A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/179633
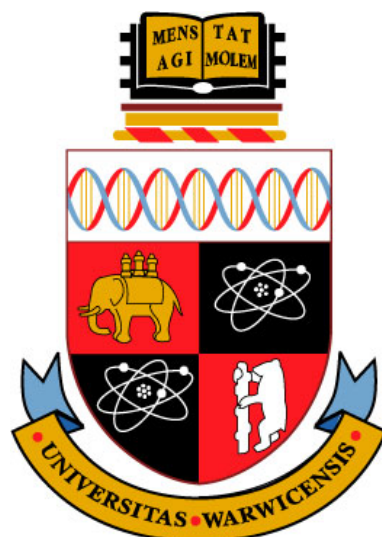
warwick.ac.uk/lib-publications

Computational Simulation of Metal Nucleation on Diamond Electrodes

by

Shayantan Chaudhuri

Thesis

Submitted to the University of Warwick

in partial fulfilment of the requirements

for admission to the degree of

Doctor of Philosophy in Chemistry

Department of Chemistry

December 2022

THE UNIVERSITY OF
WARWICK

Contents

List of Tables

List of Figures

xiii

xiv

Acknowledgments

Firstly, I would like to thank my supervisor, Dr. Reinhard Maurer, for providing me with the opportunity to conduct my PhD in his group, and for all his advice, guidance and reassurance. Your expertise and patience have been crucial to my development and you have been a wonderful mentor. Also, many thanks to the members of the Maurer Group, both new and old, for all their help, banter and company. I would also like to thank Dr. Andrew Logsdail and Dr. Julie Macpherson, both of whom have been secondary supervisors to me and provided me with a lot of advice and direction.

This thesis would not have been possible without the constant support, kindness and counsel I received from my fiancée Gurneet Dhanoa. I cannot express how grateful I am for everything, though I am sorry for all the conversations about computational chemistry that you had to endure! I am super excited for us to get married soon, hopefully both as doctorates!

I would also like to thank my grandfathers, Biswanath Chatterjee and Arun Chaudhuri for being role model figures and for inspiring me both personally and professionally. And finally, last but by no means least, my eternal gratitude to my parents Ruma and Anjan Chaudhuri. মা and বাবা, this would not have been possible without your blessings, and I hope this thesis goes some way toward reflecting everything you have done for me.

Declarations

This thesis is submitted to the University of Warwick in support for the degree of Doctor of Philosophy. It has been composed by the author and has not been submitted in any previous application for any degree. All work presented has been carried out by the author apart from work carried out by collaborators, which have been explicitly stated at the start of each chapter and referenced in the main text.

Publications

Parts of the work within this thesis have been previously published by the author in the following publications (listed in reverse chronological order):

- J. Westermayr, S. Chaudhuri, A. Jeindl, O. T. Hofmann, R. J. Maurer (2022) 'Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic–inorganic interfaces' *Digital Discovery* **1**: 463-475

- S. Chaudhuri, S. J. Hall, B. P. Klein, M. Walker, A. J. Logsdail, J. V. Macpherson, R. J. Maurer (2022) 'Coexistence of carbonyl and ether groups on oxygen-terminated (110)-oriented diamond surfaces' *Commun. Mater.* **3**: 6

Other publications by the author during his doctoral studies but not included within this thesis are listed below (in reverse chronological order):

- G. F. Wood, I. M. Terrero Rodríguez, J. J. Tully, S. Chaudhuri, J. V. Macpherson (2021) 'Electrochemical Ozone Generation Using Compacted High Pres-

sure High Temperature Synthesized Boron Doped Diamond Microparticle Electrodes' *J. Electrochem. Soc.* **168**: 126514

- D.-Q. Liu, M. Kang, D. Perry, C.-H. Chen, G. West, X. Xia, <u>S. Chaudhuri</u>, Z. P. L. Laker, N. R. Wilson, G. N. Meloni, M. M. Melander, R. J. Maurer, P. R. Unwin (2021) 'Adiabatic versus non-adiabatic electron transfer at 2D electrode materials' *Nat. Commun.* **12**: 7110

Grants

This thesis was typeset with LaTeX 2_ε, which is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The style package *warwickthesis* was used by the author.

Abstract

Metal nanocluster deposition is an important fabrication process that can be used to grow nanoarchitectures for use in various applications such as electronics and electrocatalysis. Classical nucleation theory is a powerful theory that can be used to qualitatively capture nucleation thermodynamics for many systems, but has been found to be inappropriate to describe the initial stages of nucleation due to its inherent assumptions. Despite the development of atomistic nucleation theories, much still remains unclear about the early stages of metal deposition and the role of the atomic-scale structure on the surface. In this regard, atomistic simulations based on electronic structure methods can play an important role in the elucidation of the initial nucleation processes and mechanisms. This thesis will use density functional theory and its derived methods to characterise the adsorption of gold nanoclusters on polycrystalline boron-doped diamond surfaces. First, a detailed investigation into the most stable oxygenation state of diamond (110) surfaces is conducted, and the most stable surface phase is shown to comprise coexistent and adjacent carbonyl and ether functional groups. Afterwards, the structural stability of single gold atoms on oxygen-terminated diamond (110) surfaces is investigated, and defects and dopants within the diamond surface are shown to significantly increase the adsorption energy and diffusion barriers of single gold atoms. The atom-by-atom growth of gold dimers, trimers and tetramers on diamond surfaces is subsequently studied by analysing their stabilities and identifying preferred morphologies. Finally, machine learning-based interatomic potentials are developed to facilitate accurate and computationally efficient geometry optimisations, and are used to predict the structures and stabilities of larger gold nanoclusters ranging from 6 to 147 atoms. This thesis is part of a scientific effort to develop modern atomistic theories of atom-by-atom particle growth, and will help guide the controlled design of nanostructured catalysts in the future.

Abbreviations

Abbreviations are listed in alphabetical order:

BDD: Boron-doped diamond

CVD: Chemical vapour deposition

DFA: Density-functional approximation

DFT: Density functional theory

DFTB: Density functional tight-binding

EM: Electron microscopy

GGA: Generalised gradient approximation

HGGA: Hybrid generalised gradient approximation

LDA: Local-density approximation

MAE: Mean absolute error

MBD: Many-body dispersion

MGGA: Meta-generalised gradient approximation

ML: Machine learning

MLIP: Machine-learned interatomic potential

MM: Molecular mechanics

NAO: Numeric atomic orbital

NC: Nanocluster

NP: Nanoparticle

QM: Quantum mechanics

QM/MM: Quantum mechanics/molecular mechanics

RMSD: Root-mean-square deviation

SCOV: Saturated carbonyl oxygen vacancy

STEM: Scanning transmission electron microscopy

TEM: Transmission electron microscopy

TS: Tkatchenko-Scheffler

vdW: van der Waals

XC: Exchange-correlation

XPS: X-ray photoelectron spectroscopy

Chapter 1

Introduction

1.1 Electrodeposition of Metal Nanostructures

Electrodeposition is the formation of solid structures on the surface of an electrode after an electrochemical potential is applied. Electrodeposition is a viable nanofabrication process alongside more established methods such as nanoimprint lithography [1; 2], pH-driven precipitation [3; 4] and directed assembly [5; 6; 7; 8; 9; 10; 11; 12], and can be used to create metal nanostructures for a variety of applications such as carbon dioxide reduction catalysis [13], water splitting [14], fuel cell [15] and materials for energy storage and conversion.

Metal nanoparticle (NP) electrodeposition is an important method for the growth of thin films and nanoarchitectures, which has prospective applications in numerous fields such as electronics, electrocatalysis and quantum computing. Metal NPs have been shown to exhibit unique electrocatalytic properties due to finite size effects [16; 17], while bimetallic NPs have been shown to have greater electrocatalytic activity than monometallic NPs due to geometric effects [17; 18; 19]. The key challenge in electrodeposition is to be able to control the structure, size, and stability of surface-adsorbed NPs on an atomistic scale, which in turn define the reactivity and electrochemical properties of the resulting materials [20; 21]. The establishment of an atomic-scale understanding of the electrodeposition process is thus crucial to enable the controlled fabrication of NPs and other nanostructures.

The process of electrodeposition is often described in four stages: diffusion through the solvent, adsorption at the electrode via electron-transfer, surface migration and nucleation, as depicted in Figure 1.1.
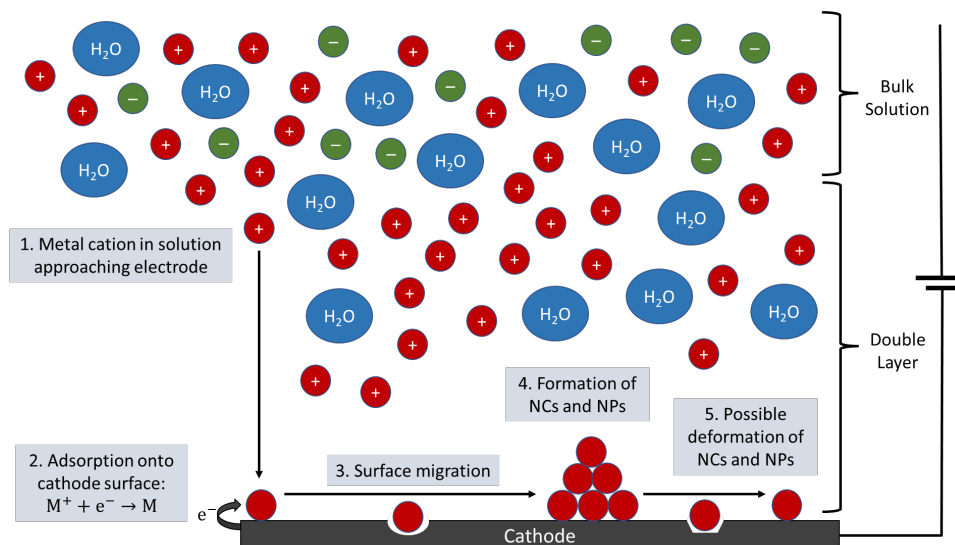
Figure 1.1. A simplified diagram showing the process of metal electrodeposition onto a cathode surface, from the nucleation of a single atom to the growth of a larger structure with possible deformation. The system can be split into two parts: the bulk system, consisting of water (H_2O) molecules (blue), anions (green) and metal cations (red), and the double layer, where charge separation occurs. When a metal cation, M^+, in the double layer gets close enough to the cathode surface (step 1), an electron-transfer reaction occurs that reduces it to a metal atom, M i.e. adsorption (step 2). Once adsorption has occurred, the adatom may migrate along the cathode surface (step 3). This surface migration can lead to adatoms coalescing, resulting in the formation of larger structures (step 4) such as nanoclusters (NCs) and nanoparticles (NPs). If energetically unfavourable, NCs and NPs may fragment (step 5) and lead to further atomic migration along the electrode surface, which may feed the growth of other NCs and NPs on the surface. Isolated atoms may also directly adsorb onto or migrate into defect sites on the electrode surface.

Experimental techniques to track and characterise the metal electrodeposition process have vastly improved over recent years, as discussed in Section 1.1.1. The complementary use of scanning transmission electron microscopy (STEM), surface spectroscopy, and electrochemical analysis provides unprecedented resolution at the nanoscale and, to a more limited extent, resolution in the time domain [22]. Simultaneously, significant advancements have been made in molecular simulation capabilities and the electronic structure theory of complex interfaces [23; 24], as summarised in Section 1.1.2. Yet a large gap between realistic applications, theoretical understanding of dynamics, and atomistic simulation remains. Both theory and experiment face challenges when it comes to bridging this gap and reaching an atomic-level understanding of electrodeposition: theoretical and simulation studies must be able to generate realistic models capable of replicating experimental

conditions, including factors such as surface heterogeneity. On the other hand, experimental studies should ideally be conducted under well-defined and idealised conditions (e.g. atomically-flat electrode interfaces) to allow for atomistic simulations and theoretical analyses to be applied [17]. The synergy between experiment and simulation has the potential to deeply enrich the field: modelling methods can be refined once information about atomic structure is attained from experiment, while simulations can also be used to make predictions that can be validated using experiments [17].

