_X is the solvent accessible surface area (SASA) of compound X. Differences in curvature are corrected by

$$\phi_X = A_X/A_X^{box} \tag{A.17}$$

where A_X^{box} is the surface area of the minimum bounding box of the SASA. This results in the simple approximation for packing spherical solvent molecules around the solute surface,

$$\mathcal{G}(\mathrm{R}_M, \mathrm{R}_S) = \frac{A_M\phi_S}{A_S\phi_M}. \tag{A.18}$$

A further correction to Equation A.16 is also required to account for the entropic penalty of creating a cavity able to fit the solute into the ideal liquid of hard spheres. The cavity entropy for the reference ideal liquid is then represented as,

$$S_c^0 = \frac{kV_M}{V_S} ln(1 - V_S n_S), \tag{A.19}$$

which yields the expression for the overall entropy of cavitation for the desired solute,

$$S_c^{solv,3D} = S_c^0 - (5.365\omega k + \Delta S_{r,S}^{solv})\mathcal{G}(\mathrm{R}_M, \mathrm{R}_S). \tag{A.20}$$

Appendix B

Multipole Corrections

B.1 Multipole Corrections

The plane wave approach inherently involves the use of periodic boundary conditions (PBC). In order to model isolated molecules, the supercell technique was developed, which involves adding large quantities of vacuum to the simulation cell in order to separate their periodic images. Although this method is adequate for neutral molecules, charged systems and systems with significant multipoles require additional considerations. Monopole and multipole moments decay in accordance to the power law, where point charges decay with $\frac{1}{r^{-1}}$, dipoles with $\frac{1}{r^{-2}}$ etc. Using the supercell method for systems with a net charge or significant dipoles require large volumes of vacuum to eliminate the electrostatic interactions between the system of the unite cell and its periodic images. However, as traditional plane wave codes extend across the entire cell, adding large quantities of vacuum is incredibly computationally costly. With respect to the work herein, the supercell technique suffices for the neutral metallic surface, whereas calculations performed for the surface/water interface involve significant dipoles depending on the net orientation of the water layer.

Various dipole corrections such as the Coulomb cut-off (CC) [237], Continuum Screening Method [238] and Gaussian Counter Charge model [239] have been developed in order to isolate the electrostatic interactions of the unit cell. In this work, we will limit our discussion to the Coulomb cut-off and its implementation in ONETEP.

The Hartree potential of a system is defined as:

$$V_H(\mathbf{r}) = \int_{space} d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \tag{B.1}$$

where $n(\mathbf{r}')$ describes the charge density of the system. Through convolution theory, applying a Fourier transformation to this term yields the reciprocal space representation V_H:

$$V_H(\mathbf{G}) = V(\mathbf{G})n(\mathbf{G}), \tag{B.2}$$

where $V(\mathbf{G}) = \frac{4\pi}{|\mathbf{G}|^2}$ represents the Fourier transform of the Coloumb interaction.

While the above expressions are taken over all space (∞ to $-\infty$), the Coulomb cut-off sets the Coulomb interaction between charges to zero beyond a specified radius. By selecting an appropriate cut-off radius, R_C, one can retain the correct electrostatic interactions of between charges in the unit cell, while eliminating spurious interactions between the periodic images. In the original formulation of CC, the truncation is performed in all directions as a sphere:

$$V^{3D}(\mathbf{r}, \mathbf{r}') = \begin{cases} \frac{1}{|\mathbf{r}-\mathbf{r}'|} & \text{for } R_C < |\mathbf{r}-\mathbf{r}'| \\ 0 & \text{for } R_C > |\mathbf{r}-\mathbf{r}'| \end{cases} \tag{B.3}$$

Performing an analytic Fourier Transformation of $V(\mathbf{r}, \mathbf{r}')$ with modified boundaries yields the following inverse space representation of the Coulomb interaction:

$$V^{3D}(\mathbf{G}) = \frac{1 - \cos(|\mathbf{G}| - R)}{|\mathbf{G}|^2} \tag{B.4}$$

Another important component of the Coulomb cut-off is the appropriate choice of cut-off radius and simulation cell length. These are selected to ensure two conditions are met: 1) The correct electrostatic interactions are represented in the cell and 2) the interaction between charges of the periodic image and the simulation cell are set to zero. The first condition is satisfied by setting the cut-off distance equal to or greater than the distance between any two non-zero charges. This is typically satisfied by $R_C > \sqrt{3}L_{cell}$. The second condition requires that the distance between non-zero charges of the simulation cell and the periodic image must be greater than or equal to the cut-off length. In ONETEP, this is achieved by placing the unit cell inside a larger padded cell, in which the charge density $\rho(\mathbf{r})$ is set to 0. The second condition is satisfied when the total cell length, $L_{total} = L_{cell} + L_{pad} \geq R_C + L_{cell}$.

Although this approach is satisfactory for systems with no periodicity, additional considerations must be made for the infinitely extended planes under study in this report. For such systems, the Coulomb cut-off must only be applied in the out-of-plane direction, while retaining PBC in the xy-plane. Originally, this was implemented in ONETEP

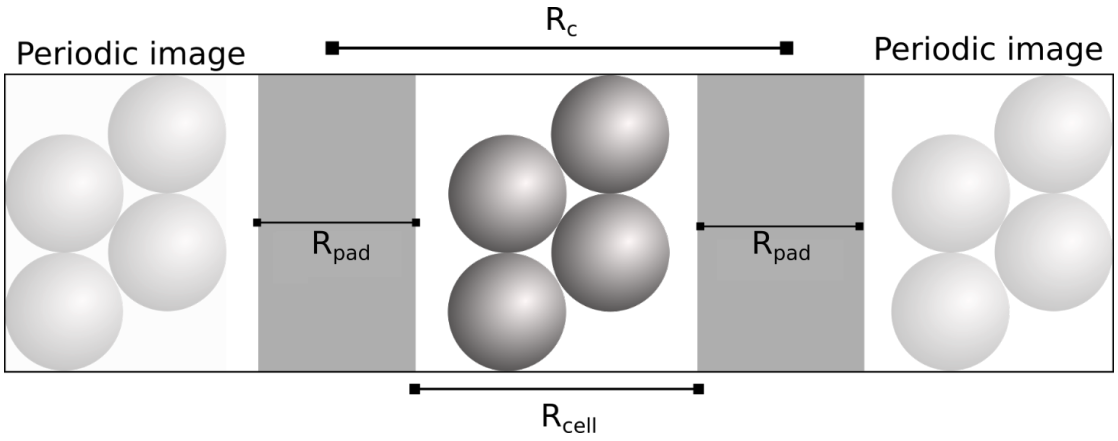

FIGURE B.1: Demonstrates the 2D-periodic Coulomb cut-off across the xz-plane. The configuration shown satisfies the conditions of the Coulomb cut-off, $L_{total} = L_{cell} + L_{pad} \geq R_C + L_{cell}$, required to isolate the system from its periodic images. In the padded regions, the electron density is set to 0.

from the formulation of *Rozzi et al.* [184], where the Coulomb interaction $V^{3D}(\mathbf{G})$ is recast to the following expression:

$$V^{2D}(\mathbf{G}_{\parallel}, G_z) = \frac{4\pi}{\mathbf{G}^2}\left[1 + e^{-G_{\parallel}R_C}\frac{G_z}{G_{\parallel}}\sin(G_z R_C) - e^{-G_{\parallel}R_C}\cos|G_z|R_C)\right] \tag{B.5}$$

Where $G_{\parallel} = \sqrt{G_x^2 + G_y^2}$ and G_z represent the in-plane and out-of-plane reciprocal space vectors respectively. This formulation requires additional expressions which handle the singularities at

1) $G_{\parallel} = 0$:

$$V^{2D}(\mathbf{G}_{\parallel}, G_z) = \frac{4\pi}{\mathbf{G_z}^2}[1 - \cos(\mathbf{G}_z R_C) - G_z R sin(G_z R_C)] \tag{B.6}$$

and: 2) $G_{\parallel} = G_z = 0$:

$$V^{2D}(\mathbf{G}_{\parallel}, G_z) = -2\pi R_C^2 \tag{B.7}$$

However, applying this implementation in our calculations resulted in unphysical electrostatic potentials. Instead of the expected smooth electrostatic potentials, our systems showed an oscillating electrostatic potential of 1 eV between every grid point in the system. These instabilities arise from an erroneous step-like function in the Coulomb interaction at $z = R_C$, resulting from equation B.6.

Therefore, as part of our work, we altered the originally implemented formulation to that proposed by *Tribaulet et al.* [185]. Their formulation requires $R_C = \frac{L_{total}}{2}$. This corrects the Coulomb cut-off by considering G_z as a multiple of $\frac{2\pi}{L_{total}}$, where $G_z \times R$ becomes a multiple of π. This eliminates the sine term, yielding a new form of the Coulomb interaction:

$$V^{2D}(\mathbf{G}_{\parallel}, G_z) = \frac{4\pi}{\mathbf{G_z}^2}[1 - e^{-G_{\parallel}R_C}\cos(\mathbf{G}_z R_C)] \tag{B.8}$$

This simplified expression eliminates the need to represent the singularity at $G_{\parallel} = 0$ with a separate equation.

A test to validate this implementation was performed for a small platinum supercell slab system, for a quantity which is sensitive to the value of the electrostatic potential (the work function). This is performed using the calculation parameters outlined in Chapter 5. For a (4x4) Pt system of 4 layers compared calculated with the Coulomb cut-off and standard periodic boundary conditions. We further test the implementation for a system with a significant dipole moment, i.e. a Pt(4x4x4) slab with a water bulk of 52 H_2O molecules, which we compare to results obtained with the Gaussian Counter Charge correction of Quantum Espresso [240], as implemented in the Environ module [111]. These calculations are performed using identical parameters (PAW psuedopotentials and PBE GGA functional) as their corresponding ONETEP calculations where applicable, with a 330 Ryd kinetic energy cut-off.

Method	Work Function / eV
QE - GCC	5.61
ONETEP - CC	5.66
ONETEP - Full PBC	5.48

TABLE B.1: Values of the work function for a Pt(4x4x4) cell in vacuum under different boundary conditions.

Method	Work Function / eV
QE - GCC	5.30
ONETEP - CC	5.25
ONETEP - Full PBC	-
ONETEP - Old CC	-

TABLE B.2: Values of the work function for a Pt(4x4x4)/52 H_2O interface with different methods of dipole interaction. Values cannot be obtained under PBC and the original implementation of the Coulomb cut-off.

The work function is taken as the $\psi = E_F - E_{vac}$, where E_{vac} represents the flat potential far in the vacuum region. First, the simple validation for the system in vacuum (Table

B.1) shows that the work function aligns well with the calculation in periodic boundary conditions and the GCC multipole correction.

Next, tests are conducted to ensure that the CC correctly replicates the electrostatic potential of an isolated system. This is performed using the average work function of 20 AIMD snapshots of Pt(4x4x4) slabs with a system of 52 water molecules (Equivalent to those used in Chapter 5). This ensures that a range of dipole moments are accounted for. Under standard periodic boundary conditions (Figure: B.1), the intrinsic dipoles of each periodic image spuriously interact with one another, meaning the electrostatic potential in the vacuum region shows an unphysical linear decay. Meanwhile, the electrostatic potential in the original implementation of CC has such large quantities of noise, no useful information can be extracted from the plot.

Contrastingly, the new implementation of CC gives the correct flat of electrostatic potential in the vacuum region. This suggests that the coulombic interactions between the periodic images have been correctly eliminated. Furthermore, the average work function of the CC agrees closely with Quantum Espresso's GCC, showing a small discrepancy of only 0.05 eV (Table B.2).