1.1.1 Experimental Methodologies

A variety of microscopy techniques have been used to study metal electrodeposition. Two such techniques are electron microscopy (EM) and scanning probe microscopy. EM is able to attain high magnification and resolution due to the use of electrons over light waves, which can therefore be used to analyse structures in immense detail and generate high-quality images [25]. Disadvantages of EM include the large equipment size, the cost of operation and the long training time required to ensure proper operation; furthermore, the high kinetic energy of the beamed incident electrons can lead to sample degradation [25]. One variant of EM is scanning EM, where the electron beam is focused to a spot and is scanned sequentially across the sample [25]. At each location, scattered electrons are emitted from the sample and are collected by detectors; electrons are typically beamed at energies of 1–30 keV [25]. In transmission EM (TEM), the electron beam is focused on a defined area of the sample [25]. Electrons are typically beamed at higher energies than in scanning EM (80–300 keV) to ensure they are transmitted through the sample; this transmitted signal is then collected by parallel detectors below the sample [25]. A hybrid form of the two methods, STEM, also exists, where scanning electron microscopes are fitted with detectors to collect transmitted electrons [25]. STEM techniques have been a popular choice to study metal electrodeposition due to the high resolution they offer [22; 26; 27; 28; 29]. STEM has been shown to be capable of dynamically visualising the early stages of electronucleation for metals such as gold, with structural resolution on the atomic scale and time resolution defined by the sequential analysis of short electrodeposition runs (several milliseconds of deposition per run) [22]. TEM has also been used to study the electrodeposition of metals such as lithium, silver, nickel and platinum at submicroscopic resolutions [26; 27; 28; 29; 30; 31]. Several studies also report the use of scanning EM to investigate metal electrodeposition [28; 32; 33]. While liquid cell TEM has made lots of progress in monitoring dynamic electrochemical systems [34; 35], it has limited resolution due to factors

3

such as electron-beam-induced gas bubble formation and electron scattering in the liquid [22]; in contrast, *ex situ* aberration-corrected STEM is not only capable of resolving single atoms but can quantify the number of atoms within a particle [22; 36]. Despite the high resolution from STEM, the resulting two-dimensional (2D) projection does not allow for a full assessment of the 3D structure of surface-adsorbed NPs. Some machine learning (ML) and neural network (NN) methodologies have been used to tackle this problem by analysing experimental data and extracting 3D structural data [37; 38; 39; 40; 41].

In contrast, scanning probe methods such as scanning tunnelling microscopy and atomic force microscopy generate images of surfaces using a physical probe that scans the sample [42]. The advantages of scanning probe microscopy over EM include the capability to extract three-dimensional (3D) topographic information at atomistic scales and the sensitivity of the different environments it can operate inmethodshowevertypically require much longer scan times than EM methods [42]. Both atomic force microscopy [43] and scanning tunnelling microscopy [44; 45; 46; 47; 48; 49; 50] are capable of analysing the influence of current on the electrodeposited structure at submicroscopic resolutions. Scanning electrochemical cell microscopy has also been used to study the initial electronucleation stages and mobility of metals such as platinum [51], copper [52], iron [53] and silver [54], and has gained much attention [55; 56; 57] due to its ability to routinely operate at submicroscopic scales [52; 58; 59]. By mapping electrochemical phenomena directly and locally via a nanoscopic meniscus cell, scanning electrochemical cell microscopy can be used to make measurements over a large array of discrete areas over an electrode surface [52]. Comparing microscopy techniques, the resolution obtained using scanning probe techniques such as atomic force microscopy and scanning electrochemical cell microscopy is typically limited to the NP level though, and while scanning tunnelling microscopy can resolve individual atoms, studies typically focus on 2D growth and dissolution [22; 44]. Furthermore, while all microscopy techniques provide information on geometrical and electronic structure to some degree, scanning electrochemical cell microscopy can provide information about reactivity.

Other microscopy techniques such as surface plasmon resonance microscopy [60] and dark-field scattering microscopy [61; 62] also exist have been used to investigate metal electrodeposition. Surface plasmon resonance microscopy is a non-intrusive optical technique that is highly sensitive to nanoscopic objects, while dark-field scattering microscopy has a relatively simple experimental setup that can directly

probe the plasmonic properties of individual structures and allow for correlation with EM [61; 62]. Their performancehoweveris restricted by the very small field of view which increases the difficulty in acquiring quantitative data on electronucleation [60; 61]. Wide-field surface plasmon resonance microscopyhoweverremoves this constraint and allows for the growth of hundreds of nuclei to be tracked simultaneously at a reasonable time resolution ($\sim 1\,$s) [60]. Dark-field scattering microscopy also requires working electrode materials to be optically-transparent and for the nanostructure being analysed to at least include a plasmonic metal; furthermore, the structures have to be large enough to ensure detection [62].

1.1.2   Simulation Methodologies

Electrodeposition comprises many different aspects that must be accounted for in computational simulations, such as the motion of atoms, interactions and electronic structure at the interface, electrolyte and ions, and the electrochemical potential itself. Various computational approaches exist and have been used to model these aspects, including classical molecular dynamics [63] and Monte Carlo methods [64] (for atomic motion), quantum mechanical (QM) electronic structure methods such as density functional theory (DFT) [65; 66] or tight-binding methods (for surface interactions), solvation methods (for the electrolyte and ions) and grand-canonical ensemble simulations [23; 67] (to account for the electrochemical potential).

DFT is the most commonly used *ab initio* electronic structure methodto describe extended surfaces [68] and materials [69], with numerous software packages containing DFT implementations [70; 71; 72; 73; 74; 75; 76; 77]. Semi-empirical tight-binding methods [78; 79; 80; 81] often a good compromise between computational cost and accuracy, have also been used to explore metal electrodeposition [82] the structure of metallic NCs [83]. a few reliable parameterisations currently exist for metal-organic interfaces [84; 85; 86], which is why DFT is typically more popular than tight-binding methods. The extended surface model and the choice of density-functional approximation (DFA) within DFT, discussed more in Section 2.1.4, need to correctly account for the rich diversity of interactions that are present in such surfaces. This includes interactions between metals and non-metals, long-range dispersion interactions, and long-range electrostatic interactions of charged species [68; 69; 87]. Increasingly accurate DFAs are being developed that can represent the energetics and electronic structure of complex materials [68; 69; 88; 89], however the intrinsic computational scalability of DFT provides a challenge for systems comprising more than a few hundred atoms. To ensure an

efficient description of such interfaces requires a pragmatic selection of well-tested DFAs that balance computational efficiency and predictive accuracy.

The large configuration space of possible NP structures at a particular electrode surface must also be considered. Various approaches such as ML [90; 91; 92], genetic algorithms [93; 94; 95; 96], cluster regularisation [97] and global optimisation methods such as basin- [98; 99] and minima-hopping [100], can automatically explore the configuration space, and the structure and stability of NPs at surfaces, not many of these approaches have been applied in the context of electrodeposition [101]. Genetic algorithms can be used to perform direct searches within the solution space, while ML can be used to extend existing solution spaces by recognising existing patterns from training datasets. However, it is important to note that while methods such as ML can provide a lot of important information, even if trained on high-level *ab initio data*, they can only provide approximations to the physical characteristics of a system, which can only be fully captured through experiments [102].

1.1.3   Nucleation Theories

Theoretical considerations of electrodeposition in literature study two processes: the movement and adsorption of solvated ions and electron-transfer reactions, and the process of electronucleation. When it comes to modelling the process of electronucleation, both classical and atomistic theories exist to describe the formation of stable nuclei [22; 103; 104; 105]. Classical nucleation theory relies on macroscopic physical quantities that are applicable to sufficiently large clusters such that their size$n$ (number of atoms)can be considered a continuous variable. In this case, the free energy of nucleation, $\Delta \overline{G}(n)$, is differentiable [104; 105], which allows for the prediction of the nucleation rate [103], as expressed in Equation (1.1).

$$\varrho = NZJ \exp\left(- \frac{\Delta \overline{G}}{k_{\mathrm{B}}T} \right), \tag{1.1}$$

where ϱ is the nucleation rate, N is the number of nucleation sites, Z is the Zeldovich factor, J is the rate at which atoms attach to the nucleus, k_{B} is the Boltzmann constant and T is the temperature [103].

When it comes to experimental electrochemical measurements, techniques such as chronoamperometry provide mostly macroscopic information, from which nanoscopic behaviour such as nucleation rates can be inferred [22; 106; 107]. Such inferenceshoweverhave been found to be inappropriate to describe the initial stages

of nucleation where individual atoms and few-atom clusters are present [22; 54; 108; 109; 110; 111; 112] due to the assumptions within classical nucleation theory. Classical nucleation theory assumes the holding of the capillarity approximation, which treats the nucleus interior as a bulk incompressible liquid and assumes that the surface tension of a small liquid droplet is equal to the surface tension of a flat surface [104; 105; 113]. This has been shown to break down for small systems [114; 115; 116]. Furthermore, clusters are assumed to either grow or shrink via single-particle absorption or emission respectively, which places kinetic restrictions on the nucleation pathways [104; 105; 113]. This does not hold in reality as entire clusters can merge or fragment, and these kinetic pathways cannot be ignored. While improvements to classical nucleation theory do exist, such as dynamical nucleation theory [117], mean-field kinetic nucleation theory [118], coupled flux theory [119; 120; 121; 122] and diffuse interface nucleation theory [123; 124], these have mostly been applied to describing the condensation of supersaturated vapours into the liquid phase and crystal nucleation studies rather than investigating metal electronucleation [113; 125]. Despite its shortcomings, classical nucleation theory is still a powerful theory and has been shown to be capable of qualitatively capturing nucleation thermodynamics and kinetics for many systems [113].

In contrast, atomistic nucleation theory can be applied to clusters so small such that n is no longer continuous, as is the case with first-order phase transitions at high supersaturation levels resulting in $\Delta\overline{G}(n)$ being non-differentiable [104; 105]. Figure 1.2 shows the intrinsic differences between the two theories as well as the (dis)continuity of cluster size. Atomistic theories allow for high $\Delta\overline{\mu}$ levels to be modelled and have been validated against experimental studies [105]. Despite the existence of such theories, much remains unclear regarding the initial stages of electronucleation and the role of the atomic-scale structure of the electrode. In this regard, explicit atomistic simulations can play an important role in elucidating the initial processes and mechanisms.
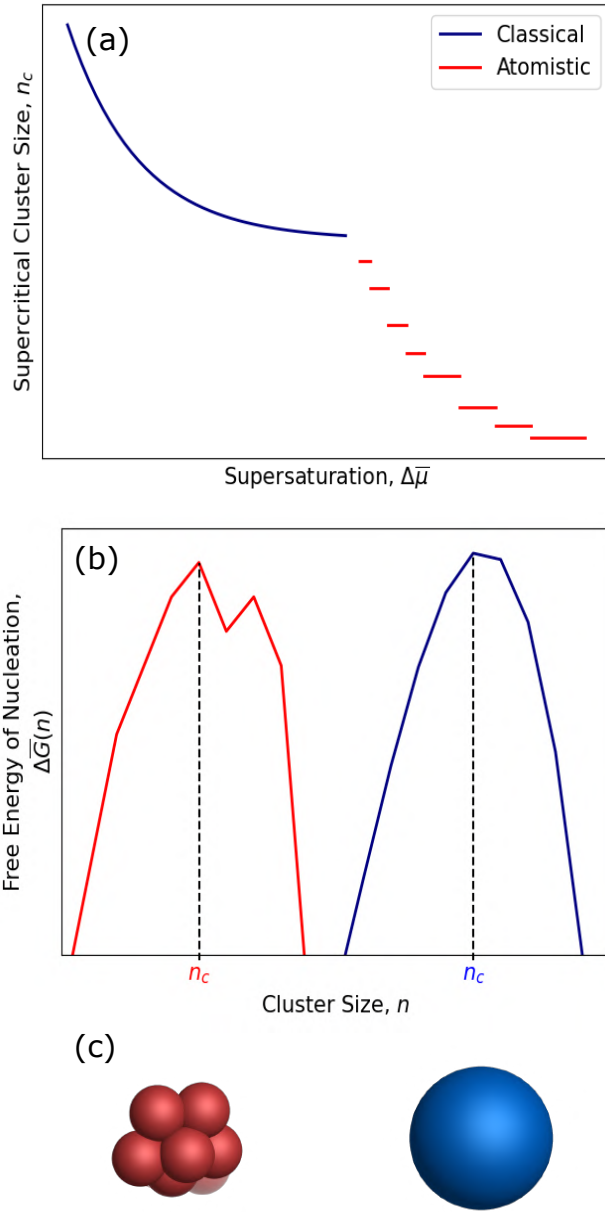
Figure 1.2. Schematics showing the differences between classical (blue) and atomistic (red) nucleation theories. Shown in (a) is the relationship between supersaturation, $\Delta\overline{\mu}$, and the supercritical cluster size (n_c atoms) according to both theories. In classical/atomistic nucleation theory, $\Delta\overline{\mu}$ is low/high enough such that n_c is continuous/discontinuous. Shown in (b) is the relationship between the free energy of nucleation, , and the cluster size of a cluster (n atoms) for a constant $\Delta\overline{\mu}$. In classical/atomistic theories, $\Delta\overline{\mu}$ is small/large enough to result in $\Delta\overline{G}(n)$ being differentiable/non-differentiable. In both theories, $n = n_c$ [105] is the size at which $\Delta\overline{G}(n)$ has a global maximum. Shown in (c) are visualisations of both theories treat clusters: atomistic theories are applicable for small clusters, whereas classical theories can be used for large clusters which can be treated spherically.

Once , may migrate along the electrode surface and coalesce to form metastable nanoclusters (NCs). In order for NPs to form, these NCs will rearrange to form crystalline NPs; closely-spaced NCs may also dissemble and feed atoms into existing NPs. Both for the formation of NCs and NPs, atoms will migrate along the electrode surface. In short, the on-surface dynamics of electronucleation can be extremely rich and complex. The nucleation and growth of NPs on electrode surfaces involves a number of factors including thermodynamics and kinetics, which define the NC size distribution, growth rate, and the rate-determining steps [126]. Kinetics can be determined using transition state theory, for which the energy barriers between given initial and final states need to be calculated. This can be done via minimum energy path methods such as nudged elastic band [127] to determine the energy required to move isolated metal atoms over an electrode surface [22].

A full simulation pipelinethusneeds to be established that is both efficient and can capture all the relevant effects for metal NC growth. Furthermore, it is prudent that an understanding is gained as to how nucleation is affected by variables such as surface defects, and if certain atomistic details govern nucleation in certain systems and conditions. Once has been established, optimisation of the conditions in which nucleation is kinetically controlled by on-surface processes, such that NC and NP growth can be selectively controlled by surface modification and subtle parameter tuning, can be attempted. Understanding the factors that affect the electrodeposition and electronucleation of metals, and possessing the ability to control their growth on an atomistic scale, has the potential to increase the efficiency of many technologies and applications and is therefore of paramount importance.

1.2 Diamond as a Material

Diamond is a solid allotrope of carbon that possesses numerous extreme properties, such as the highest hardness [128], thermal conductivity [129] and Debye temperature [130] of any material. Diamond also possesses a wide band gap, which results in a high electrical resistivity [131]. These properties have been exploited across a large variety of fields including electrochemistry [21; 131], photonics [132; 133], and quantum computing [134; 135; 136].

Diamond is only a metastable allotrope of carbon at room temperature and pressure, with graphite being the most stable instead [137]. However, the graphitisation rate of diamond is almost negligible under standard conditions, which means

9

diamond and graphite can coexist. Natural terrestrial diamonds typically form deep within the mantle of the Earth [138], where pressures and temperatures are high enough such that diamond becomes [137]. atural diamonds normally contain many contaminants and are therefore unfit for most industrial applicationsynthetic diamonds are typically used for technological applications. Synthetic diamonds are used either as single crystals, where the crystal arrangement is continuous, or as a polycrystalline material, where grain boundaries exist between crystal arrangements of different orientations. However, it is typically much more difficult [139; 140] and expensive to grow single-crystal diamonds. For large-area applications, it is much cheaper to grow polycrystalline diamond than single-crystal diamond, though other types such as nanocrystalline and microcrystalline diamonds have also found use [141]. Polycrystalline diamond has numerous industrial applications, including thermal management [142] and electrochemistry [143]. Synthetic diamonds can be grown using two methodologies, which are briefly summarised in Section 1.2.1.

1.2.1 Diamond Synthesis

For many technological applications, the most common approach to growing diamond is via the process of chemical vapour deposition (CVD) [144]. With CVD growth, it is relatively easy to control impurity content and grow large-area samples [145], hence is commonly used to grow samples for industrial and technological applications. CVD occurs in the region where diamond is metastable compared to graphite, and is therefore driven by kinetics rather than thermodynamics [145]. CVD is typically performed using a small fraction of a carbonaceous gas such as methane in an excess of molecular hydrogen [144; 145]. The excess hydrogen is crucial as it etches away graphite but not diamond, and helps to stabilise the diamond surface [145]. The gas mixture is then thermally activated, and this heating can be done using various methods, though a hot filament or microwave plasma are the most commonly used [145]. With single-crystal diamond, growth of the required crystal face is achieved by using a single crystal substrate of the same orientation for epitaxial growth [140] or by laser cutting along the required axis post-growth. As the diamond grows thicker, or the growth substrate moves away from atomic smoothness, achieving monocrystallinity becomes more challenging [139; 140]. Other substrates such as silicon, molybdenum or tungsten [145] can be used to grow polycrystalline diamond.

In contrast , high pressure high temperature [146] seeks to, in principle, reproduce the conditions that result in the growth of natural diamonds with some practical adjustments [145]. This method of synthesis operates within the regime

where diamond is thermodynamically stable and allows control over the geometry of the sample, though it is more difficult than CVD when it comes to controlling impurity content. igh pressure high temperature-grown diamonds typically containing a few hundred parts per million of single substitutional nitrogen dopants [147]. For this reason, CVD is typically preferred for growing crystals for technological applications [148].

1.2.2 Diamond Crystallography

Within a diamond crystal, each carbon atom is covalently bonded to four other carbon atoms via sp^3 hybridisation in a tetrahedral manner. Diamond has a cubic crystal structure with eight atoms per unit cell. The unit cell can be thought of as the amalgamation of two face-centred cubic lattices, as can be seen in Figure 1.3. To derive the positions of the eight atoms within the full unit cell, consider an enlargement of the unit cell by a factor of 4. Then, atoms within the cell lie at positions $(x, y, z) \in \mathbb{Z}^3$ such that $x \equiv y \equiv z \pmod 2$ [149]. For the first face-centred cubic, positions must also satisfy the condition $x + y + z \equiv 0 \pmod 4$ [149], to which there are four solutions (modulo 4): $(0, 0, 0), (0, 2, 2), (2, 0, 2), (2, 2, 0)$. For the second face-centred cubic, atomic positions must satisfy the aforementioned $x \equiv y \equiv z \pmod 2$ condition, as well as $x + y + z \equiv 1 \pmod 4$ [149]. Once again, there are four solutions to this congruence relation (modulo 4), which are $(3, 3, 3), (3, 1, 1), (1, 3, 1)$ and $(1, 1, 3)$. Reducing this down to a unit cell, given any position (x_1, y_1, z_1), its nearest atom can be found at $(x_1 \pm 1/4, y_1 \pm 1/4, z_1 \pm 1/4)$, in fractional coordinates. For this reason, $\{(0,0,0), (1/4, 1/4, 1/4)\}$ is considered to be the two-atom *motif* for atomic positions within the diamond unit cell.
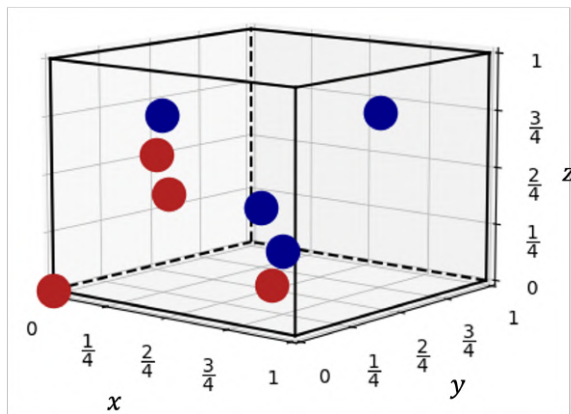
Figure 1.3. Atomic positions within a diamond unit cell, expressed as fractional coordinates. The unit cell can be thought of as the amalgamation of two face-centred cubics, with their atoms shown in red and blue respectively.

When diamond crystals grow, either naturally or synthetically, they can do so in any direction. However, there are three primary crystallographic orientations that dominate crystal growth, which are (111), (100) and (110) (in Miller index notation). Each crystallographic orientation has a different surface arrangement of carbon atoms, and Figure 1.4 visualises the atomic structure of these facets. Of these orientations, the (110) facet typically has the fastest growth rate [139; 150; 151] and thus serves as a useful model of the more reactive sites during diamond growth [152]. The faster growth rate of the (110) facet can be explained by comparing the different growth mechanisms of the different facets [151]. (110) faces are stepped, where the nucleation process is one-dimensional as these faces grow by a unidirectional addition of atoms [139; 151]. (111) faces, however, are flat and grow via a layer-by-layer mechanism [139; 151]. The nucleation process is therefore two-dimensionalwhich slows the growth kinetics but makes it easier to grow atomically smooth surfaces [139; 151]. On the other hand, (100) faces are kinked and have a very small nucleation barrier [139; 151]. Growth of such faces typically occurs via a random incorporation of atoms and tend to be the fastest-growing [139; 151]. However, for many materials such as silicon, germanium, gallium arsenide and diamond, their (100) surfaces have been shown to typically (2×1)-reconstruct, which confers a flat-like character to the face and hence slows their growth rate [139; 153].
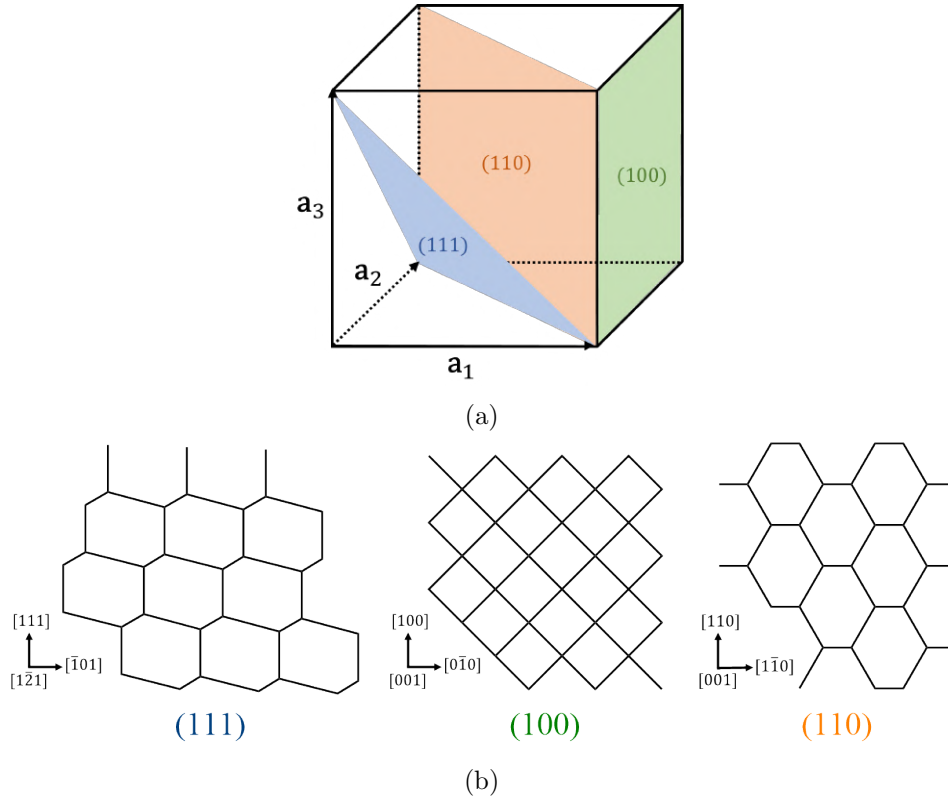
Figure 1.4. three primary crystallographic facets of diamond, shown (a) as planes within a cubic crystal with orthogonal lattice vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, and (b) as skeletal formulae from a 'front-on' view with axes presented.