Appendix C

Parameterization of Vacuum Slab

C.1 Pt Vacuum Benchmarks

Ab initio methods provide a powerful computational tool to accurately simulate surface properties. Several benchmark studies [202, 241, 242] using the periodic supercell model have been conducted using a range of Density Functional Theory (DFT) methods. Such simulations are often done with the plane wave (PW) formalism for DFT calculations, which allows one to calculate charge density, energies, and band structures through integration across the Brillouin zone (BZ) [243]. Sampling techniques such as the Monkhorst-Pack scheme [87] allow for the efficient integration of the BZ via a summation of special **k**-points evenly distributed throughout the first Brillouin zone (1BZ) [87]. Typically, surface properties tend to be highly dependent on both the thickness of the slab and sampling of the 1BZ. It is therefore important to test for convergence by both sequentially increasing the number of layers of the surface slab and increasing the density of **k**-point sampling.

Using standard PW methods, *Singh-Miller et al.* [202] and *Da Silva et al.* [242] demonstrated that convergence of both the interlayer relaxation and work function occurs with 6 layers for the Pt(111) surface to within 0.02 Å and 0.01 eV respectively. However, each study emphasizes care when calculating the surface energy. As these calculations use the bulk properties as a reference, one can introduce errors through inconsistencies in 1BZ sampling between the bulk and surface calculations [244]. However, one can drastically reduce errors by using i) a sufficiently dense **k**-point mesh (typically around 30x30x30 or more) across both calculations or ii) the method outlined by Fiorentini and Methfessel [244]; where the reference bulk energy is extracted from a set of successively thicker surfaces as opposed to a single bulk calculation.

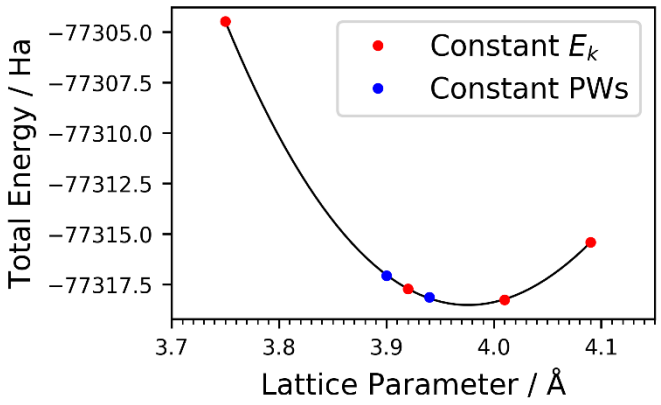

FIGURE C.1: The Birch-Murnaghan fitting procedure is carried out for an 864 Pt atom simulation cell. The quality of the basis is kept constant for each of the blue markers and the number of *psinc* basis functions vary. Red markers show the psinc kinetic energy cut off E_k is varied by approximately 10 eV to the nearest adjacent while the number of psinc basis functions remains constant.

One can also achieve adequate reciprocal space sampling by increasing the size of the simulation cell, which correspondingly reduces the volume of the 1BZ. For a sufficiently large simulation cell, the variation across the 1BZ decreases such that only one **k**-point (known as the Γ-point at **k**=0) is required to achieve convergence with respect to reciprocal space sampling. However, as the plane wave basis extends across the entire simulation cell, the memory requirement of the calculation increases dramatically with simulation cell size. This is especially punishing in slab supercell simulations, where large volumes of vacuum must be included in the system to avoid self-interaction between adjacent periodic images. *Santarossa et al.* [241] demonstrated the equivalence between the Γ-point and the Monkhorst-Pack sampling scheme for Pt(111) and Pt(110). Their study shows close agreement with the dense **k**-point approaches with (1x1) units cells, provided the bulk and surfaces are constructed with upwards of (6x6) repeating units of Pt. The memory cost of vacuum was made more manageable using the Gaussian Plane Wave (GPW) [188] basis approach as implemented in the CP2K code, which represents the charge density of Gaussian basis functions through an auxiliary plane wave basis set. The DFT formalism implemented in the linear-scaling DFT code, ONETEP [91] presents another approach for reducing the vacuum cost. By representing the plane waves as a set of highly localized periodic sinc (psinc) basis functions which are used to expand atom-centered NGWFs strictly localized within spherical regions, unnecessary computations are avoided in vacuum regions of the simulation cell.

TABLE C.1: Bulk properties obtained with different DFT methods, all using the PBE functional. PW approaches shown use a (1×1) unit cell of the fcc bulk with (21×21×21) **k**-point sampling. Both ONETEP and the GPW calculations were performed with **Γ**-point only sampling, with a (6×6×6) unit cell.

Calculation Method	Bulk Modulus / GPa	Lattice Constant / Å
ONETEP (6×6×6 unit cell)	268	3.98
Plane Wave (1)[a]	246	3.99
GPW (6×6×6 unit cell)[b]	270	3.97
Plane Wave (2)[c]	241	3.97
Experimental Value[d]	278	3.92

[a] *Singh-Miller et al.* [202]. [b] *Santarossa et al.* [241]. [c] *Da Silva et al.* [242]. [d] *Kittel et al.* [154].

C.2 Pt Bulk Properties

The bulk is constructed as a set of (6×6×6) FCC unit cells containing 864 Pt atoms, where the simulation cell volume is varied through a lattice constant of 3.75 to 4.09 Å. The Birch-Murnaghan equation of state9 is then used to obtain the equilibrium lattice constant and bulk modulus. Variations in cell size require careful treatment to keep the underlying basis set description as invariant as possible. One can either i) fix the basis quality by keeping the kinetic energy cut-off E_k constant, but increasing the number of plane waves, as outlined by *Skylaris et al.* [245], or ii) fix the number of basis functions, but varying the value of E_k. Here we combine both approaches as shown in Figure C.1.

In ONETEP, E_k is inversely proportional to the square of the grid spacing of the underlying psinc functions. In order to keep the value of E_k constant, the psinc spacing must also remain unchanged. This is done by varying the simulation cell size in discrete multiples of the underlying grid spacing (0.258 Å). An exception to this is the two points highlighted in blue in Figure C.1. Here, instead of keeping the psinc spacing constant and varying the number of grid points, we keep the number of grid points constant and vary the E_k by expanding the cell by a small amount (0.14 $Å^3$), with the number of plane waves kept constant. This only leads to an increase in psinc spacing of 0.0015 Å (or E_k ∼10 eV). As we can see from Figure C.1, the method of varying E_k produces far smaller fluctuations in total energy compared to the changes due to the lattice parameter, so we will take also these points into account in our calculation of the equilibrium lattice constant. The points shown in Figure C.1 correspond to lattice constants of 3.75 Å (with 87 grid points in each dimension), 3.90 Å (91 grid points with increased spacing), 3.92 Å (91 grid points), 3.94 Å (91 grid points with decreased spacing), 4.01 Å (93 grid points) and 4.09 Å (95 grid points). These points were selected as they are multiples of prime numbers and allow efficient FFTs for the calculation of the Hartree potential. The points with 89 grid points (3.83 Å) were not possible to use as they lead to very slow calculations as they are not divisible to smaller numbers of points for efficient FFTs.

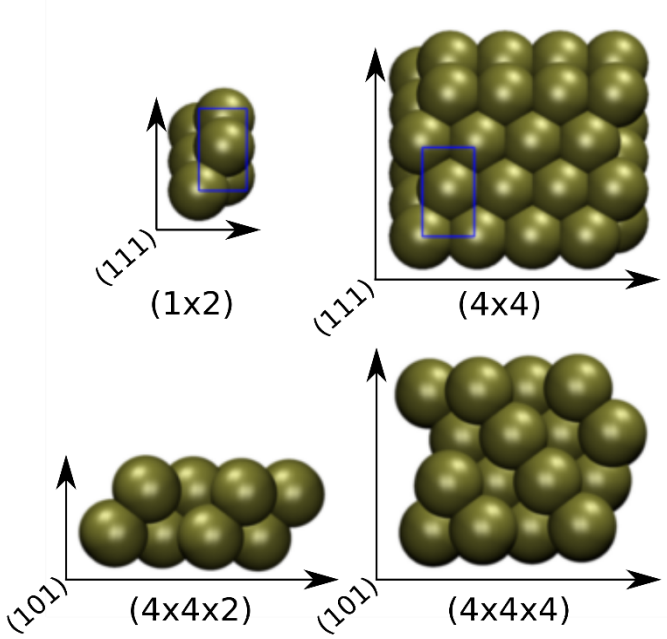

FIGURE C.2: Illustrates the (4×4) supercell of the Pt(111) facet with its underlying (1×2) unit cell viewed through the z-axis (top panel, (111) facet) and its corresponding view through the y-axis (bottom panel, (101) facet) showing two cells of N=2 and N=4 thickness.

Table C.1 shows the values of the lattice constant obtained via the Birch-Murnanghan equation from our points in Figure C.1 compared to PBE calculations with other methods such as the GPW **Γ**-point only approach used by *Santarossa et al.* [241] and plane wave approaches with k-point sampling. There is close agreement in the lattice constants between ONETEP (3.98 Å) and the other methods which produce values in the range 3.97-3.99 Å. We also quote the experimental value which is 3.92 Å, where it is well known that PBE overestimates this lattice constant [246]. The values of the bulk modulus show a wider spread in the range 246-270 GPa with the PW results in the range 241-246 GPa and ONETEP 268 GPa being very close to the value of GPW which is 270 GPa and close to the experimental value of 278 GPa. The much closer agreement with experiment of the GPW and ONETEP calculations may be an artefact of the use of localized orbitals in these approaches.

C.3 Surface Calculations in Vacuum

Surface calculations were performed with the periodic supercell method with a slab model. This approach emulates an infinitely extended metallic surface of a specified thickness through a set of repeated unit cells. To avoid self-interaction, the periodic images were separated by at least 20 Å of vacuum in the z-direction. Surfaces were constructed from sets of (1×2), orthorhombic unit cells for Pt(111) (as shown in Figure C.2). In order to test convergence with respect to **Γ**-point sampling, calculations were

TABLE C.2: Energetic properties for the Pt(7×6×N) set of slabs, including the work function and the surface energy, as calculated with the Fiorrentini method [244].

Species	Vacuum ϕ / eV	Surface Energy, σ / eV Å^{-2}
Pt(7×6×2)	5.72	0.69
Pt(7×6×3)	5.82	0.68
Pt(7×6×4)	5.70	0.66
Pt(7×6×5)	5.71	0.68
Pt(7×6×6)	5.64	0.68
Pt(7×6×7)	5.65	0.69
Pt(7×6×8)	5.64	0.68
PW[a]	5.69	0.65
Experiment	6.10 ± 0.06[b]	0.96[c]

[a]PW with 13-layer slab [202]. [b]UHV LEED [200]. [c]*Kumikov et al.* [247].

performed with a 5-layer slab with increasing numbers of unit cells in the x and y directions (ie. (4×4), (5×6), (6×6), (7×6) and (8×8). Energetic and structural properties were also evaluated with respect to the number of layers in the slab in order to validate our structures for use in solvation calculations. This was performed with the (7×6) surface slab.

C.3.1 Work Function and Surface Energy

The work function (ϕ) represents the minimum energy required to remove an electron from the metallic surface to a location in the vacuum where the variation of the electrostatic (Hartree) potential is no longer influenced by the charge density of the metallic species. This work will follow the convention where ϕ is calculated as the difference between the Fermi level E_F and the plane-averaged electrostatic potential V at a midpoint in the vacuum region between the Pt slab and its adjacent periodic:

$$\phi = E_{vac} - E_F, \tag{C.1}$$

where E_{vac} is calculated as an average of the electrostatic potential in the xy-plane 10 Å above the surface.

For slabs with N=2 to N=5, the value of the work function fluctuates between 5.82 to 5.70 eV (Table C.2). This implies finite size effects introduce significant errors to the work function for N<6. However, the work function values for N>6 converge within ±0.1 eV of the thickest surface slab. Furthermore, our converged value of 5.64 ± 0.1 eV compares favourably to the value of 5.61 ± 0.1 eV for the GAPW approach2, and the 5.69 ± 0.05 eV value of the PW approach with multiple *k*-points [202].