The (110) surface possesses C_2 symmetry [152], which means the surface remains identical after a 180° rotation. Every surface carbon atom is bonded to three other carbon atoms and therefore has one dangling bond. (110)-oriented single crystals have applications in fields like photonics [154], though these are much more difficult to grow as the crystal face decreases in size as the crystal gets thicker [155]. Furthermore, the occurrence of (110) faces during single-crystal growth can lead to a partial break-up of the crystal beyond a certain thickness [150]. In contrast, the slower growth rates of the (111) and (100) facets [139; 150; 151] make it much easier to grow single-crystal samples of such orientations. For this reason, many single-crystal diamond studies tend to use (111)- or (100)-oriented samples, hence why the (110) facet has not been as well studied in literature.

In polycrystalline material, the (110) facets can be revealed by mechanically polishing to a surface roughness where the surface is predominantly (110)-textured. This has been experimentally demonstrated using electron backscatter

diffraction [156] and STEM [22]. Such surfaces have found practical use in electro-chemical applications, where the polycrystalline diamond material can be treated as a textured surface with a dominant (110) orientation [22; 156].

1.2.3 Boron-Doped Diamond

Other elements are frequently added into diamond as impurities to modify the properties exhibited by the material [131]. Substitutional boron is often added Boron acts as a p-type dopant in diamond and bestows diamond with semiconductive properties [21; 131; 157]. Boron is a triel with three valence electrons, whereas carbon atoms have four valence electrons. Therefore, if a boron atom displaces a carbon atom in diamond, the valence electrons of boron form covalent bonds with three of its carbon neighbours, but no bond is formed with its fourth neighbour, thus introducing a formal +1 charge within the diamond.

Typically only one atom in a thousand carbon atoms needs to be replaced by boron for diamond to exhibit semi-metallic properties [131]. Boron-doped diamond (BDD), typically in polycrystalline form, has thus found great use in electrochemistry as a working electrode material [21; 131]. BDD has a very wide solvent window, low background and capacitive currents, as well as reduced fouling as compared to other electrode materials [131], which makes it ideal for a variety of applications such as sensing, waste water treatment and chemical disinfection [158]. Along with its electrical conductivity, BDD is highly stable [21; 131; 143; 157; 158] and can be used as a substrate for STEM measurements [22]. It is therefore an attractive substrate to study metal electrodeposition on, and has a lot of potential as a support for metal nanostructures for purposes such as electrocatalysis.

1.3 Gold as a Material

Gold is a precious, noble metal that possesses many unique properties, such as the highest malleability [159] and nobleness [160] of all metals. Furthermore, gold nanoclusters are known to adopt unique electronic and geometric structures [161; 162]. The high electrocatalytic activity of monometallic [163; 164; 165; 166; 167; 168], bimetallic [169; 170; 171; 172; 173; 174; 175; 176; 177; 178; 179] and multimetallic [180; 181; 182] gold-based nanostructures has been well-established in literature. Recently, supported single atom electrocatalysts [183; 184; 185] have also grown in popularity due to their optimal atom utilisation and unique electronic properties, often outperform supported metal nanostructures. Single gold atoms, supported by

materials such as copper [186], magnesium oxide [187; 188; 189; 190], cerium(IV) oxide [190; 191; 192; 193; 194; 195] and graphene/graphite [196; 197; 198; 199; 200; 201; 202], have been shown to be very efficient electrocatalysts for a variety of key chemical processes, including nitrogen reduction [203; 204; 205] and oxygen reduction and evolution [206].

The formation of gold nanostructures on BDD [22; 207] and their combined electroanalytical applications [208; 209; 210; 211; 212; 213; 214] have been investigated extensively, though only one study considered their combined electrocatalytic ability [164]. a lot of potential for BDD-supported gold-based nanomaterials fabricated via electrodeposition to be used for electrocatalytic purposes. One particular study of note is Hussein *et al.*, where STEM was used to track the electrodeposition dynamics of gold on polycrystalline BDD surfaces, from the nucleation and growth of a single atom to a crystalline NP [22]. Here, gold atoms were found to be stable atop BDDs surface and even withstand the momentum transfer from the highly energetic electron beam [22]. However, the diffusion barriers for single gold atoms were calculated to be very low [22], which indicates that the high stability of gold atoms observed by Hussein *et al.* is likely due to defects and dopants that were not seen within their STEM images or investigated further. Atomistic simulations based on DFT can provide many insights in this regard to elucidate how the atomic-scale structure of the BDD surface influences the adsorption and nucleation of gold NCs.

1.4   Thesis Aim

This thesis will detail an investigation into the initial stages of gold nucleation and aim to characterise the adsorption of gold NCs on polycrystalline BDD surfaces by making use of DFT and its derived methodologies. Chapter 2 will provide a theoretical background to the underlying principles of the various computational methods that were used herein, as well as introduce the various software packages that were used to conduct this work. Chapter 3 will describe a joint computational-experimental investigation into the characterisation of the most stable oxygenation state of the diamond (110) surface. Building on this, Chapter 4 will present an investigation into the adsorption of single gold atoms onto oxygen-terminated diamond (110) surfaces and establish mechanisms that promote the stability of adsorbed single atoms. This will be followed by studying the atom-by-atom growth of supported gold dimers, trimers and tetramers, and identifying energetically preferred structures. Chapter 5 will showcase the development of an MLIP approach that can be

used to facilitate fast and accurate structure searches, and efficiently explore the stabilities of supported gold NC structures of various sizes. Chapter 6 will then present an overall discussion of the work conducted in this thesis and finish with an outlook. This thesis will help contribute to the development of modern atomistic theories of atom-by-atom particle growth and will help guide the controlled design of nanostructured catalysts in the future.

Chapter 2

Computational Methods

2.1 Density Functional Theory

DFT is a method based on quantum mechanics (QM) that has been widely used to study the electronic structure of numerous systems. In the following, the fundamentals of DFT are discussed and how it can be used to calculate the ground-state energy of a many-body system based on its density. Atomic units are used throughout this section.

2.1.1 The Many-Body Schrödinger Equation

The Schrödinger equation can be used to fully describe the many-body behaviour and energy of a system. Equation (2.1) expresses a time-independent and non-relativistic form of the Schrödinger equation, which can be used to describe stationary states.

$$\hat{\mathcal{H}} \left| \Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N, \mathfrak{r}_1, \mathfrak{r}_2, \cdots, \mathfrak{r}_M) \right\rangle = \mathcal{E} \left| \Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N, \mathfrak{r}_1, \mathfrak{r}_2, \cdots, \mathfrak{r}_M) \right\rangle, \quad (2.1)$$

where $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N \in \mathbb{R}^3$ denote the positions of $N \in \mathbb{N}$ electrons, $\mathfrak{r}_1, \mathfrak{r}_2, \cdots, \mathfrak{r}_N \in \mathbb{R}^3$ denote the positions of $M \in \mathbb{N}$ nuclei, Ψ is the many-body wavefunction, $\hat{\mathcal{H}}$ is the Hamiltonian operator and $\mathcal{E}$ is the energy of the system. $\hat{\mathcal{H}}$ can be expressed as the sum of five terms, as shown in Equation (2.2).

$$\hat{\mathcal{H}} = \hat{\mathcal{T}}_{\text{elec}} + \hat{\mathcal{T}}_{\text{nuc}} + \hat{\mathcal{U}}_{\text{elec}} + \hat{\mathcal{U}}_{\text{nuc}} + \hat{\mathcal{V}}_{\text{ext}} \quad (2.2)$$

where $\hat{\mathcal{T}}_{\text{elec}}$ and $\hat{\mathcal{T}}_{\text{nuc}}$ are the electronic and nucleic kinetic energy operators respectively, $\hat{\mathcal{U}}_{\text{elec}}$ and $\hat{\mathcal{U}}_{\text{nuc}}$ are the electron-electron and nucleus-nucleus repulsion operators respectively, and $\hat{\mathcal{V}}_{\text{ext}}$ is the external potential operator which describes

electron-nucleus attractions.

The Born-Oppenheimer approximation [215] assumes that nuclei move much slower than electrons due to their larger sizes and masses. The positions of the nuclei can therefore be assumed to be fixed, and electrons can be assumed to move within a field of nuclei, with electronic positions being parameterised by nucleic positions. Furthermore, this approximation means that $\hat{\mathcal{T}}_{\mathrm{nuc}}$ can be equated to zero and $\hat{\mathcal{U}}_{\mathrm{nuc}}$ can be equated to a constant. Using the Born-Oppenheimer approximation, Equation (2.1) can be reduced to the electronic Schrödinger equation to evaluate the electronic energy, as shown in Equation (2.3).

$$\hat{\mathcal{H}}_{\mathrm{elec}} \left| \Psi_{\mathrm{elec}}(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) \right\rangle = \mathcal{E}_{\mathrm{elec}} \left| \Psi_{\mathrm{elec}}(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) \right\rangle, \tag{2.3}$$

where Ψ_{elec} is the electronic many-body wavefunction, $\hat{\mathcal{H}}_{\mathrm{elec}}$ is the electronic Hamiltonian operator and $\mathcal{E}_{\mathrm{elec}}$ is the electronic energy of the system. $\hat{\mathcal{H}}_{\mathrm{elec}}$ can then be expressed as the sum of three terms, as expressed in Equation (2.4), which are the electronic kinetic energy operator ($\hat{\mathcal{T}}$), the electron-electron repulsion operator ($\hat{\mathcal{U}}$) and the external potential operator ($\hat{\mathcal{V}}$), which describes the attraction between electrons and nuclei. The total energy of a system can then be calculated by summing $\mathcal{E}_{\mathrm{elec}}$ and the now-constant nucleus-nucleus repulsion.

$$\hat{\mathcal{H}}_{\mathrm{elec}} = \hat{\mathcal{T}}_{\mathrm{elec}} + \hat{\mathcal{U}}_{\mathrm{elec}} + \hat{\mathcal{V}}_{\mathrm{ext}} \tag{2.4}$$

While $\hat{\mathcal{T}}_{\mathrm{elec}}$ and $\hat{\mathcal{U}}_{\mathrm{elec}}$ do not change for any N-electron system and are therefore universal operators, $\hat{\mathcal{V}}_{\mathrm{ext}}$ is dependent on the nuclei within the system, which makes it difficult to calculate Ψ_{elec}. Furthermore, it quickly becomes infeasible, if not impossible, to solve Equation (2.3) for systems comprising more than a few electrons as Ψ_{elec} is dependent on $3N$ coordinates, which makes it difficult to store the wavefunction.

2.1.2 Hohenberg-Kohn Theorems

Hohenberg and Kohn [65] were able to overcome the calculation of Ψ_{elec} by expressing $\mathcal{E}_{\mathrm{elec}}$ as a functional of the electron density, $\rho(\mathbf{r})$, as shown in Equation (2.5) via bra-ket notation [216]. The electron density describes the probability distribution of electrons in space and is a function of only 3D space, which makes it much easier to store than the many-body wavefunction.

$$\mathcal{E}_{\mathrm{elec}} = E[\rho] = \langle \Psi[\rho] | \hat{\mathcal{T}}_{\mathrm{elec}} + \hat{\mathcal{U}}_{\mathrm{elec}} + \hat{\mathcal{V}}_{\mathrm{ext}} | \Psi[\rho] \rangle \tag{2.5}$$

Equation (2.5) holds as Hohenberg and Kohn proved that there exists an invertible one-to-one mapping between $\hat{\mathcal{V}}_{\text{ext}}$ and $\rho(\mathbf{r})$ [65]. Then, the external potential can be expressed as a functional of ρ, as shown in Equation (2.6).

$$\hat{\mathcal{V}}_{\text{ext}} = V_{\text{ext}}[\rho] = \int v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) \, d\mathbf{r} \tag{2.6}$$

Using Equation (2.6), Equation (2.5) can be rewritten as Equation (2.7) [65], which is the first Hohenberg-Kohn theorem. The $\langle \Psi[\rho] | \hat{\mathcal{T}}_{\text{elec}} + \hat{\mathcal{U}}_{\text{elec}} | \Psi[\rho] \rangle$ term in Equation (2.5) is referred to as the Hohenberg-Kohn functional, $F_{\text{HK}}[\rho]$.

$$\mathcal{E}_{\text{elec}} = E[\rho] = F_{\text{HK}}[\rho] + \int v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) \, d\mathbf{r} \tag{2.7}$$

The second Hohenberg-Kohn theorem, as summarised in Equation (2.8), states that the Hohenberg-Kohn functional will yield the ground-state energy, $\mathcal{E}_0$, of a system if and only if $\rho = \rho_0$, where ρ_0 is the ground-state density [65]. Furthermore, for any $\rho \neq \rho_0$, $E[\rho]$ would provide an upper bound on $\mathcal{E}_0$ [65].

$$\mathcal{E}_0 = E[\rho_0] \leq E[\rho] \tag{2.8}$$

There is still the question of how to formulate $F_{\text{HK}}[\rho]$ in Equation 2.7, as its constituent operators are not solely density-dependent. However, $F_{\text{HK}}[\rho]$ can be recast as powers of the electronic charge, e^2, as shown in Equations (2.9)–(2.11). The implication between Equations (2.10) and (2.11) follows after equating coefficients $F^{(i)}$. The zeroth-order term can be equated to the kinetic energy of a non-interacting set of electrons, T_s, while the first-order terms correspond to Coulomb and exchange terms that can be found from Hartree-Fock theory, where only the classical Hartree-Coulomb contribution, E_{H}, can be expressed as a simple functional of ρ. All other higher-order terms are grouped within the exchange-correlation functional, E_{XC}. In Equation (2.11), only the $E_{\text{H}}[\rho]$ term is known, while many approximations to E_{XC} exist and are discussed in Section 2.1.4.

$$F_{\text{HK}}[\rho] = \sum_{j=0}^{\infty} \left(e^2\right)^j F_{\text{HK}}{}^{(j)}[\rho] \tag{2.9}$$

$$= F_{\text{HK}}{}^{(0)}[\rho] + e^2 F_{\text{HK}}{}^{(1)}[\rho] + e^4 F_{\text{HK}}{}^{(2)}[\rho] + \cdots \tag{2.10}$$

$$= T_s[\rho] + E_{\text{H}}[\rho] + E_{\text{XC}}[\rho] \tag{2.11}$$

2.1.3 Kohn-Sham Equations

The Hohenberg-Kohn theorems prove that any ρ_0 can be mapped to only one single external potential within that mapping [65]. Trial electron densities must satisfy mathematical conditions such as V-representability, which comprises being normalisable, positive and compliant with the Pauli principle [217], to ensure they can be associated with an external potential. However, any ρ can be connected to two different external potentials via two different mappings: one for interacting electrons and one for non-interacting electrons. Kohn and Sham [66] proposed that the non-interacting ground-state density could first be calculated in a known, fictitious Kohn-Sham potential, V_s, for a non-interacting set of electrons. This ground-state density can then be used to evaluate the ground-state energy of a system [66] as described by V_{ext}. This can be done by introducing fictitious one-electron orbitals, ψ_i, for the non-interacting set of electrons, in turn can be used to construct a Slater determinant for the non-interacting wavefunction [66]. The Kohn-Sham approach allows for T_s to be expressed as Equation (2.12).

$$T_s[\rho] = \sum_{i=1}^{N} \langle \psi_i | -\frac{1}{2}\nabla^2 | \psi_i \rangle \tag{2.12}$$

The Kohn-Sham equations are then a set of one-electron equations that are essentially eigenvalue problemfor each electron i [66], which can be summarised as Equation (2.13).

$$\left(-\frac{1}{2}\nabla^2 + V_s \right)\psi_i = \epsilon_i \psi_i \tag{2.13}$$

where ϵ_i is the Kohn-Sham eigenenergy for electron i. The ground-state density, for both interacting and non-interacting systems, can be expressed as:

$$\rho_0(\mathbf{r}) = \sum_{i=1}^{N} |\langle \mathbf{r} | \psi_i \rangle \langle \psi_i | \mathbf{r} \rangle|^2 \tag{2.14}$$

Using the variational principle, the one-particle V_s potential can be defined as the variational derivatives of the energy functional, as shown in Equation (2.15):

$$V_s[\rho] = V_{\text{ext}}[\rho] + v_{\text{H}}[\rho] + V_{\text{XC}}[\rho] \tag{2.15}$$

2.1.4 Density-Functional Approximations

Many DFAs to the true E_{XC} functional have been developed [218], and these are typically arranged hierarchically depending on the complexity of their formulation

and their accuracy [219]. This hierarchical arrangement of DFAs has been referred to as Jacob's ladder [219], and at its lowest rung lies local-density approximations (LDAs), which assume that E_{XC} depends solely upon the value of ρ at each point in space [69]. The general LDA approximation to E_{XC} is shown in Equation (2.16),

$$E_{XC}[\rho] \approx E_{XC}^{LDA}[\rho] = \int \rho(\mathbf{r}) e_{xc}^{LDA}(\rho) \, d\mathbf{r} \tag{2.16}$$

where $e_{xc}(\rho)$ is the exchange and correlation energy of a homogeneous electron gas at ρ. The exchange component, e_x^{LDA}, for a homogeneous electron gas is known and can be expressed as Equation (2.17), but different numerical parameterisations can be used to calculate the correlation component, e_c, and this is where different LDAs differ [68]. In practice, different LDAs differ only in the functional form and parameterisation, and all typically yield similar results [68].

$$e_x^{LDA}(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \int \rho(\mathbf{r})^{4/3} \, d\mathbf{r} \tag{2.17}$$

LDAs have been shown to generally calculate lattice constants and band structures of simple metals well, but they tend to perform less well for molecules and semi-conductors [68] by overestimating band gaps and overestimating adsorption energies [69]. This occurs because LDAs exhibit an artificial energy minimum between sub-systems, even if they are not covalently or ionically bonded [68]. This was often taken to mimic van der Waals (vdW) interactions, although this overbinding is due to the wrong physical reason [89; 220; 221; 222]. This has been demonstrated for many hybrid inorganic-organic interfaces [220; 223; 224; 225; 226; 227] as well as for intermolecular bond [228], where LDAs result in adsorption distances that are much than experimental data.

The subsequent rung comprises generalised gradient approximations (GGAs), which assume that E_{XC} depends on ρ as well as the density gradient, $\nabla\rho$, as shown in Equation (2.18). GGAs are very popular in condensed matter physics due to their ability to typically produce fairly accurate results without too much computational expense [68; 69] as opposed to more complex approaches.

$$E_{XC}[\rho] \approx E_{XC}^{GGA}[\rho] = \int \rho(\mathbf{r}) e_{xc}^{GGA}(\rho, \nabla\rho) \, d\mathbf{r} \tag{2.18}$$

The $e_{xc}^{GGA}(\rho, \nabla\rho)$ term in Equation (2.18) can be expressed as Equation (2.19), where $\mathcal{F}$ is an enhancement factor and can be split into exchange, $\mathcal{F}_x$, and correlation, $\mathcal{F}_c$,

components.

$$e_{\text{xc}}^{\text{GGA}}(\rho, \nabla\rho) = e_{\text{x}}^{\text{LDA}}(\rho)\mathcal{F}_{\text{xc}}(\rho, \nabla\rho) \tag{2.19}$$

Different GGAs typically differ in their choice of $\mathcal{F}_{\text{xc}}$ formulation; for example, the PBE [229] GGA, which is one of the most popular DFAs [68; 69; 230] and is the main DFA used herein, uses Equation (2.20) to calculate $\mathcal{F}_{\text{x}}$, where s is the reduced density gradient, and κ and χ are constants. Within the PBE GGA, κ is specified to be 0.804 [229]. The PBE-derived revPBE GGA also uses Equation (2.20) to calculate $\mathcal{F}_{\text{x}}$ but softens κ to 1.245 [231]. The PBEsol GGA only differs from the (rev)PBE formulation by reducing the s-dependence of $\mathcal{F}_{\text{x}}$ after reducing χ [232]. The RPBE GGAhoweverpossesses a different mathematical form for $\mathcal{F}_{\text{x}}$ but has the same formulation for all other constituent components [233].

$$\mathcal{F}_{\text{x}}^{\text{PBE}} = 1 + \kappa - \frac{\kappa}{1 + \chi s^2/\kappa}, \tag{2.20}$$

Meta-GGAs (MGGAs) make up the third rung on Jacob's ladder and seek to account for higher-order dependencies by including the density of the Kohn-Sham orbital kinetic energy, t_s, as shown in Equation (2.21) [69]. MGGAs tend to improve upon GGAs and LDAs with more accurate binding energies, along with better cohesive and structural properties [69], but are also typically more computationally costly. The TPSS MGGA uses a similar equation to Equation (2.20) to calculate $\mathcal{F}_{\text{x}}$, but uses a different parameter to χs^2 [234]. he M06-L MGGA also includes the PBE exchange energy density within its formulation for the exchange energy [235]. In contrast, the TPSSloc MGGA uses a localised PBE-like DFA for $\mathcal{F}_{\text{c}}$ within a TPSS-like formulation [236], while the revTPSS MGGA is based on the PBEsol modification to $\mathcal{F}_{\text{c}}$ with respect to PBE [232]. Within the SCAN [237] MGGA, $\mathcal{F}_{\text{x}}$ obeys the conjectured $\mathcal{F}_{\text{x}} \leq 1.174$ bound for all densities [237; 238; 239], while within (rev)TPSS and even PBE, $\mathcal{F}_{\text{x}}$ monotonically tends to the general Lieb-Oxford bound (1.804) [240].

$$E_{\text{XC}}[\rho] \approx E_{\text{XC}}^{\text{MGGA}}[\rho] = \int \rho(\mathbf{r})e_{\text{xc}}^{\text{MGGA}}(\rho, \nabla\rho, t_s)\,\mathrm{d}\mathbf{r} \tag{2.21}$$

Hybrid functionals, the fourth rung along Jacob's ladder, aim to improve upon lower rungs by mixing a certain amount of exact exchange into their formulations using Hartree-Fock theory, which are typically GGA-based (HGGAs) [241; 242; 243; 244] but can also be LDA- [245] or MGGA-based [246]. However, this admixing typically increases the computation effort associated with hybrid functionals by at least one order of magnitude over lower-rung DFAs [68]. Equation (2.22) details

the mixing between exact and GGA-exchange for the calculation of the exchange and correlation density, where a is the mixing parameter and a constant. For both the PBE0 [243] and PBEsol0 [244] HGGAs, $a = 0.25$, and the PBE and PBEsol GGAs are used to calculate $E_{\text{x}}^{\text{GGA}}$ respectively. Range-separated HGGAs such as HSE03 [241] and HSE06 [242] build on this and also use a screening parameter to split $e_{\text{x}}^{\text{GGA}}$ in Equation (2.22) into short- and long-ranged components.

$$e_{\text{xc}}^{\text{HGGA}} = a e_{\text{x}}^{\text{exact}} + (1 - a)e_{\text{x}}^{\text{GGA}} + e_{\text{c}}^{\text{GGA}} \tag{2.22}$$

2.1.5 Basis Sets

Apart from the choice of DFA, another factor that must be taken into consideration when running DFT calculations is the choice of basis set. A basis set is a set of functions that can be used to represent the Kohn-Sham orbitals in a way that is suitable for computational implementation. The accuracy of a DFT calculation increases as the basis set size increases, but so does its computational cost. There are at least two sources of error that occur from basis sets. The first is a basis set incompleteness error [247; 248], which is the difference between results obtained with, or extrapolated to, a complete basis set and results obtained with a smaller basis set. The second error that can arise is a basis set superposition error [249], which occurs when atoms are in close proximity to each other. In this situation, the basis functions of the atoms can overlap, resulting in the basis set size for each atom artificially increasing as it borrows basis functions from its neighbouring atoms. This can be result in erroneous results such as overestimated binding energies [248; 249]. Basis sets should therefore be chosen to minimise these sources of error and have a suitable degree of accuracy.