The surface energy of a metallic facet is defined as the free energy per unit area relative to the bulk for a particular crystal facet. In essence, this quantity describes the free

energy of formation of a surface compared to the bulk crystal. The simplest way to calculate surface energy, σ, is through the expression:

$$\sigma = \frac{1}{2\,A}\left(E^{N}_{surf} - NE_{bulk}\right) \tag{C.2}$$

where E^{N}_{surf} represents the total energy of the metallic surface of N Pt atoms, and E_{bulk} represents the energy of a single Pt atom within the bulk. Due to the symmetry of the slab, the surface energy per unit area can be obtained by a division of the total slab surface area, 2A. As mentioned in the introduction, the accuracy of σ is limited by the correspondence of **k**-point sampling of the surface and bulk calculations. The method of *Fiorrenttini et al.* [244] circumvents this issue by obtaining E_{bulk} from as a linear regression of E^{N}_{surf} for a set of surface slabs of increasing numbers of layers N_{layer}, but with a constant number of constituent unit cells and **k**-points.

Applying this method to the Pt(7×6×N) surfaces, σ converges to a value of 0.68 ± 0.01 eV/Å^2 with 5 layers (Table C.2), where the value continues to fluctuate marginally for the surface with 7 layers. These values are consistent with similar studies carried out with other PW methods. For example, our values are especially compatible with the 13-layer converged surface of *Singh-Miller et al.* [202], where agreement was found within 0.02 eV/Å^2.

C.3.2 Interlayer Spacing

The interlayer spacing (or interlayer relaxation) describes the change in distance between each of the metallic layers relative to the unrelaxed surface (in the bulk positions):

$$\Delta d_{ij}\,\% = \frac{d^{relaxed}_{ij} - d^{unrelaxed}}{d^{relaxed}} \times 100\,\% \tag{C.3}$$

where $d^{relaxed}_{ij}$ is the distance between layers i and j and $d^{unrelaxed}_{ij}$ is taken as $a_0/\sqrt{3}$ (along the (111) direction of a FCC metal), where a_o is the equilibrium bulk lattice constant. Throughout the paper, a positive value of Δd_{ij} % corresponds to an expansion and a negative to a compression of the surface slab relative to the unrelaxed structure.

Previous benchmark studies by *Singh-Miller et al.* [202] using the PWSCF method [202] found that the interlayer spacing for the first layer Pt(111) expands by ∼1.0%, while the second layer contracts ∼0.5 % relative to the bulk value.

Although convergence with respect to the numbers of unit cells was achieved with Pt(7×6×5), the magnitude of Δd_{12} % is smaller than Δd_{23} % for both of these surfaces. This is quantitatively inconsistent with the results of previous benchmark studies, but this is shown to be an artefact of finite size effects, as shown by the increasing the number of layers of the Pt(7×6×N) slab from N=2 to N=9 (Table C.3). For slabs $N < 6$, the

TABLE C.3: Interlayer spacing after full geometry relaxation for a set of increasingly thicker Pt(111) slabs from 2 to 9 layers. These values are compared against **Γ**-point only calculations with GPW (Quickstep), and a PW (PWSCF) calculation with an (8×8) Monkhorst-Pack grid and 13 Pt layers.

Slab Dimensions	Δd_{12} / Å	Δd_{23} / Å	Δd_{12} %	Δd_{23} %
Pt(7×6×2)	2.310	-	0.69	-
Pt(7×6×3)	2.290	2.290	-0.18	-0.18
Pt(7×6×4)	2.313	2.240	0.82	-2.36
Pt(7×6×5)	2.304	2.274	0.43	-0.88
Pt(7×6×6)	2.312	2.272	0.77	-0.97
Pt(7×6×7)	2.313	2.281	0.82	-0.58
Pt(7×6×8)	2.314	2.284	0.86	-0.45
Pt(7×6×9)	2.317	2.280	0.99	-0.62
Pt(8×8×8)[a]	2.315	2.278	1.13	-0.49
PW 13-layers[b]	-	-	0.85	-0.56

[a] GPW [241]. [b] *Singh-Miller et al.* [202].

interlayer spacing fluctuates significantly from the N=9, with Δd_{23} and Δd_{12} varying by up to 0.032 Å. Although the qualitative expansion of Δd_{12} and contraction of Δd_{23} are observed for N=4 to N=6, the magnitude of Δd_{23} % for these thicknesses are relatively larger than Δd_{23} % for $N > 6$. This is quantitatively inconsistent with interlayer spacing changes for N=9 and the thickest slabs of the PW and GPW approaches, where Δd_{23} % contractions are consistently 0.3-0.5% smaller than the expansions of the first and second layer.

However, interlayer spacing changes converge with slab thicknesses $N \geq$ [188], where the maximum fluctuation in Δd_{12} and Δd_{23} is small (0.003 Å). Also, of note is the similarity between our results and those of the GPW **Γ**-point only approach. Here, we see the differences for the Δd_{12} between the Pt(8×8×8) and Pt(7×6×9) surfaces are 0.001 Å, and 0.006 Å for Δd_{23}, respectively. We conclude that a surface slab such as Pt(7×6×7) with the **Γ**-point approach is sufficiently large to provide geometric relaxation parameters comparable with a conventional calculation of a small cell with extensive k-point sampling.

C.4 k-point Convergence of Pt(111)

We performed additional convergence tests with respect to Brillouin Zone sampling by using increasingly large supercells in the xy-direction. Here, we demonstrate that the value of the work function shows relatively small variation from Pt(4×4×5) to Pt(8×8×5), showing a difference of 0.03 eV between the smallest and the largest slabs (Table C.4).

Conversely, the interlayer spacing is more sensitive to BZ sampling. Compared to the converged Pt slab, the smaller systems (Pt(5×6) and Pt(6×6)) exhibit a large contraction

TABLE C.4: The convergence of the work function for slabs with increasing size in the xy-plane to demonstrate convergence is achieved in the **Γ**-point approximation.

Slab Dimension	Work Function ϕ / eV
Pt(4×4×5)	5.65
Pt(5×6×5)	5.66
Pt(6×6×5)	5.66
Pt(7×6×5)	5.70
Pt(8×8×5)	5.68

TABLE C.5: The convergence of interlayer spacing for a 5 layer Pt slab with increasing size in the xy-plane to demonstrate convergence is achieved in the **Γ**-point approximation.

Slab Dimensions	Δd_{12} /Å	Δd_{23} / Å	Δd_{12} %	Δd_{23} %
Pt(4×4×5)	2.330	2.291	1.63	−0.03
Pt(5×6×5)	2.297	2.244	0.21	−2.10
Pt(6×6×5)	2.292	2.252	−0.01	−1.78
Pt(7×6×5)	2.304	2.274	0.43	−0.88
Pt(8×8×5)	2.302	2.263	0.42	−1.30
Pt(8×8×5)[a]	3.202	2.268	0.59	−0.93
PW 13-layers[b]	-	-	0.85	−0.56

[a] *Santarossa et al.* [241]. [b] *Singh-Miller et al.* [202].

of the second layer while the expansion of the first layer tends to be smaller (with the exception of Pt(4×4)). For slabs larger than Pt(7×6), the geometry converges to within 0.01 Å (Table C.5).

Appendix D

Entropy and Free Energies of Adsorption

D.1 Entropy of Adsorption Values Tabulated

Akinola et al. [59] propose two major potential sources of error associated the experimental data of Table D.2. First is the use of the HCl electrolyte, where Cl^- is known to bind to the Pt surface [47, 49, 248]. The co-adsorption of Cl^- to the surface is known to disrupt the packing and orientation of interfacial water, which introduces an error due to a dependence of the adsorbate binding energies on pH and voltage [249, 250]. Secondly, the Van't Hoff linear regression covers a relatively small range of K_{eq} values (Between 1-2 over a range of 283-403 K), meaning the error of the linear fit is relatively large [59]. The presence of these errors for ΔS_{ads}^{solv} is supported by substantially lower entropy values for studies performed with a weakly binding acetate buffer, which reduces the error due to electrolyte. For the adsorption of phenol on the Pt(100), Rh(111) and Rh(100) surfaces, the ΔS_{ads}^{solv} for Pt(100) is predicted to be $+59 \pm 39$ J K^{-1} mol^{-1} relative to $+188$ JK^{-1} mol^{-1} of Bockris et al. for Pt(111) [46]. However, measurements with the acetate buffer still predict a high entropy value for phenol on the Pt(111) facet, but with a substantial error estimate ($+174 \pm 64$ J K^{-1} mol^{-1}). Further evidence shows that water on the Pt(100) and Pt(111) facets have similar dynamics and adhesion strength [226], suggesting that increased ordering of water does not rationalise the large entropy differences between the two facets. Deciding whether the upper or lower bounds better physically reflects adsorption at the solvated Pt surface requires more data measured with the acetate buffer. The large experimental errors associated with the with the Van't Hoff approach for weakly adsorbing species adds further ambiguity.

TABLE D.1: Entropy of adsorption values in the gaseous and aqueous phase for molecules adsorbed horizontally to the Pt(111) surface. Lower bound (LB) calculcated by statistical mechanical arguments and accentric factor approximations of the cavitation entropy. Upper bound (UB) calculated as the entropy of fusion for the approximate number of H_2O molecules displaced based on the intersecting surface area of the metallic surface and adsorbate. All values stated in J K^{-1} mol^{-1}.

Molecule	ΔS_{ads}^{vac}	ΔS_{ads}^{solv} LB	ΔS_{ads}^{solv} UB	ΔS_{ads}^{solv} Expt. [47]	# H_2O Displaced
Benzoic Acid	-141	103	175	191	4.05
Naphtholic Acid	-150	162	271	254	5.55
Naphthol	-146	136	254	244	5.68
Phenol	-135	75	151	188	4.05
Butanol	-144	90	151	59	3.21
Valeric Acid	-153	109	180	81	3.74

TABLE D.2: Entropy of adsorption values in the gaseous and aqueous phase for molecules adsorbed vertically to the Pt(111) surface. Lower bound (LB) calculated by statistical mechanical arguments and accentric factor approximations of the cavitation entropy. Upper bound (UB) calculated as the entropy of fusion for the approximate number of H_2O molecules displaced based on the intersecting surface area of the metallic surface and adsorbate. All values stated in J K^{-1} mol^{-1}.

Molecule	ΔS_{ads}^{vac}	ΔS_{ads}^{solv} LB	ΔS_{ads}^{solv} UB	ΔS_{ads}^{solv} Expt. [47]	# H_2O Displaced
Benzoic Acid	141	57	74	191	1.29
Naphtholic Acid	150	89	115	254	1.92
Naphthol	146	68	89	244	1.63
Phenol	135	37	54	188	1.31
Butanol	144	49	68	59	1.50
Valeric Acid	153	66	90	81	1.82

D.2 Coverage Geometries

The slab geometries used through out are based on a four layer, orthogonal Pt(111) facet with dimensions of (7×6), (8×8) or (10×8) where appropriate for the desired coverage. Coverage was also varied by increasing the number of adsorbates on the surface, and is normalised to the saturation coverage of each compound. Large simulations cells were necessary as ONETEP uses the Γ-point only for total energy calculations, meaning the unit cell in real space must be large enough to sufficiently sample reciprocal space for a single **k**-point.