Most modern DFT software packages either use plane waves or atomic orbitals to represent the electronic structure [68]. Approaches that use the former rely on delocalised basis functions and the basis can be systematically improved to achieve monotonic convergence [68]. However, core electrons need to treated carefully, and various approaches exist to treat them such as pseudopotentials [250; 251] and projector-augmented waves [72; 73; 252]. In contrast, atom-centred basis functions are centred on atomic nuclei and can easily describe core electrons, but the convergence of results with respect to basis-set complexity is typically not strictly monotonic [68]. However, for systems such as surfaces exposed to vacuum and low-dimensional systems, atom-centred basis functions have the benefit that basis functions are only placed around atoms and do not cover vacuum regions [68], which

is why atomic orbitals are used in this work. It should be noted though that if well-converged settings are used, plane-wave and atom-centred basis functions should result in identical results [253].

There are many types of atomic orbitals that can be used to describe basis sets for DFT calculations. Nuclear-centred Slater-type orbitals [254], which are solutions to the Schrödinger equation for hydrogen-like atoms, have the correct asymptotic form to exhibit exponential decay when they are far away from the nucleus, as well as satisfy Kato's nuclear cusp condition [255] for interactions close to the nucleus [256]. However, Slater-type orbitals struggle to accurately describe many-electron interactions, as these are not encountered by hydrogen-like atoms, and are very computationally demanding. Gaussian-type orbitals have since gained a lot of popularity as they facilitate much faster evaluations than Slater-type orbitals whilst still being able to approximate them [256; 257; 258]. Numeric atomic orbitals (NAOs) have recently gained popularity as a suitable parameterisation can represent Slater- or Gaussian-type orbitals exactly [256]. Furthermore, a single NAO can simultaneously satisfy the Kato nuclear cusp condition [255] for interactions close to the nucleus as well as demonstrate the aforementioned exponential decay [256]. NAOs are typically the most accurate atomic orbital basis set choice [70; 256; 259] and for this reason are used herein to conduct DFT calculations.

2.2 Dispersion Corrections

While many DFAs are able to capture around 99% of the total electronic energy (depending on the system) [260], many of them fail to account for long-range non-covalent interactions [261] such as hydrogen bonds and vdW forces, which often play a large part in the formation and stability of molecular complexes and nanostructured materials [262]. Despite having a minimal contribution to the absolute electronic energy, these long-range dispersion effects have been shown to have a significant effect on relative interaction energies [263; 264; 265], binding energies [222; 266; 267; 268; 269] and structural features [265; 270; 271], as well as the electronic [272; 273], thermodynamic [274], kinetic [275; 276; 277] and mechanical [278; 279] properties of intermolecular complexes, aggregates and nanostructures [260]. There have been several attempts to account for long-range dispersion interactions, such as vdW density functionals [280; 281], though these methods tend to be computationally demanding. A popular method is to account for long-range effects via an *a posteriori* correction scheme, of which there are many varieties,

which are discussed below.

2.2.1 Tkatchenko-Scheffler Scheme

In the Tkatchenko-Scheffler (TS) correction scheme, the dispersion energy, E_{vdW}, is defined as a pairwise interatomic term, as shown in Equation (2.23) [262], which is subsequently added to the total energy.

$$E_{\text{vdW}} = \sum_{\text{A,B}} f_{\text{damp}}(R_{\text{A,B}}) \frac{C_6(\text{A,B})}{R_{\text{A,B}}^6},$$

(2.23)

where f_{damp} is a damping function which controls short-range behaviour and corrects for singularities at short distances, $R_{\text{A,B}}$ is the distance between atoms A and B, and C_6 is an diatomic-specific coefficient for A and B. C_6 coefficients for every pair of atoms can be calculated using various methods, such as empirical interpolations as is done within the pairwise DFT-D3 [282] scheme. The TS method removes this empirical dependence by computing static atomic polarisabilities, $\alpha_{\text{A}}^{\text{free}}$, and C_6 coefficients for each free atom A, C_6^{free}, using tabulated values [283]. Both $\alpha_{\text{A}}^{\text{free}}$ and C_6^{free} are then scaled via the Hirshfeld atoms-in-molecules partitioning scheme [284]. This scheme partitions the electron density from a DFT calculation and can be used to calculate the effective volume of an atom A within a molecule, $\nu_{\text{A}}^{\text{eff}}$. This quantity can then be divided by its known volume as a free atom, $\nu_{\text{A}}^{\text{free}}$, to calculate a Hirshfeld volume ratio, H_{A}, as shown in Equation (2.24).

$$H_{\text{A}} = \frac{\nu_{\text{A}}^{\text{eff}}}{\nu_{\text{A}}^{\text{free}}}$$

(2.24)

Using H_{A}, the atomic polarisability of atom A, α_{A}, within a system can be calculated using Equation (2.25), while $C_6(\text{A,A})$ can be calculated using Equation (2.26):

$$\alpha_{\text{A}} = H_{\text{A}} \alpha_{\text{A}}^{\text{free}}$$

(2.25)

$$C_6(\text{A,A}) = H_{\text{A}}^2 C_6^{\text{free}}(\text{A,A})$$

(2.26)

Then, for any two atoms A and B, their $C_6(\text{A,B})$ coefficient can be calculated using Equation (2.27):

$$C_6(\text{A,B}) = \frac{2C_6(\text{A,A})C_6(\text{B,B})}{\frac{\alpha_{\text{B}}}{\alpha_{\text{A}}}C_6(\text{A,A}) + \frac{\alpha_{\text{A}}}{\alpha_{\text{B}}}C_6(\text{B,B})}$$

(2.27)

The tabulated vdW radius of each atom, $R_{\text{A}}^{\text{free}}$, is also scaled using H_{A} to

calculate an effective radius, R_A^{eff}, as shown in Equation (2.28):

$$R_A^{\text{eff}} = H_A^{1/3} R_A^{\text{free}} \tag{2.28}$$

Then, f_{damp} in Equation (2.23) can be defined using Equation (2.29):

$$f_{\text{damp}}(R_{\text{A,B}}, R_{\text{A,B}}^{\text{free}}) = \frac{1}{1 + \exp\left(-d\left(\dfrac{R_{\text{A,B}}}{s_R R_{\text{A,B}}^{\text{free}}} - 1\right)\right)} \tag{2.29}$$

where d is a parameter to control the damping steepness and s_R is a screening parameter that depends on the DFA to which the TS scheme is coupled.

2.2.2 Many-Body Dispersion Schemes

Many-body dispersion (MBD) [285; 286; 287; 288; 289] correction schemes have become popular as they extend upon Equation (2.23) by accounting for long-range correlation beyond pairwise atomic interactions. There are many variants of MBD that exist [287; 288], but they all account for long-range correlations in some way. In any MBD method, an atomic system is mapped to a model Hamiltonian of quantum harmonic oscillators located at $\mathbf{r}$ with static polarisabilities α and non-interacting uncoupled frequencies ω. In a finite system of N oscillators, where each oscillator represents an atom, the MBD Hamiltonian, $\hat{\mathcal{H}}_{\text{MBD}}$, can be defined as Equation (2.30):

$$\hat{\mathcal{H}}_{\text{MBD}} = \sum_{j=1}^{N} -\frac{1}{2}\nabla^2(\xi_j) + \sum_{j=1}^{N} \frac{1}{2}\omega_j{}^2\xi_j{}^2 + \sum_{j>k}^{N} \omega_j\omega_k\sqrt{\alpha_j\alpha_k}\xi_j\mathbf{T}_{jk}\xi_k \tag{2.30}$$

where ξ_j are the displacements of the oscillator charges weighted by the oscillator masses, and $\mathbf{T}$ is a damped dipole interaction tensor. The first summation in Equation (2.30) corresponds to the kinetic energy, the second summation represents the potential energy, while the final summation is for dipole-dipole interactions between the oscillators. Diagonalisation of $\hat{\mathcal{H}}_{\text{MBD}}$ will yield a set of interacting dipole-coupled oscillation frequencies, $\tilde{\omega}$. The MBD energy, E_{MBD}, can then be obtained as the change in the zero-point energies of the coupled and uncoupled oscillations induced by the dipole interaction, as shown in Equation (2.31).

$$E_{\text{MBD}} = \sum_{l=1}^{3N} \frac{\tilde{\omega}_l}{2} - \sum_{j=1}^{N} \frac{3\omega_j}{2} \tag{2.31}$$

Range-Separated Self-Consistent Screening Variant

The range-separated self-consistent screening variant of MBD (MBD@rsSCS), is the many-body extension of the TS method outlined in Section 2.2.1. In this variant, a screening procedure is used to obtain a refined set of parameters before they are used within $\hat{\mathcal{H}}_{\mathrm{MBD}}$ in Equation (2.30). To do this, a frequency-dependent polarisability, $\alpha_A(i\omega)$, for each atom A at an imaginary frequency $i\omega, i \in \mathbb{I}$, is calculated using Equation (2.32).

$$\alpha_A(i\omega) = \frac{\alpha_A[\rho]}{1 + \left(\dfrac{\omega}{\omega_A[\rho]}\right)^2} \tag{2.32}$$

The short-range atomic polarisabilities can then be extracted after applying a Dyson-like screening, as shown in Equation (2.33), where α^{TS} is the sum of effective polarisabilities calculated using the TS method, as shown in Equation (2.25).

$$\alpha^{\mathrm{SCS}}(\mathbf{r}, i\omega) = \alpha^{\mathrm{TS}}(\mathbf{r}, i\omega) + \alpha^{\mathrm{TS}}(\mathbf{r}, i\omega) \int \mathbf{T}(\mathbf{r} - \mathbf{r}_1)\alpha^{\mathrm{SCS}}(\mathbf{r}_1, i\omega)\,\mathrm{d}\mathbf{r}_1 \tag{2.33}$$

As α^{TS} is continuous over space, and by representing N atoms using quantum harmonic oscillators, Equation (2.33) can be rewritten as the discretised Equation (2.34) after integration over $\mathbf{r}$. The α_A^{SCS} term is finally made isotropic before entering Equation (2.30).

$$\alpha_A^{\mathrm{SCS}}(\mathbf{r}, i\omega) = \alpha_A^{\mathrm{TS}}(\mathbf{r}, i\omega) + \alpha_A^{\mathrm{TS}}(\mathbf{r}, i\omega) \sum_{A \neq B} \mathbf{T}_{A,B}\alpha_B^{\mathrm{SCS}}(\mathbf{r}_1, i\omega) \tag{2.34}$$

Non-Local Variant

Rather than use an atomic response model in the form of static polarisabilities as MBD@rsSCS does, the non-local variant of MBD (MBD-NL) uses the Vydrov and Van Voorhis polarisability functional [290], α^{VV}, which is a functional of ρ, to parameterise atomic responses [288]. While MBD-NL retains a similar level of empiricism as MBD@rsSCS, the need for tabulated vdW radii and short-range screening is removed [288]. The polarisability density is coarse-grained into atomic fragments and can be expressed as Equation (2.35):

$$\alpha^{\mathrm{VV}}[\rho](\mathbf{r}, i\omega) = \frac{\rho(\mathbf{r})}{\dfrac{4\pi}{3}\rho(\mathbf{r}) + C\dfrac{|\nabla\rho(\mathbf{r})|^4}{\rho(\mathbf{r})^4} + \omega^2} \tag{2.35}$$

where C is an empirical parameter. Hirshfeld volume ratios, $H_A(\mathbf{r})$, are then used to partition $\alpha^{VV}[\rho](\mathbf{r}, i\omega)$ into atomic polarisabilities, $\alpha_A^{VV}(\mathbf{r}, i\omega)$, for each atom A, as shown in Equation (2.36).

$$\alpha_A^{VV}(\mathbf{r}, i\omega) = \int H_A(\mathbf{r})\alpha^{VV}[\rho](\mathbf{r}, i\omega)\, \mathrm{d}\mathbf{r} \tag{2.36}$$

The C_6(A,A) coefficients, if needed, can then be calculated using the Casimir-Polder formula [291; 292], as shown in Equation (2.37):

$$C_6(\mathrm{A,A}) = \frac{3}{\pi} \int_0^\infty \alpha_A^{VV}(i\omega)^2\, \mathrm{d}\omega \tag{2.37}$$

2.3 Density Functional Tight-Binding

Density functional tight-binding (DFTB) is an electronic structure method based on an approximation to the Kohn-Sham DFT expansion of the total energy functional, as expressed in Equation (2.13) [79]. In DFTB, ρ is expressed as the perturbation, $\delta\rho$, of a carefully chosen reference density, ρ_0 i.e. $\rho = \rho_0 + \delta\rho$. The energy functional can then be expanded as a Taylor series, as expressed in Equation (2.38).

$$E[\rho_0 + \delta\rho] = \sum_{i=0}^\infty E^i[\rho_0, (\delta\rho)^i] \tag{2.38}$$

There are several DFTB approximations, with each based on what order the Taylor series in Equation (2.38) is truncated to. The first-order, non-self-consistent DFTB1 [293; 294] approach only takes the $E^0[\rho_0]$ and $E^1[\rho_0, \delta\rho]$ terms into account. The second-order, self-consistent DFTB2 [78] approach also includes the $E^2[\rho_0, (\delta\rho)^2]$ term, while the newer DFTB3 [295; 296] approach includes the third-order term. Only the DFTB2 method is used in this work, and is expanded upon below.

2.3.1 DFTB2

To evaluate the total DFTB2 energy, the E^0, E^1 and E^2 terms in Equation (2.38) need to be evaluated. E^0 is a functional of ρ_0, which it means it is essentially universal and does not depend on the surrounding chemical environment [79]. Therefore, it can be evaluated for a suitable 'reference system' and be applied to any environment. In DFTB, E^0 is approximated as a sum of repulsive pairwise potentials, as

expressed in Equation (2.39).

$$E^0[\rho_0] \approx \frac{1}{2} \sum_{\text{A,B}} V_{\text{rep}}(\text{A,B}), \qquad (2.39)$$

where $V_{\text{rep}}(\text{A,B})$ is the repulsive energy term between atoms A and B. These repulsive pair potentials can be computed by fitting to either experimental data [297] or DFT calculations [294].

Equation (2.40) shows the exact expression for the E^1 functional in terms of one-electron orbitals ψ_i, occupation numbers n_i, and a Hamiltonian operator, $\hat{\mathcal{H}}$ [79].

$$E^1[\rho_0, \delta\rho] = \sum_i n_i \langle \psi_i | \hat{\mathcal{H}}[\rho_0] | \psi_i \rangle \qquad (2.40)$$

Two approximations are made to calculate E^1: firstly, ψ_i is expressed using a valence-only minimal basis set, $\{\phi_m\}$, within a linear combination of atomic orbitals *ansatz* [79], as expressed in Equation (2.41).

$$\psi_i = \sum_m c_{mi} \phi_m \qquad (2.41)$$

where c_{mi} are expansion coefficients and $\{\phi_m\}$ is explicitly calculated from DFT calculations by solving the Kohn-Sham equations for an isolated atom with an additional (usually harmonic) confining potential [79]. Secondly, a two-centre approximation to $\hat{\mathcal{H}}$ is used in a linear combination of atomic orbitals basis, as expressed in Equation (2.42) [79].

$$\langle \phi_m | \hat{\mathcal{H}}[\rho_0] | \phi_v \rangle \approx \langle \phi_m | - \frac{1}{2}\nabla^2 + V[\rho_A + \rho_B] | \phi_v \rangle \qquad (2.42)$$

where $m \in \text{A}$ and $v \in \text{B}$. This representation can be computed by evaluating Kohn-Sham equations for dimers [79].

To calculate the E^2 term, $\delta\rho$ is expressed as a superposition of atomic contributions, as expressed in Equation (2.43) [79].

$$\delta\rho(\mathbf{r}) = \sum_{\text{A}} \delta\rho_{\text{A}}(\mathbf{r} - R_{\text{A}}) \qquad (2.43)$$

Then, the E^2 term can be approximated as a function of charge fluctuations,

Δq, for each atom, as shown in Equation (2.44) [78; 79].

$$E^2(R_{\mathrm{AB}}) \approx \frac{1}{2} \sum_{A \neq B} \overline{\gamma}_{\mathrm{AB}}(R_{\mathrm{AB}}) \Delta q_{\mathrm{A}} \Delta q_{\mathrm{B}} \tag{2.44}$$

where $\overline{\gamma}$ is an analytical function and R_{AB} is the distance between atoms A and B. Here, Δq is the difference between the Mulliken charge with respect to the neutral atom.

2.3.2 Limitations of DFTB

As can be seen in Section 2.3.1, there are many approximations made in order to calculate DFTB energies. While these help to significantly reduce the computational expense of DFTB as compared to DFT, they are also responsible for some inaccuracies. While accurate fitting to the E^0 term in Equation (2.39) can resolve some inaccuracies, not all approximations can be compensated for via parameterisations [79]. For example, different bonding environments between two atoms cannot be differentiated within one parameterisation. Single, double and triple bonds therefore have to be accounted for within one repulsive potential, which is not a realistic way to account for the different bonding interactions [79]. As DFTB is a DFT-derived method, some DFT limitations are also inherited by DFTB, such as the inability to account for long-range, non-covalent interactions [79] (see Section 2.2 for a detailed explanation). The use of a minimal basis set can also be problematic for solids such as silicon, which require d orbital contributions to describe the conduction band [79].

2.4 Hybrid Quantum Mechanics/Molecular Mechanics

While methods like DFT can provide detailed insights into the structural and electronic properties of many materials [68; 69], their computational scalability as standalone methods generally means that higher-rung DFAs cannot be used to study large, periodic models without incurring high, and often infeasible, computational costs. Furthermore, while the use of periodic boundary conditions can be helpful to study extended surfaces [68], they become problematic when modelling isolated defects or dopants as they acquire a periodicity under such boundary conditions, when they should in fact be aperiodic.

To address the fact that surface chemistry is intrinsically local, as well as potentially aperiodic in nature in the presence of surface defects on semiconductors,

embedded cluster approaches are a viable alternative. In embedded cluster models, the material is modelled as a finite system and is embedded into a surrounding medium that is typically treated at a lower level of theory [68]. Embedded cluster models have been treated using a variety of approaches, such as hybrid quantum mechanics/quantum mechanics [298], molecular mechanics/molecular mechanics [299] and quantum mechanics/molecular mechanics (QM/MM) [300]. Hybrid QM/MM is an approach that aims to combine the accuracy of QM methods with the speed of MM methods, whereby a region of interest within a system, such as a defect, dopant or reaction site, is treated using a QM method such as DFT, while the surrounding environment is treated using a classical forcefield [301], as shown in Figure 2.1. This is typically done by cutting a cluster model from a periodic unit cell, which itself can be treated using QM or MM, and partitioning it into QM and MM regions, as shown in Figure 2.1. Apart from being partitioned into QM and MM regions, the embedded cluster is also usually partitioned into an active region, which includes the entire QM region and part of the MM region, and defines the volume of the cluster that is to be optimised; and a frozen region, where atoms are constrained during optimisation to ensure the active region is not affected by the edges of the cluster. The QM/MM methodology allows for local chemical processes to be modelled with the accuracy of QM, whilst circumventing the computational expense required to treat the full system with QM [301; 302].
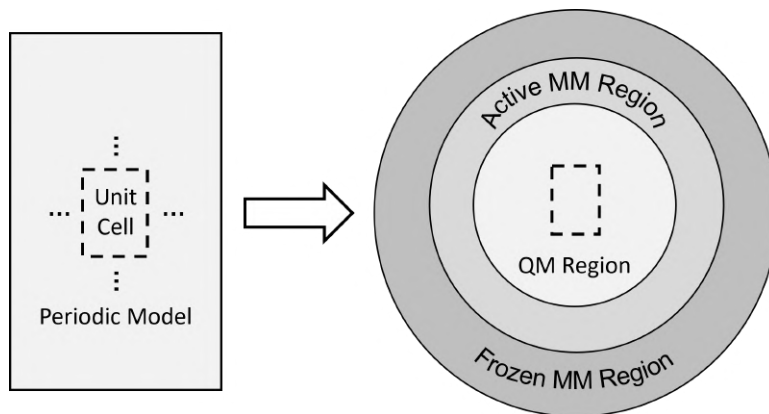
Figure 2.1. Process of cutting an infinite periodic surface model, defined by a unit cell, into a finite, embedded cluster surface model partitioned into quantum mechanical (QM) and molecular mechanical (MM) regions. The MM region itself is split into active and frozen parts. The position of the unit cell from the periodic model within the embedded cluster model is shown using dotted black lines, but note that the cluster is not defined by a unit cell.

There are two ways to calculate the total QM/MM energy of a system. The

first is via a subtractive scheme, as shown in Equation (2.45).

$$E = E_{\text{QM}}(\text{QM}) + E_{\text{MM}}(\text{QM} + \text{MM}) - E_{\text{MM}}(\text{QM}) \qquad (2.45)$$

where $E_{\text{QM}}(\text{QM})$ denotes the energy of the QM region as calculated using QM, $E_{\text{MM}}(\text{QM})$ is the energy of the QM region as calculated using MM, and $E_{\text{MM}}(\text{QM}+$ MM) is the energy of the combined QM and MM regions as calculated using MM. An alternative approach is an additive scheme, as shown in Equation (2.46), which is typically more popular [302; 303; 304] and is used herein. Despite its greater popularity, it should be noted that if properly set up, Equations (2.45) and (2.46) should provide identical results [304].

$$E = E_{\text{QM}}(\text{QM}) + E_{\text{MM}}(\text{MM}) + E(\text{QM/MM}) \qquad (2.46)$$

where $E_{\text{MM}}(\text{MM})$ is the energy of the MM region as calculated using MM and $E(\text{QM/MM})$ is the interaction energy between the QM and MM regions and includes the bonded, electrostatic and vdW interactions between the atoms in the two regions [305].