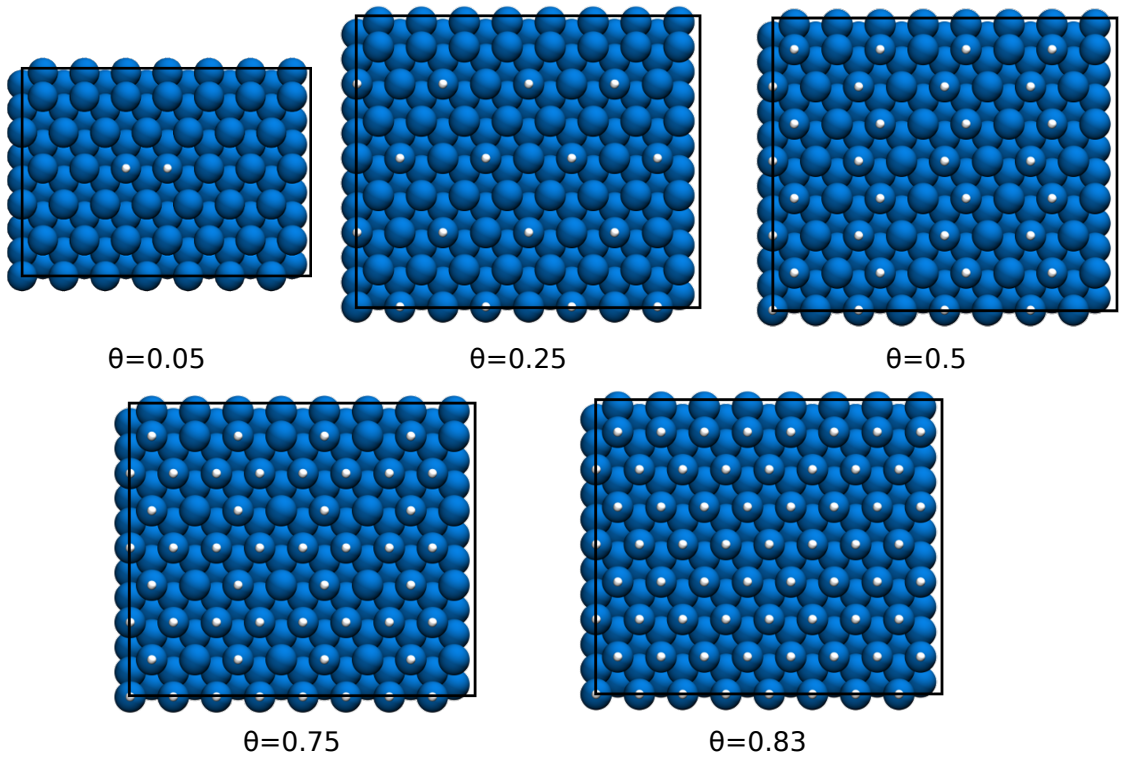

FIGURE D.1: The coverages of hydrogen on the Pt(111) facet.

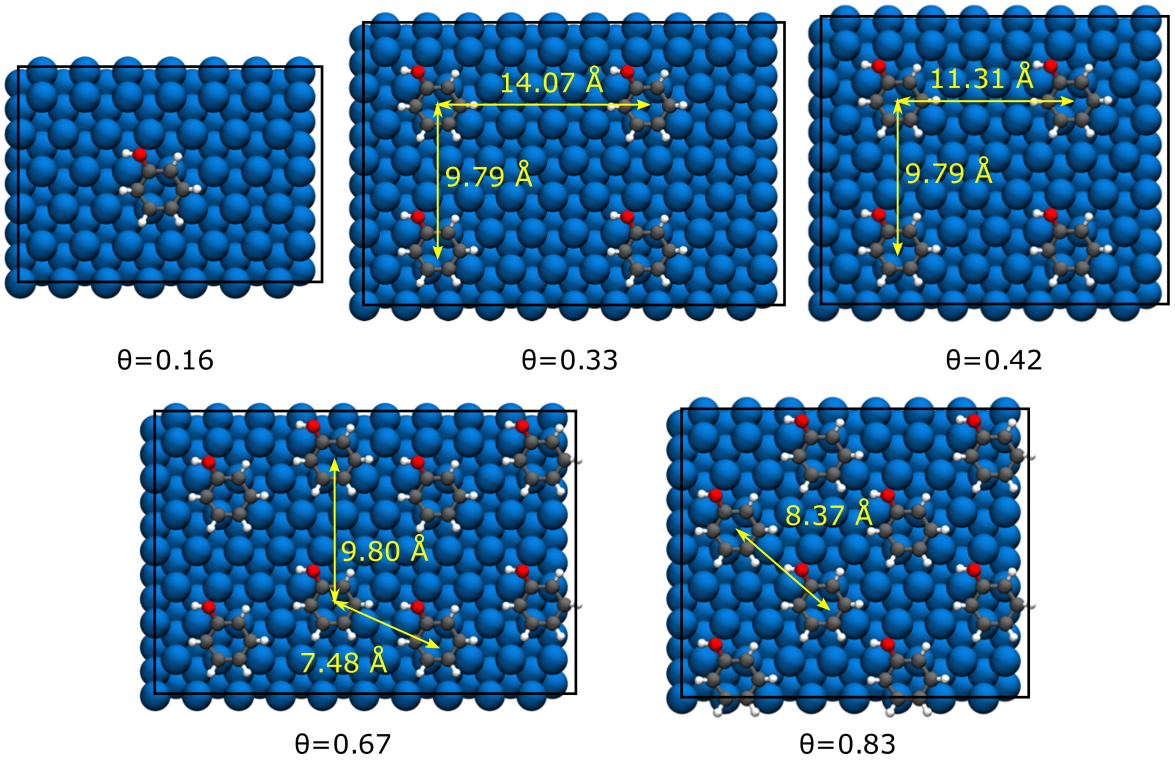

FIGURE D.2: The coverages of phenol on the Pt(111) facet. All adsorbates horizontal with respect to the surface.

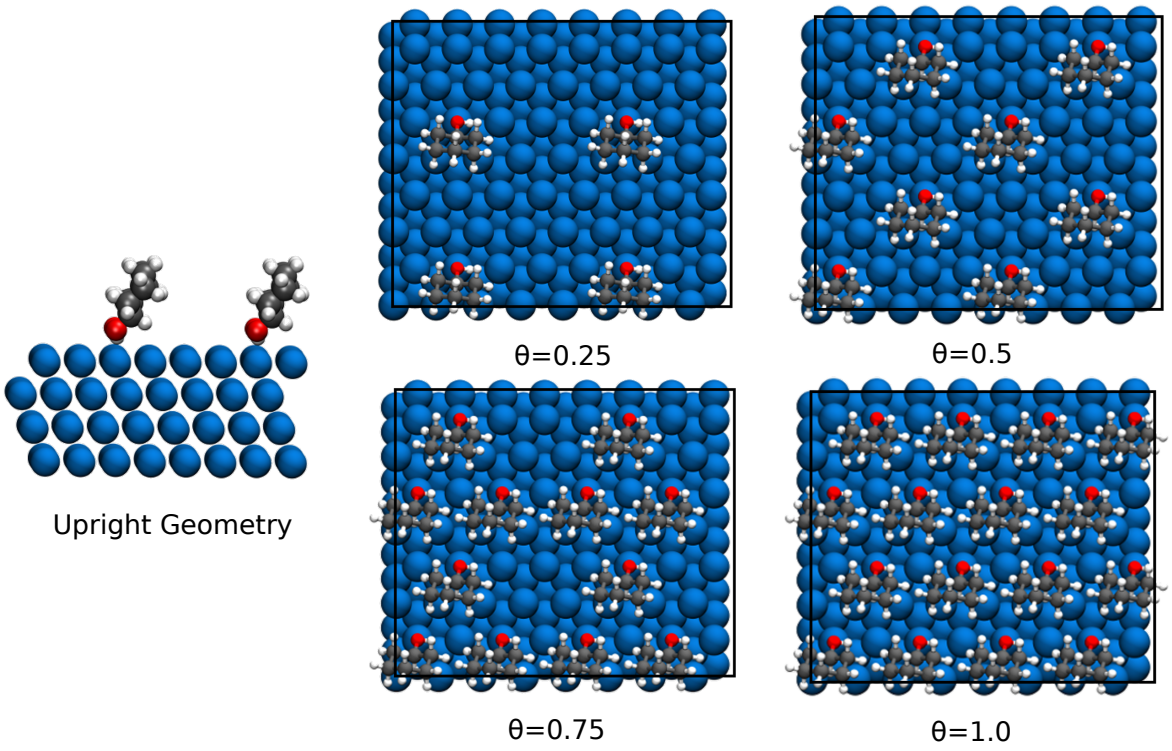

FIGURE D.3: The coverages of cyclohexanol on the Pt(111) facet. All adsorbates horizontal with respect to the surface.

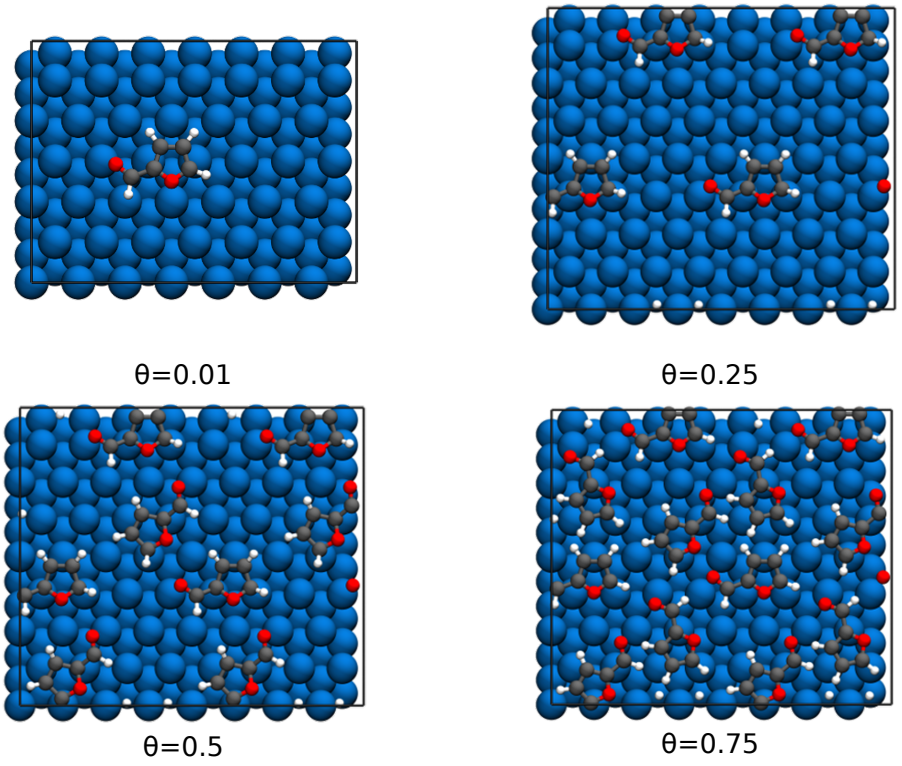

FIGURE D.4: The coverages of furfural on the Pt(111) facet. All adsorbates vertical with respect to the surface.

D.3 Entropy of Solvation and Adsorption Values

TABLE D.3: The translational ($S^{vac,3D}_{Mol,t}$ Equation A.2) and rotational ($S^{vac,3D}_{Mol,r}$ Equation A.4) components of the standard 3D entropies of all measured molecules in vacuum ($S^{vac,3D}_{Mol}$).

Molecule	m /kg	V /m^3	$S^{vac,3D}_{Mol,\ t}$ /J K^{-1} mol^{-1}	σ_r	I /kg m^2	$S^{vac,3D}_{Mol,\ r}$ /J K^{-1} mol^{-1}	$S^{vac,3D}_{Mol}$ /J K^{-1} mol^{-1}
Benzaldehyde	1.76E-25	3.99E-26	167	1	3.91E-45	121	288
Benzene	1.30E-25	3.99E-26	163	6	1.87E-45	97	260
Benzoic Acid	2.03E-25	3.99E-26	169	1	5.17E-45	124	293
Butanol	1.23E-25	3.99E-26	162	1	2.04E-45	113	275
Cyclohexanol	1.66E-25	3.99E-26	166	1	3.37E-45	119	285
Furfural	1.60E-25	3.99E-26	166	1	2.97E-45	117	283
H_2	3.35E-27	3.99E-26	117	2	8.37E-48	38	156
Naphthoic Acid	2.86E-25	3.99E-26	173	1	1.11E-44	134	307
Phenol	1.56E-25	3.99E-26	165	1	2.97E-45	117	283
Naphthol	2.39E-25	3.99E-26	171	1	7.28E-45	129	299
Valeric Acid	1.70E-25	3.99E-26	166	1	4.27E-45	122	288

TABLE D.4: The translational ($S^{vac,2D}_{Mol,t}$ Equation A.6) and rotational ($S^{vac,2D}_{Mol,r}$ Equation A.8) components of the standard 2D entropies of all measured molecules in solvent ($S^{vac,2D}_{Mol}$). The area (A) is taken as the total area of the Pt(7×6) metallic slab.

Molecule	m /kg	A /m^2	$S^{vac,2D}_{Mol,t}$ /J K^{-1} mol^{-1}	σ_r	I /kgm^2	$S^{vac,2D}_{Mol,r}$ /J K^{-1} mol^{-1}	$S^{vac,2D}_{Mol}$ /J K^{-1} mol^{-1}
Benzaldehyde	1.76E-25	4.54E-19	87	1	2.93E-45	49	136
Benzene	1.30E-25	3.76E-19	83	6	1.49E-45	31	114
Benzoic Acid	2.03E-25	4.22E-19	88	1	8.01E-45	53	141
Butanol	1.23E-25	3.35E-19	81	1	4.34E-45	63	144
Cyclohexanol	1.66E-25	1.96E-19	80	1	2.72E-45	59	138
Furfural	1.60E-25	3.59E-19	84	1	2.01E-45	48	132
H_2	1.67E-27	7.30E-20	66	-	-	0	66
Naphthoic Acid	2.86E-25	5.76E-19	93	1	1.94E-44	57	150
Phenol	2.39E-25	5.84E-19	92	1	1.15E-44	55	146
Naphthol	1.56E-25	4.27E-19	86	1	2.83E-45	49	135
Valeric Acid	1.70E-25	4.00E-19	86	1	7.46E-45	67	153

TABLE D.5: The translational ($S^{solv,3D}_{Mol,t}$ A.2) and rotation ($S^{solv,3D}_{Mol,r}$ A.13) components of the standard 3D entropies of all measured molecules in solvent ($S^{solv,3D}_{Mol}$). $V_{free} = 8.91E-30\text{m}^3$ for the calculation of ν_c (Equation A.11).

Molecule	ν_c /m^3	$S^{solv,3D}_{Mol,t}$ /J K^{-1} mol^{-1}	r_c /m	r_g /m	$S^{solv,3D}_{Mol,r}$ /J K^{-1} mol^{-1}
Benzaldehyde	3.31E-28	127	4.29E-10	2.28E-10	98
Benzene	2.90E-28	122	4.11E-10	2.03E-10	76
Benzoic Acid	3.47E-28	129	4.36E-10	2.46E-10	99
Butanol	2.94E-28	121	4.12E-10	2.18E-10	90
Cyclohexanol	3.43E-28	127	4.34E-10	2.14E-10	98
Furfural	2.87E-28	125	4.09E-10	2.07E-10	96
H_2	8.30E-29	66	2.71E-10	5.00E-11	29
Naphthoic Acid	4.41E-28	135	4.72E-10	3.04E-10	104
Phenol	4.02E-28	132	4.58E-10	2.75E-10	101
Naphthol	3.12E-28	125	4.21E-10	2.16E-10	95
Valeric Acid	3.35E-28	127	4.31E-10	2.28E-10	99

TABLE D.6: The cavity ($S^{solv,3D}_{Mol,c}$ Equation A.20) component of the standard 3D entropies of all measured molecules in solvent ($S^{solv,3D}_{Mol}$) and the solvation entropy relative to the vacuum phase ($\Delta S^{solv,3D}_{Mol}$). $\phi_S = 0.51$ (Equation A.17), $\Delta S^{solv}_{r,S} = -1.13 \times 10^{-23}$J K^{-1} mol^{-1} (Equation A.13), and $V_S = 2.1 \times 10^{-29} m^3$. $S^{solv,3D}_{Mol}$ includes the entropy of bringing 1 atm of gas to 1M solution $\Delta S_{conc} = -k \ln (c_f/c_i) = -53.20$ J K^{-1} mol^{-1}.

Molecule	V_m /m^3	A_m /m^2	A_{box} /m^2	S_0 /J K^{-1}mol^{-1}	S_c /J K^{-1} mol^{-1}	$S^{solv,3D}_{Mol}$ /J K^{-1} mol^{-1}	$\Delta S^{solv,3D}_{Mol,r}$ /J K^{-1} mol^{-1}
Benzaldehyde	1.14E-28	1.32E-18	2.23E-18	−54	−81	117	−198
Benzene	9.41E-29	1.11E-18	1.93E-18	−45	−68	103	−157
Benzoic Acid	1.22E-28	1.41E-18	2.60E-18	−58	−89	112	−180
Butanol	9.57E-29	1.20E-18	2.10E-18	−46	−71	114	−161
Cyclohexanol	1.20E-28	1.38E-18	2.76E-18	−57	−90	107	−178
Furfural	9.26E-29	1.13E-18	1.93E-18	−44	−67	126	−157
H_2	1.20E-29	2.54E-19	4.05E-19	−6	−11	58	−98
Naphthoic Acid	1.70E-28	1.85E-18	3.31E-18	−81	−121	92	−215
Naphthol	1.50E-28	1.64E-18	2.67E-18	−72	−103	104	−196
Phenol	1.04E-28	1.22E-18	2.15E-18	−50	−76	118	−165
Valeric Acid	1.16E-28	1.39E-18	2.95E-18	−55	−90	108	−180

D.4 Free Energy of Adsorption Values

The free energies of adsorption, heats of adsorption and entropies of adsorption are presented below for all coverages measured. Geometries are shown in Section S4. In the calculation of $\Delta G^{int,solv}_{ads}$, the entropy of adsorption is taken as ΔS^{vac}_{ads}, where $S^{vac,2D}_{Mol,t}$ is calculated using an area corresponding to the cross-sectional area of the molecule at the binding interface (A_{ov}).

TABLE D.7: The free energies and enthalpies of adsorption in vacuum and solvent for hydrogen, normalised to a single hydrogen atom ($0.5H_2 + Pt \rightarrow Pt/H$).

Coverage	$\Delta H_{ads}^{int,vac}$ /kJ mol^{-1}	ΔS_{ads}^{vac} /J K^{-1} mol^{-1}	$\Delta G_{ads}^{int,vac}$ /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ LB /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ UB /kJ mol^{-1}
1.00	−34	−39	−22	−18	−18
0.75	−37	−37	−26	−21	−21
0.50	−38	−36	−27	−22	−24
0.25	−41	−39	−29	−25	−28
0.05	−44	−14	−40	−31	−38

TABLE D.8: The free energies and enthalpies of adsorption in vacuum and solvent for phenol.

Coverage	$\Delta H_{ads}^{int,vac}$ /kJ mol^{-1}	ΔS_{ads}^{vac} /J K^{-1} mol^{-1}	$\Delta G_{ads}^{int,vac}$ /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ LB /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ UB /kJ mol^{-1}
0.83	−185	−158	−138	−99	−108
0.67	−187	−144	−144	−102	−109
0.42	−199	−152	−153	−106	−113
0.33	−200	−150	−155	−106	−115
0.16	−234	−132	−195	−134	−151

TABLE D.9: The free energies and enthalpies of adsorption in vacuum and solvent for furfural.

Coverage	$\Delta H_{ads}^{int,vac}$ /kJ mol^{-1}	ΔS_{ads}^{vac} /J K^{-1} mol^{-1}	$\Delta G_{ads}^{int,vac}$ /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ LB /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ UB /kJ mol^{-1}
1.00 - Vertical	−91	−153	−45	−36	−39
0.75	−177	−151	−132	−107	−113
0.50	−192	−147	−148	−120	−125
0.25	−179	−142	−136	−87	−95
0.01	−205	−133	−165	−107	−122

TABLE D.10: The free energies and enthalpies of adsorption in vacuum and solvent for cyclohexanol.

Coverage	$\Delta H_{ads}^{int,vac}$ /kJ mol^{-1}	ΔS_{ads}^{vac} /J K^{-1} mol^{-1}	$\Delta G_{ads}^{int,vac}$ /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ LB /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ UB /kJ mol^{-1}
1.00	−82	−144	−40	−14	−18
0.75	−78	−142	−36	0	−1
0.50	−68	−139	−27	7	3
0.25	−72	−133	−33	−1	−2
0.01 - Horizontal	−102	−125	−65	−17	−40

TABLE D.11: The free energies and enthalpies of adsorption in vacuum and solvent for two molecules for a single coverage.

Molecule	Coverage	$\Delta H_{ads}^{int,vac}$ /kJ mol^{-1}	ΔS_{ads}^{vac} /J K^{-1} mol^{-1}	$\Delta G_{ads}^{int,vac}$ /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ LB /kJ mol^{-1}	$\Delta G_{ads}^{int,solv}$ UB /kJ mol^{-1}
Benzaldehyde	0.16	−249	−153	−203	−143	−162
Benzene	0.16	−227	−146	−184	−141	−157

Bibliography

(1) G. Yang, S. A. Akhade, X. Chen, Y. Liu, M. S. Lee, V. A. Glezakou, R. Rousseau and J. A. Lercher, *Angewandte Chemie - International Edition*, 2019, **58**, 3527–3532.

(2) S. Sakong, K. Forster-Tonigold and A. Groß, *Journal of Chemical Physics*, 2016, **144**, 194701.

(3) Y. Yoon, R. Rousseau, R. S. Weber, D. Mei and J. A. Lercher, *Journal of the American Chemical Society*, 2014, **136**, 10287–10298.

(4) C. J. Bodenschatz, S. Sarupria and R. B. Getman, *Journal of Physical Chemistry C*, 2015, **119**, 13642–13651.

(5) X. Zhang, R. S. DeFever, S. Sarupria and R. B. Getman, *Journal of Chemical Information and Modeling*, 2019, **59**, 2190–2198.

(6) H. Hu, Z. Lu and W. Yang, *Journal of Chemical Theory and Computation*, 2007, **3**, 390–406.

(7) M. Faheem, S. Suthirakun and A. Heyden, *Journal of Physical Chemistry C*, 2012, **116**, 22458–22462.

(8) M. Zare, M. Saleheen, S. K. Kundu and A. Heyden, *Communications Chemistry*, 2020, **3**, 187.

(9) S. N. Steinmann, R. Ferreira De Morais, A. W. Götz, P. Fleurat-Lessard, M. Iannuzzi, P. Sautet and C. Michel, *Journal of Chemical Theory and Computation*, 2018, **14**, 3238–3251.

(10) P. Clabaut, P. Fleurat-Lessard, C. Michel and S. N. Steinmann, *Journal of Chemical Theory and Computation*, 2020, **16**, 4565–4578.

(11) P. Clabaut, B. Schweitzer, A. W. Götz, C. Michel and S. N. Steinmann, *Journal of Chemical Theory and Computation*, 2020, **16**, 6539–6549.

(12) C. T. Campbell and J. R. Sellers, *Journal of the American Chemical Society*, 2012, **134**, 18109–18115.

(13) J. Tomasi, B. Mennucci and R. Cammi, *Chemical Reviews*, 2005, **105**, 2999–3093.

(14) A. Klamt and G. Schüürmann, *Journal of the Chemical Society, Perkin Transactions 2*, 1993, **2**, 799–805.

(15) B. Mennucci, J. Tomasi, R. Cammi, J. R. Cheeseman, M. J. Frisch, F. J. Devlin, S. Gabriel and P. J. Stephens, *Journal of Physical Chemistry A*, 2002, **106**, 6102–6113.

(16) W. J. Yin, M. Krack, X. Li, L. Z. Chen and L. M. Liu, *Progress in Natural Science: Materials International*, 2017, **27**, 283–288.

(17) J. Lu, S. Behtash, O. Mamun and A. Heyden, *ACS Catalysis*, 2015, **5**, 2423–2435.

(18) G. W. Huber, S. Iborra and A. Corma, *Synthesis of transportation fuels from biomass: Chemistry, catalysts, and engineering*, 2006.

(19) K. Ohno, *Tracking Clean Energy Progress*, International Energy Agency, 2021, vol. 63, pp. 283–286.

(20) J. Zetterholm, E. Wetterlund, K. Pettersson and J. Lundgren, *Energy*, 2018, **144**, 564–575.

(21) T. R. Brown, R. Thilakaratne, R. C. Brown and G. Hu, *Fuel*, 2013, **106**, 463–469.

(22) R. E. Sims, W. Mabee, J. N. Saddler and M. Taylor, *Bioresource Technology*, 2010, **101**, 1570–1580.

(23) D. P. Ho, H. H. Ngo and W. Guo, *Bioresource Technology*, 2014, **169**, 742–749.

(24) T. Koizumi, in *SpringerBriefs in Applied Sciences and Technology*, Pergamon, 2014, vol. 52, pp. 103–121.

(25) M. C. Rulli, D. Bellomi, A. Cazzoli, G. De Carolis and P. D'Odorico, *Scientific Reports*, 2016, **6**, 22521.

(26) S. V. Vassilev, D. Baxter, L. K. Andersen, C. G. Vassileva and T. J. Morgan, *Fuel*, 2012, **94**, 1–33.

(27) A. V. Bridgwater and G. V. Peacocke, *Renewable and sustainable energy reviews*, 2000, **4**, 1–73.

(28) Y. Solantausta, N. O. Nylund, M. Westerholm, T. Koljonen and A. Oasmaa, *Bioresource Technology*, 1993, **46**, 177–188.

(29) P. M. Mortensen, J. D. Grunwaldt, P. A. Jensen, K. G. Knudsen and A. D. Jensen, *Applied Catalysis A: General*, 2011, **407**, 1–19.

(30) C. A. Mullen, A. A. Boateng, N. M. Goldberg, I. M. Lima, D. A. Laird and K. B. Hicks, *Biomass and Bioenergy*, 2010, **34**, 67–74.

(31) A. Oasmaa and E. Kuoppala, *Energy and Fuels*, 2003, **17**, 1075–1084.

(32) D. Yue, F. You and S. W. Snyder, *Computers and Chemical Engineering*, 2014, **66**, 36–56.

(33) E. Furimsky, *Applied Catalysis A: General*, 2000, **199**, 147–190.

(34) J. Wildschut, F. H. Mahfud, R. H. Venderbosch and H. J. Heeres, *Industrial & Engineering Chemistry Research*, 2009, **48**, 10324–10334.

(35) F. E. Massoth, P. Politzer, M. C. Concha, J. S. Murray, J. Jakowski and J. Simons, *The Journal of Physical Chemistry B*, 2006, **110**, 14283–14291.

(36) K. Li, R. Wang and J. Chen, *Energy and Fuels*, 2011, **25**, 854–863.

(37) D. D. Laskar, M. P. Tucker, X. Chen, G. L. Helms and B. Yang, *Green Chemistry*, 2014, **16**, 897–910.

(38) C. R. Lee, J. S. Yoon, Y. W. Suh, J. W. Choi, J. M. Ha, D. J. Suh and Y. K. Park, *Catalysis Communications*, 2012, **17**, 54–58.

(39) Y. Yu, X. Li, L. Su, Y. Zhang, Y. Wang and H. Zhang, *Applied Catalysis A: General*, 2012, **447-448**, 115–123.

(40) C. Ju, M. Li, Y. Fang and T. Tan, *Green Chemistry*, 2018, **20**, 4492–4499.

(41) E. A. Roldugina, E. R. Naranov, A. L. Maximov and E. A. Karakhanov, *Applied Catalysis A: General*, 2018, **553**, 24–35.

(42) W. Zhang, J. Chen, R. Liu, S. Wang, L. Chen and K. Li, *ACS Sustainable Chemistry and Engineering*, 2014, **2**, 683–691.

(43) Z. Liu, I. A. Hamad, Y. Li, Y. Chen, S. Wang, R. E. Jentoft and F. C. Jentoft, *Applied Catalysis A: General*, 2019, **585**, 117199.

(44) Y. Dong, H. Zhao, Z. Liu, M. Yang, Z. Zhang, T. Zhu and H. Cheng, *RSC Advances*, 2020, **10**, 11039–11045.

(45) N. Singh, U. Sanyal, J. L. Fulton, O. Y. Gutiérrez, J. A. Lercher and C. T. Campbell, *ACS Catalysis*, 2019, **9**, 6869–6881.

(46) M. Bockris, M. Green, D. A. J Swinkels, E. A. Blomgren and C. Jesch, *Relationships of Organic Compounds*, tech. rep. 6, 1960, p. 1948.

(47) J. O. Bockris and K. T. Jeng, *Journal of Electroanalytical Chemistry*, 1992, **330**, 541–581.

(48) N. Singh and C. T. Campbell, *ACS Catalysis*, 2019, **9**, 8116–8127.

(49) J. Akinola, I. Barth, B. R. Goldsmith and N. Singh, *ACS Catalysis*, 2020, **10**, 4929–4941.

(50) K. D. Etzel, K. R. Bickel and R. Schuster, *Review of Scientific Instruments*, 2010, **81**, 34101.

(51) K. R. Bickel, K. D. Etzel, V. Halka and R. Schuster, *Electrochimica Acta*, 2013, **112**, 801–812.

(52) M. P. Soriaga and A. T. Hubbard, *Journal of the American Chemical Society*, 1982, **104**, 3937–3945.

(53) M. P. Soriaga, J. H. White, D. Song and A. T. Hubbard, *Journal of Electroanalytical Chemistry*, 1984, **171**, 359–363.

(54) N. De Jonge and F. M. Ross, *Nature Nanotechnology*, 2011, **6**, 695–704.

(55) M. Saleheen and A. Heyden, *ACS Catalysis*, 2018, **8**, 2188–2194.

(56) S. K. Iyemperumal and N. A. Deskins, *ChemPhysChem*, 2017, **18**, 2171–2190.

(57) N. Ayawei, A. N. Ebelegi and D. Wankasi, *Journal of Chemistry*, 2017, **2017**, ed. W. Guo, 3039817.

(58) S. J. Carey, W. Zhao, Z. Mao and C. T. Campbell, *Journal of Physical Chemistry C*, 2019, **123**, 7627–7632.

(59) J. Akinola and N. Singh, *Journal of Applied Electrochemistry*, 2021, **51**, 37–50.

(60) N. Singh, M. T. Nguyen, D. C. Cantu, B. L. Mehdi, N. D. Browning, J. L. Fulton, J. Zheng, M. Balasubramanian, O. Y. Gutiérrez, V. A. Glezakou, R. Rousseau, N. Govind, D. M. Camaioni, C. T. Campbell and J. A. Lercher, *Journal of Catalysis*, 2018, **368**, 8–19.

(61) S. J. Carey, W. Zhao, Z. Mao and C. T. Campbell, *Journal of Physical Chemistry C*, 2019, **123**, 7627–7632.

(62) S. J. Carey, W. Zhao and C. T. Campbell, *Surface Science*, 2018, **676**, 9–16.

(63) O. Lytken, W. Lew, J. J. Harris, E. K. Vestergaard, J. M. Gottfried and C. T. Campbell, *Journal of the American Chemical Society*, 2008, **130**, 10247–10257.

(64) J. M. Gottfried, E. K. Vestergaard, P. Bera and C. T. Campbell, *Journal of Physical Chemistry B*, 2006, **110**, 17539–17545.

(65) E. M. Karp, T. L. Silbaugh, M. C. Crowe and C. T. Campbell, *Journal of the American Chemical Society*, 2012, **134**, 20388–20395.

(66) Y. Song, U. Sanyal, D. Pangotra, J. D. Holladay, D. M. Camaioni, O. Y. Gutiérrez and J. A. Lercher, *Journal of Catalysis*, 2018, **359**, 68–75.

(67) J. R. Rumptz and C. T. Campbell, *ACS Catalysis*, 2019, **9**, 11819–11825.

(68) C. McGreavy, *Introduction to statistical thermodynamics, Terrell L. Hill, Adison-wesley publishing company, incorporated, burlington, massachusetts(1960). 508 pages.$9.75*, John Wiley & Sons, Ltd, 1963, vol. 9, pp. 144–144.

(69) M. Jørgensen and H. Grönbeck, *Journal of Physical Chemistry C*, 2017, **121**, 7199–7207.

(70) H. Terrell, *An Introduction to Statistical Thermodynamics*, Addison-Wesley Publishing Company, Inc., Reading, Massachuetts, 1960, pp. 124–128.

(71) L. H. Sprowl, C. T. Campbell and L. Árnadóttir, *The Journal of Physical Chemistry C*, 2016, **120**, 9719–9731.

(72) C. T. Campbell, L. H. Sprowl and L. Árnadóttir, *Journal of Physical Chemistry C*, 2016, **120**, 10283–10297.

(73) P. J. Dauenhauer and O. A. Abdelrahman, *ACS Central Science*, 2018, **4**, 1235–1243.

(74) *WebBook NIST (2017) Standard Reference Database Number 69*, National Institute of Standards and Technology, Gaithersburg MD, 2017, p. 20899.

(75) L. G. Verga, A. E. Russell and C. K. Skylaris, *Physical Chemistry Chemical Physics*, 2018, **20**, 25918–25930.

(76) E. Schrödinger, *Annalen der Physik*, 1926, **384**, 489–527.

(77) A. Szabo and N. S. Ostlund, *Modern quantum chemistry : introduction to advanced electronic structure theory*, Dover Publications, 1996, p. 466.

(78) P. Hohenberg and W. Kohn, *Physical Review*, 1964, **136**, B864–B871.

(79) E. Gey, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, 1995, vol. 191, pp. 277–278.

(80) W. Kohn and L. J. Sham, *Physical Review*, 1965, **140**, A1133–A1138.

(81) G. Kresse and J. Furthmüller, *Physical Review B - Condensed Matter and Materials Physics*, 1996, **54**, 11169–11186.

(82) J. F. Annett, *Computational Materials Science*, 1995, **4**, 23–42.

(83) Á. Ruiz-Serrano and C. K. Skylaris, *Journal of Chemical Physics*, 2013, **139**, 054107.

(84) N. D. Mermin, *Physical Review*, 1965, **137**, 1441–1443.

(85) M. S. Rogalski and S. B. Palmer, *Solid state physics*, Brooks/Cole, Australia ; 2014, pp. 1–467.

(86) F. Bloch, *Zeitschrift für Physik*, 1929, **52**, 555–600.

(87) K. Hu, M. Wu, S. Hinokuma, T. Ohto, M. Wakisaka, J. I. Fujita and Y. Ito, *Journal of Materials Chemistry A*, 2019, **7**, 2156–2164.

(88) E. Prodans and W. Kohn, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 11635–11638.

(89) W. Kohn, *Physical Review Letters*, 1996, **76**, 3168–3171.

(90) P. D. Haynes and M. C. Payne, *Physical Review B - Condensed Matter and Materials Physics*, 1999, **59**, 12173–12176.

(91) J. C. Prentice, J. Aarons, J. C. Womack, A. E. Allen, L. Andrinopoulos, L. Anton, R. A. Bell, A. Bhandari, G. A. Bramley, R. J. Charlton, R. J. Clements, D. J. Cole, G. Constantinescu, F. Corsetti, S. M. Dubois, K. K. Duff, J. M. Escartín, A. Greco, Q. Hill, L. P. Lee, E. Linscott, D. D. O'Regan, M. J. Phipps, L. E. Ratcliff, Á. R. Serrano, E. W. Tait, G. Teobaldi, V. Vitale, N. Yeung, T. J. Zuehlsdorff, J. Dziedzic, P. D. Haynes, N. D. Hine, A. A. Mostofi, M. C. Payne and C. K. Skylaris, *Journal of Chemical Physics*, 2020, **152**, 174111.

(92) X. P. Li, R. W. Nunes and D. Vanderbilt, *Physical Review B*, 1993, **47**, 10891–10894.

(93) G. Boschetto, H.-T. Xue, J. Dziedzic, M. Krompiec and C.-K. Skylaris, *The Journal of Physical Chemistry C*, 2017, **121**, 2529–2538.

(94) Y.-C. Hung, C.-Y. Chao, C.-A. Dai, W.-F. Su and S.-T. Lin, *The Journal of Physical Chemistry B*, 2013, **117**, 690–696.

(95) C. F. Melius and W. A. Goddard, *Physical Review A*, 1974, **10**, 1528–1540.

(96) A. Zunger and M. L. Cohen, *Physical Review Letters*, 1978, **41**, 53–56.

(97) D. R. Hamann, M. Schlüter and C. Chiang, *Physical Review Letters*, 1979, **43**, 1494–1497.

(98) P. E. Blöchl, *Physical Review B*, 1994, **50**, 17953–17979.

(99) C. Rostgaard, *arXiv: Materials Science*, 2009.

(100) D. Vanderbilt, *Physical Review B*, 1990, **41**, 7892–7895.

(101) D. Joubert, *Physical Review B - Condensed Matter and Materials Physics*, 1999, **59**, 1758–1775.

(102) R. P. Feynman, *Physical Review*, 1939, **56**, 340–343.

(103) P. Pulay, *Molecular Physics*, 1969, **17**, 197–204.

(104) Á. Ruiz-Serrano, N. D. Hine and C. K. Skylaris, *Journal of Chemical Physics*, 2012, **136**, 234101.

(105) J. Nocedal and S. J. Wright, *Numerical optimization*, Springer, Madison, WI, 2006, pp. 1–664.

(106) D. Packwood, J. Kermode, L. Mones, N. Bernstein, J. Woolley, N. Gould, C. Ortner and G. Csányi, *Journal of Chemical Physics*, 2016, **144**, 164109.

(107) J. Nocedal, *Mathematics of Computation*, 1980, **35**, 773.

(108) B. G. Pfrommer, M. Côté, S. G. Louie and M. L. Cohen, *Journal of Computational Physics*, 1997, **131**, 233–240.

(109) C. Wang, J. Wang, Q. Cai, Z. Li, H.-K. Zhao and R. Luo, *Computational & theoretical chemistry*, 2013, **1024**, 34–44.

(110) J. C. Womack, L. Anton, J. Dziedzic, P. J. Hasnip, M. I. Probert and C. K. Skylaris, *Journal of Chemical Theory and Computation*, 2018, **14**, 1412–1432.

(111) O. Andreussi, I. Dabo and N. Marzari, *Journal of Chemical Physics*, 2012, **136**, 064102.

(112) J. Dziedzic, Q. Hill and C. K. Skylaris, *Journal of Chemical Physics*, 2013, **139**, 214103.

(113) R. A. Pierotti, *Chemical Reviews*, 1976, **76**, 717–726.

(114) J. Dziedzic, H. H. Helal, C. K. Skylaris, A. A. Mostofi and M. C. Payne, *Epl*, 2011, **95**, 43001.

(115) F. M. Floris, J. Tomasi and J. L. Ahuir, *Journal of Computational Chemistry*, 1991, **12**, 784–791.

(116) J. A. Wagoner and N. A. Baker, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 8331–8336.

(117) J. Zhang, H. Zhang, T. Wu, Q. Wang and D. Van Der Spoel, *Journal of Chemical Theory and Computation*, 2017, **13**, 1034–1043.

(118) R. C. Harris and B. M. Pettitt, *Journal of Chemical Theory and Computation*, 2015, **11**, 4593–4600.

(119) R. M. Levy, L. Y. Zhang, E. Gallicchio and A. K. Felts, *Journal of the American Chemical Society*, 2003, **125**, 9523–9530.

(120) J. L. Fattebert and F. Gygi, *Journal of Computational Chemistry*, 2002, **23**, 662–666.

(121) J. L. Fattebert and F. Gygi, *International Journal of Quantum Chemistry*, 2003, **93**, 139–147.

(122) D. A. Scherlis, J. L. Fattebert, F. Gygi, M. Cococcioni and N. Marzari, *Journal of Chemical Physics*, 2006, **124**, 074103.

(123) G. Fisicaro, L. Genovese, O. Andreussi, S. Mandal, N. N. Nair, N. Marzari and S. Goedecker, *Journal of Chemical Theory and Computation*, 2017, **13**, 3829–3845.

(124) A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *Journal of the American Chemical Society*, 1992, **114**, 10024–10035.

(125) A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer and D. G. Truhlar, *Minnesota Solvation Database-version 2012*, 2012.

(126) L. Gundelach, T. Fox, C. S. Tautermann and C. K. Skylaris, *Physical Chemistry Chemical Physics*, 2021, **23**, 9381–9393.

(127) J. Dziedzic, A. Bhandari, L. Anton, C. Peng, J. C. Womack, M. Famili, D. Kramer and C. K. Skylaris, *Journal of Physical Chemistry C*, 2020, **124**, 7860–7872.

(128) A. Bhandari, C. Peng, J. Dziedzic, L. Anton, J. R. Owen, D. Kramer and C. K. Skylaris, *Journal of Chemical Physics*, 2021, **155**, 24114.

(129) S. Alvarez, *Dalton Transactions*, 2013, **42**, 8617–8636.

(130) G. Fisicaro, L. Genovese, O. Andreussi, N. Marzari and S. Goedecker, *Journal of Chemical Physics*, 2016, **144**, 014103.

(131) A. V. Marenich, C. J. Cramer and D. G. Truhlar, *Journal of Physical Chemistry B*, 2009, **113**, 6378–6396.

(132) E. Aprà, E. J. Bylaska, W. A. De Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. Van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H.

Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin Del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-De-La-Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, A. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao and R. J. Harrison, *Journal of Chemical Physics*, 2020, **152**, 184102.

(133) J. P. Perdew, M. Ernzerhof and K. Burke, *Journal of Chemical Physics*, 1996, **105**, 9982–9985.

(134) T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, *Journal of Computational Chemistry*, 1983, **4**, 294–301.

(135) N. D. Hine, J. Dziedzic, P. D. Haynes and C. K. Skylaris, *Journal of Chemical Physics*, 2011, **135**, 204103–2810.

(136) N. Marzari, D. Vanderbilt and M. C. Payne, *Physical Review Letters*, 1997, **79**, 1337–1340.

(137) D. R. Hartree, *Mathematical Proceedings of the Cambridge Philosophical Society*, 1928, **24**, 111–132.

(138) M. Probert, *Electronic structure: basic theory and practical methods, 2nd edition*, Cambridge University Press, Cambridge, 2020, vol. 61, pp. 312–312.

(139) N. D. Woods, M. C. Payne and P. J. Hasnip, *Computing the self-consistent field in Kohn-Sham density functional theory*, 2019.

(140) D. Kincaid and W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing (2nd ed.)* American Mathematical Society, Providence, 3rd edn., 1996, xii, 804 p.

(141) P. Pulay, *Chemical Physics Letters*, 1980, **73**, 393–398.

(142) D. J. Thouless, *Journal of Physics C: Solid State Physics J. Phys*, 1970.

(143) G. Kresse and J. Hafner, *Physical Review B*, 1993, **47**, 558–561.

(144) G. P. Kerker, *Physical Review B*, 1981, **23**, 3082–3084.

(145) Y. Zhou, H. Wang, Y. Liu, X. Gao and H. Song, *Physical Review E*, 2018, **97**, 33305.

(146) C. Freysoldt, S. Boeck and J. Neugebauer, *Physical Review B - Condensed Matter and Materials Physics*, 2009, **79**, 241103.

(147) S. Ismail-Beigi and T. A. Arias, *Computer Physics Communications*, 2000, **128**, 1–45.

(148) M. J. Gillan, *Journal of Physics: Condensed Matter*, 1989, **1**, 689–711.

(149) J. Aarons, M. Sarwar, D. Thompsett and C.-K. Skylaris, *The Journal of Chemical Physics*, 2016, **145**, 220901.

(150) P. O. Löwdin, *The Journal of Chemical Physics*, 1950, **18**, 365–375.

(151) Y. Shiihara, O. Kuwazuru and N. Yoshikawa, *Modelling and Simulation in Materials Science and Engineering*, 2008, **16**, 035004.

(152) A. Ruiz-Serrano, Ph.D. Thesis, University of Southampton, 2013, pp. 163–188.

(153) E. Artacho and L. Milns Del Bosch, *Physical Review A*, 1991, **43**, 5770–5777.

(154) G. W. Luckey, *Introduction to Solid State Physics*, Wiley, 1957, vol. 79, p. 3299.

(155) H. R. Fang and Y. Saad, *Numerical Linear Algebra with Applications*, 2009, **16**, 197–221.

(156) P. P. Pratapa and P. Suryanarayana, *Chemical Physics Letters*, 2015, **635**, 69–74.

(157) G. Bramley, M. T. Nguyen, V. A. Glezakou, R. Rousseau and C. K. Skylaris, *Journal of Chemical Theory and Computation*, 2020, **16**, 2703–2715.

(158) M. J. Eslamibidgoli, J. Huang, T. Kadyk, A. Malek and M. Eikerling, *Nano Energy*, 2016, **29**, 334–361.

(159) J. K. Nørskov, F. Abild-Pedersen, F. Studt and T. Bligaard, *Density functional theory in surface chemistry and catalysis*, 2011.

(160) C. Michel, J. Zaffran, A. M. Ruppert, J. Matras-Michalska, M. Jdrzejczyk, J. Grams and P. Sautet, *Chemical Communications*, 2014, **50**, 12450–12453.

(161) Y. Qiao, F. Xu, S. Xu, D. Yang, B. Wang, X. Ming, J. Hao and Y. Tian, *Energy and Fuels*, 2018, **32**, 10801–10812.

(162) J. J. Varghese and S. H. Mushrif, *Reaction Chemistry & Engineering*, 2019, **4**, 165–206.

(163) C. Shang and Z. P. Liu, *Journal of the American Chemical Society*, 2011, **133**, 9938–9947.

(164) A. Cuesta, *Surface Science*, 2004, **572**, 11–22.

(165) J. Kim, H. Kim, W. J. Lee, B. Ruqia, H. Baik, H. S. Oh, S. M. Paek, H. K. Lim, C. H. Choi and S. I. Choi, *Journal of the American Chemical Society*, 2019, **141**, 18256–18263.

(166) V. Tripkovic, M. E. Björketun, E. Skúlason and J. Rossmeisl, *Physical Review B - Condensed Matter and Materials Physics*, 2011, **84**, 115452.

(167) S. Trasatti, *Pure and Applied Chemistry*, 1986, **58**, 955–966.

(168) R. Rousseau, V. De Renzi, R. Mazzarello, D. Marchetto, R. Biagi, S. Scandolo and U. Del Pennino, *Journal of Physical Chemistry B*, 2006, **110**, 10862–10872.

(169) S. Sakong and A. Groß, *Journal of Chemical Physics*, 2018, **149**, 084705.

(170) J. Le, M. Iannuzzi, A. Cuesta and J. Cheng, *Physical Review Letters*, 2017, **119**, 016801.

(171) M. Faheem, M. Saleheen, J. Lu and A. Heyden, *Catalysis Science and Technology*, 2016, **6**, 8242–8256.

(172) M. J. Gillan, D. Alfè and A. Michaelides, *Journal of Chemical Physics*, 2016, **144**, 130901.

(173) R. Jinnouchi and A. B. Anderson, *Physical Review B - Condensed Matter and Materials Physics*, 2008, **77**, 245417.

(174) R. Sundararaman and K. Schwarz, *The Journal of Chemical Physics*, 2017, **146**, 084111.

(175) N. G. Hörmann, O. Andreussi and N. Marzari, *Journal of Chemical Physics*, 2019, **150**, 41730.

(176) L. Blumenthal, J. M. Kahk, R. Sundararaman, P. Tangney and J. Lischner, *RSC Advances*, 2017, **7**, 43660–43670.

(177) N. G. Hörmann, Z. Guo, F. Ambrosio, O. Andreussi, A. Pasquarello and N. Marzari, *npj Computational Materials*, 2019, **5**, 100.

(178) J. Le, A. Cuesta and J. Cheng, *Journal of Electroanalytical Chemistry*, 2018, **819**, 87–94.

(179) O. A. Vydrov and T. Van Voorhis, *The Journal of Chemical Physics*, 2010, **133**, 244103.

(180) N. D. Mermin, *Physical Review*, 1965, **137**, A1441–A1443.

(181) K. F. Garrity, J. W. Bennett, K. M. Rabe and D. Vanderbilt, *Computational Materials Science*, 2014, **81**, 446–452.

(182) L. G. Verga, J. Aarons, M. Sarwar, D. Thompsett, A. E. Russell and C. K. Skylaris, *Physical Chemistry Chemical Physics*, 2016, **18**, 32713–32722.

(183) C. Witzgall and R. Fletcher, *Mathematics of Computation*, 1989, **53**, 768.

(184) C. A. Rozzi, D. Varsano, A. Marini, E. K. Gross and A. Rubio, *Physical Review B - Condensed Matter and Materials Physics*, 2006, **73**, 205119.

(185) T. Sohier, M. Calandra and F. Mauri, *Physical Review B*, 2017, **96**, 075448–075448.

(186) J. Hutter, M. Iannuzzi, F. Schiffmann and J. Vandevondele, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2014, **4**, 15–25.

(187) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *Journal of Chemical Physics*, 2010, **132**, 154104.

(188) G. Lippert, J. Hutter and M. Parrinello, *Molecular Physics*, 1997, **92**, 477–488.

(189) G. Bussi, D. Donadio and M. Parrinello, *Journal of Chemical Physics*, 2007, **126**, 14101.

(190) J. Cheng and M. Sprik, *Physical Chemistry Chemical Physics*, 2012, **14**, 11245–11267.

(191) S. Trasatti, *Electrochimica Acta*, 1991, **36**, 1659–1667.

(192) J. Cheng, X. Liu, J. VandeVondele, M. Sulpizi and M. Sprik, *Accounts of Chemical Research*, 2014, **47**, 3522–3529.

(193) J. Cheng and J. Vandevondele, *Physical Review Letters*, 2016, **116**, 86402.

(194) J. Cheng, M. Sulpizi and M. Sprik, *Journal of Chemical Physics*, 2009, **131**, 154504.

(195) F. Costanzo, M. Sulpizi, R. G. D. Valle and M. Sprik, *Journal of Chemical Physics*, 2011, **134**, 244508.

(196) F. Ambrosio, Z. Guo and A. Pasquarello, *Journal of Physical Chemistry Letters*, 2018, **9**, 3212–3216.

(197) W. R. Fawcett, *Langmuir*, 2008, **24**, 9868–9875.

(198) R. Gómez, V. Climent, J. M. Feliu and M. J. Weaver, *Journal of Physical Chemistry B*, 2000, **104**, 597–605.

(199) M. Dodia, T. Ohto, S. Imoto and Y. Nagata, *Journal of Chemical Theory and Computation*, 2019, **15**, 3836–3843.

(200) G. N. Derry and Z. Ji-zhong, *Physical Review B*, 1989, **39**, 1940–1941.

(201) A. Patra, J. E. Bates, J. Sun and J. P. Perdew, *Proceedings of the National Academy of Sciences of the United States of America*, 2017, **114**, E9188–E9196.

(202) N. E. Singh-Miller and N. Marzari, *Physical Review B - Condensed Matter and Materials Physics*, 2009, **80**, 235407.

(203) D. C. Cantu, Y. G. Wang, Y. Yoon, V. A. Glezakou, R. Rousseau and R. S. Weber, *Catalysis Today*, 2017, **289**, 231–236.

(204) H. Ogasawara, B. Brena, D. Nordlund, M. Nyberg, A. Pelmenschikov, L. G. Pettersson and A. Nilsson, *Physical Review Letters*, 2002, **89**, 276102.

(205) G. A. Bramley, M.-T. Nguyen, V.-A. Glezakou, R. Rousseau and C.-K. Skylaris, *Journal of Chemical Theory and Computation*, 2022, DOI: 10.1021/acs.jctc.1c00894.

(206) N. Chaudhary, A. Hensley, G. Collinge, Y. Wang and J. S. McEwen, *Journal of Physical Chemistry C*, 2019, **124**, 356–362.

(207) M. S. Lee and M. A. Olson, *Journal of Chemical Physics*, 2013, **139**, 044119.

(208) X. Wang, B. Deng and Z. Sun, *Journal of Molecular Modeling*, 2019, **25**, 1–18.

(209) G. Fisicaro, S. Filice, S. Scalese, G. Compagnini, R. Reitano, L. Genovese, S. Goedecker, I. Deretzis and A. La Magna, *Journal of Physical Chemistry C*, 2020, **124**, 2406–2419.

(210) D. van der Spoel, J. Zhang and H. Zhang, *WIREs Computational Molecular Science*, 2021, 1560.

(211) H. H. Heenen, J. A. Gauthier, H. H. Kristoffersen, T. Ludwig and K. Chan, *Journal of Chemical Physics*, 2020, **152**, 144703.

(212) A. J. Garza, *Journal of Chemical Theory and Computation*, 2019, **15**, 3204–3214.

(213) G. Collinge, S. F. Yuk, M. T. Nguyen, M. S. Lee, V. A. Glezakou and R. Rousseau, *ACS Catalysis*, 2020, **10**, 9236–9260.

(214) B. M. Wong, G. Collinge, A. J. Hensley, Y. Wang and J. S. McEwen, *Progress in Surface Science*, 2019, **94**, 100538.

(215) C. K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Diéguez and M. C. Payne, *Physical Review B - Condensed Matter and Materials Physics*, 2002, **66**, 1–12.

(216) M. Dion, H. Rydberg, E. Schröder, D. C. Langreth and B. I. Lundqvist, *Physical Review Letters*, 2004, **92**, 246401.

(217) N. D. Hine, M. Robinson, P. D. Haynes, C. K. Skylaris, M. C. Payne and A. A. Mostofi, *Physical Review B - Condensed Matter and Materials Physics*, 2011, **83**, 195102.

(218) G. Kresse and J. Hafner, *Physical Review B*, 1994, **49**, 14251–14269.

(219) K. Mathew, R. Sundararaman, K. Letchworth-Weaver, T. A. Arias and R. G. Hennig, *Journal of Chemical Physics*, 2014, **140**, DOI: 10.1063/1.4865107.

(220) K. Mathew, V. S. Kolluru, S. Mula, S. N. Steinmann and R. G. Hennig, *Journal of Chemical Physics*, 2019, **151**, 234101.

(221) A. A. Mostofi, Ph.D. Thesis, University of Cambridge, 2003, pp. 185–186.

(222) M. A. Halim, D. M. Shaw and R. A. Poirier, *Journal of Molecular Structure: THEOCHEM*, 2010, **960**, 63–72.

(223) D. L. Doering and T. E. Madey, *Surface Science*, 1982, **123**, 305–337.

(224) H. Ogasawara, B. Brena, D. Nordlund, M. Nyberg, A. Pelmenschikov, L. G. Pettersson and A. Nilsson, *Physical Review Letters*, 2002, **89**, 276102.

(225) W. Lew, M. C. Crowe, E. Karp and C. T. Campbell, *Journal of Physical Chemistry C*, 2011, **115**, 9164–9170.

(226) S. Gim, K. J. Cho, H. K. Lim and H. Kim, *Scientific Reports*, 2019, **9**, 14805.

(227) M. S. Dionísio, J. J. M. Ramos and R. M. Gonçalves, *Canadian Journal of Chemistry*, 1990, **68**, 1937–1949.

(228) P. Tereshchuk and J. L. F. Da Silva, *The Journal of Physical Chemistry C*, 2013, **117**, 16942–16952.

(229) J. Akinola, C. T. Campbell and N. Singh, *Journal of Physical Chemistry C*, 2021, **125**, 24371–24380.

(230) N. Chaudhary, A. Hensley, G. Collinge, Y. Wang and J. S. McEwen, *Journal of Physical Chemistry C*, 2019, **124**, 356–362.

(231) H. S. Frank and M. W. Evans, *The Journal of Chemical Physics*, 1945, **13**, 507–532.

(232) L. Yan, Y. Yamamoto, M. Shiga and O. Sugino, *Physical Review B*, 2020, **101**, 165414.

(233) Y. Mao, M. Loipersberger, K. J. Kron, J. S. Derrick, C. J. Chang, S. M. Sharada and M. Head-Gordon, *Chemical Science*, 2021, **12**, 1398–1414.

(234) C. Y. Li, J. B. Le, Y. H. Wang, S. Chen, Z. L. Yang, J. F. Li, J. Cheng and Z. Q. Tian, *Nature Materials*, 2019, **18**, 697–701.

(235) S. A. Akhade, N. Singh, O. Y. Gutierrez, J. Lopez-Ruiz, H. Wang, J. D. Holladay, Y. Liu, A. Karkamkar, R. S. Weber, A. B. Padmaperuma, M. S. Lee, G. A. Whyatt, M. Elliott, J. E. Holladay, J. L. Male, J. A. Lercher, R. Rousseau and V. A. Glezakou, *Chemical Reviews*, 2020, **120**, 11370–11419.

(236) H. Chen and C. K. Skylaris, *Physical Chemistry Chemical Physics*, 2021, **23**, 8891–8899.

(237) M. Jarvis, I. White and R. Godby, *Physical Review B - Condensed Matter and Materials Physics*, 1997, **56**, 14972–14978.

(238) M. Otani and O. Sugino, *Physical Review B - Condensed Matter and Materials Physics*, 2006, **73**, 115407.

(239) I. Dabo, B. Kozinsky, N. E. Singh-Miller and N. Marzari, *Physical Review B - Condensed Matter and Materials Physics*, 2008, **77**, 115139.

(240) P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. De Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari and R. M. Wentzcovitch, *Journal of Physics Condensed Matter*, 2009, **21**, 395502.

(241) G. Santarossa, A. Vargas, M. Iannuzzi, C. A. Pignedoli, D. Passerone and A. Baiker, *Journal of Chemical Physics*, 2008, **129**, 234703.

(242) J. L. Da Silva, C. Stampfl and M. Scheffler, *Surface Science*, 2006, **600**, 703–715.

(243) S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. Probert, K. Refson and M. C. Payne, *Zeitschrift fur Kristallographie*, 2005, **220**, 567–570.

(244) V. Fiorentini and M. Methfessel, *Journal of Physics Condensed Matter*, 1996, **8**, 6525–6529.

(245) C. K. Skylaris and P. D. Haynes, *Journal of Chemical Physics*, 2007, **127**, 164712.

(246) U. W. Hamm, D. Kramer, R. S. Zhai and D. M. Kolb, *Journal of Electroanalytical Chemistry*, 1996, **414**, 85–89.

(247) V. K. Kumikov and K. B. Khokonov, *Journal of Applied Physics*, 1983, **54**, 1346–1350.

(248) F. Gossenberger, T. Roman and A. Groß, *Electrochimica Acta*, 2016, **216**, 152–159.

(249) N. Singh, M. S. Lee, S. A. Akhade, G. Cheng, D. M. Camaioni, O. Y. Gutiérrez, V. A. Glezakou, R. Rousseau, J. A. Lercher and C. T. Campbell, *ACS Catalysis*, 2019, **9**, 1120–1128.

(250) X. Chen, I. T. McCrum, K. A. Schwarz, M. J. Janik and M. T. Koper, *Angewandte Chemie - International Edition*, 2017, **56**, 15025–15029.