When partitioning any system into two regions, such as QM and MM regions, it is important to ensure the interface between the two regions is treated properly. For covalent systems such as diamond, the boundary between the two regions will inevitably pass through a covalent carbon-carbon bond. This therefore requires proper truncation of the embedded region, and many methodologies exist for this purpose, such as localised-orbital schemes [306; 307], where hybrid orbitals are placed at the boundary to saturate the embedded QM region, and boundary-atom schemes, where the MM atom is replaced by an atom that appears in both the QM and MM calculations [302]. The most common method however, and the one used herein, is the hydrogen link-atom approach [302; 308; 309]. In the hydrogen link-atom approach, for every covalent bond across the QM-MM boundary, a hydrogen atom is added to covalently bond with the QM atom to satisfy its valency. Forcefield terms in the MM calculation also need to be selectively deleted to ensure no interactions are double counted [310]. It is important to note that the hydrogen atom is not part of the real system and only serves to prevent the appearance of unsaturated dangling bonds within the QM calculation.

2.5 Machine Learning

ML involves the creation of an algorithmic framework that learns the statistical trends of its input data and uses these to predict outputs, whilst also possessing the ability to improve upon its predictions if additional input is provided [102]. Mathematically, most ML methods aim to find and optimise a predictive function $f : \mathbb{X} \to \mathbb{Y}$ that acts as a 'universal approximation' by mapping an input space $\mathbb{X}$ to a target space $\mathbb{Y}$ [102].

There are two primary regimes of ML: supervised and unsupervised, though others also exist [102]. Supervised ML seeks to find a function f given a known finite training set, $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, where $(x_i, y_i) \in \mathbb{X} \times \mathbb{Y}, i \in \mathbb{N}$ [102]. In the ideal case, f should ensure $f(x_i) = y_i$ for all x_i. If new x' values are provided, f should then be able to accurately predict $f(x')$ values. This is in contrast to unsupervised ML, where only $\mathbb{X}$ is known and $\mathbb{Y}$ is unknown, and the training set is only $\{x_1, x_2, \cdots, x_n\}$. Unsupervised ML is typically used to ascertain patterns within data, such as clustering or for dimensionality reduction [102]. For the purposes of this work, supervised ML is made use of to train regression models on DFT calculations for systems x_i with properties y_i.

While $f(x_i) = y_i$ is the ideal prediction, in reality this equality does not hold and there is an error between $f(x_i)$ and y_i. Many functions map $\mathbb{X}$ to $\mathbb{Y}$, and the training process seeks to find the function that minimises the expected risk. Let $F = \{g \,|\, g : \mathbb{X} \to \mathbb{Y}\}$. The accuracy of any $g \in F$ can be determined using a loss function, $\ell : g(\mathbb{X}) \times \mathbb{Y} \to \mathbb{R}$, which evaluates the error between the predicted $g(x)$ and true y values [102; 311]. It is important that ℓ is chosen carefully as it is dependent on how the ML model is trained. The expected risk, R, of any $g \in F$ can then be defined as the expectation of its loss function, $\mathbb{E} : \ell \to \mathbb{R}$, as shown in Equation (2.47),

$$R(g) = \mathbb{E}[\ell(g(\mathbb{X}), \mathbb{Y})] = \int \ell(g(\mathbb{X}), \mathbb{Y}) \, d\mathbb{P}(\mathbb{X}, \mathbb{Y}) \qquad (2.47)$$

where $\mathbb{P}$ is a joint probability distribution over $\mathbb{X}$ and $\mathbb{Y}$ [311]. Here, an assumption is made that T comprises n instances which are independently and identically distributed from $\mathbb{P}(\mathbb{X}, \mathbb{Y})$, and $\mathbb{Y}$ is assumed to be a random variable with a distribution conditional on $\mathbb{X}$. Then, the optimal predictive function, g^*, can be expressed as

Equation (2.48) [102; 311].

$$g^* = \arg\min_{g \in F}(R(g)) \tag{2.48}$$

In practice however, $\mathbb{P}$ is unknown and in supervised ML, ℓ requires knowledge of $\mathbb{Y}$. Therefore, g^* is calculated to be the function that minimises the empirical risk [102; 311; 312], R_{emp}, which is the average of ℓ over the size of T, as shown in Equation (2.49):

$$R_{\text{emp}}(g) = \frac{1}{n} \sum_{i=1}^{n} \ell(g(x_i), y_i) \tag{2.49}$$

Conditions can be applied to the ML algorithm to ensure $R_{\text{emp}} \to R$ as $n \to \infty$ in Equation (2.49) [311].

2.5.1 Neural Networks

There are many ML architectures that exist to optimise a predictive function. One such framework is NNs, which are circuit-like arrangements inspired by how neurons process information. In biological NNs, upon the application of a stimulus, a neuron receives information through its dendrites [313]. The neuron then emits a signal along its axon and at the majority of synapses, the signal is transferred from its axon to the dendrite of another neuron, and the whole process repeats [313]. This is how the framework of an artificial NN is modelled, where neurons are termed 'nodes' instead, neurites (dendrites and axons) are termed 'connections', and a block of nodes is a 'layer'. The inputs into any node are all assigned weights depending on their relative importance. The node then applies some function to the weighted sum of its inputs and maps this to its successor node. Figure 2.2 shows a comparison between a biological and artificial neuron. There are three types of layers within any NN. The first is the input layer, which only receives information and passes it on to its successive layer. The information sent from the input layer is then received by a hidden layer, where intermediate processing is done. If multiple hidden layers are used, the process is referred to as 'deep learning' [314]. The output layer processes all the weighted information received from the final hidden layer and provides an output based on this.
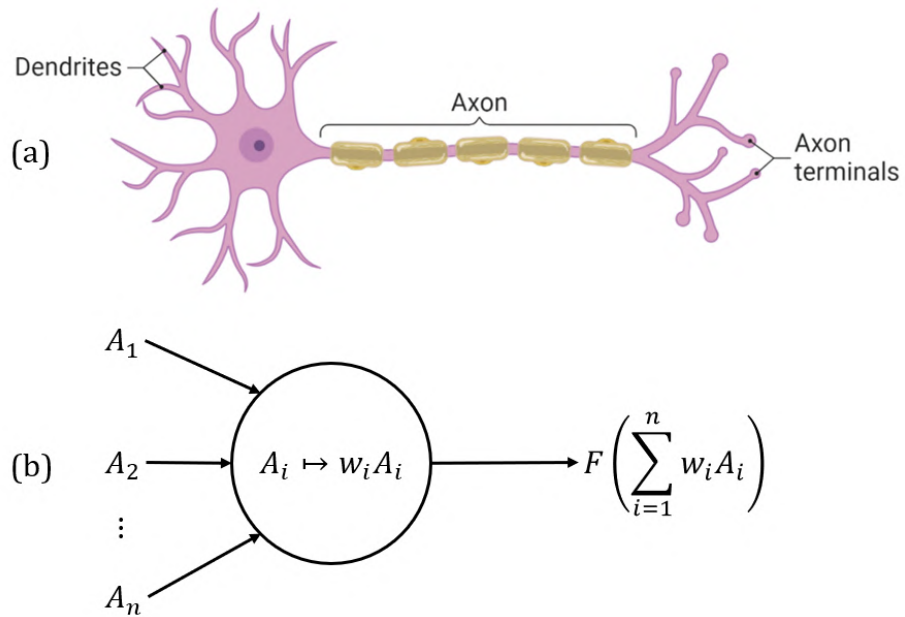
Figure 2.2. Comparison of neurons within (a) biological and (b) artificial neural networks. In (b), each input A_i is assigned a weight w_i and the node applies some function, F, to $\sum_{i=1}^{n} w_i A_i$. Figure (a) is adapted from "Neuron Anatomy" by BioRender.com (2022), retrieved from https://app.biorender.com/biorender-templates.
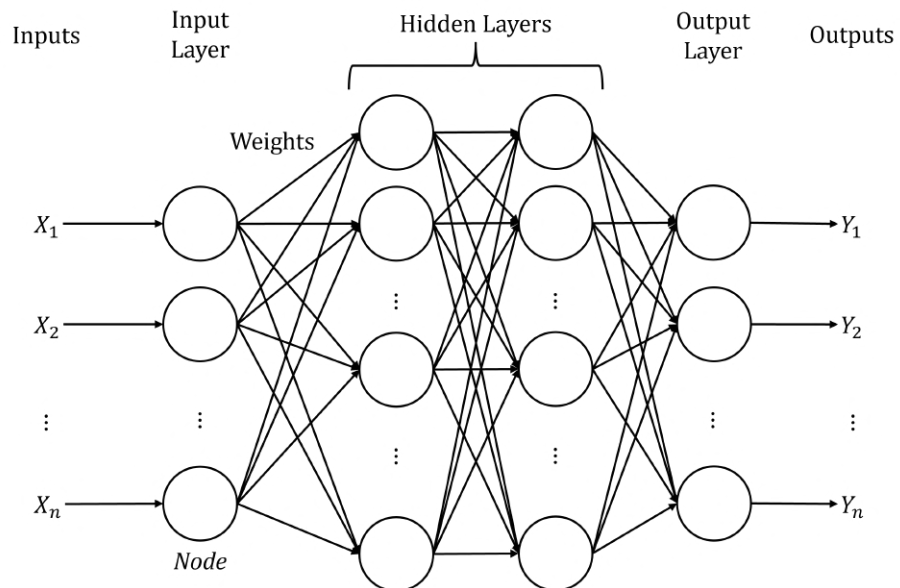
Figure 2.3. Example schematic of the structure of a neural network.

2.5.2 Machine-Learned Interatomic Potentials

Once appropriately trained, machine-learned interatomic potentials (MLIPs) can be used *in lieu* of *ab initio* methods such as DFT at a fraction of their computational cost. Furthermore, if trained using *ab initio* data, such potentials can theoretically retain the *ab initio*-level of accuracy possessed by the training data (assuming there are no prediction errors). MLIPs thus have the potential to advance many computational tasks, including structure searches and geometry optimisations of complex systems containing thousands of atoms. There are many different flavours of MLIPs, such as NN potentials [315; 316; 317; 318; 319; 320], which are used herein. It should be noted though that there is no universally best MLIP that exists for every problem [321] because any pair of optimisation algorithms should be equivalent when their performances are averaged over all possible problems, as explained within the no free lunch theorem [322].

To train any MLIP, the atomic environment first has to be transformed into a representation that is machine-learnable; this representation is termed a 'descriptor'. Unless specifically required, descriptors should generally aim to satisfy a number of properties [323], such as invariance to translation and rotation of the atomic system, as well as invariance to the permutation of the atomic indices within the system [321]. There are many different descriptors that exist to model atomic systems. A popular example is the simplified molecular-input line-entry system [324] which uses a line notation to encode molecular structures and only utilises the bonding patterns within a molecule [325]. Such graphical descriptors do not require the full 3D molecular structure, which can instead be extracted from line notations [325]. However, problems can arise if there are strong electron delocalisation effects within the system, such as in metals or π-conjugated molecules [325]. An alternative approach is to make use of geometrical features [325], which have been shown to perform well when predicting thermodynamic and electronic properties [326; 327; 328]. Global descriptors can be used to encode information about the whole atomic structure [321], examples of which include the Coulomb-like matrix [329], the Ewald sum matrix and the sine matrix [323]. However, these matrical descriptors are not permutationally invariant as the matrix columns and rows are dependent on the ordering of atomic indices, though permutational invariance can be enforced by methods such as diagonalisation or by ordering rows and columns using a norm [321; 329]. Furthermore, global descriptors depend on and typically scale unfavourably with the size of a system and the number of interactions within it [325; 330]. In contrast, local descriptors such as the smooth overlap of atomic positions [331; 332] or the one used

within the SchNet NN [315; 316; 333], the latter of which is used herein, can be used to encode a localised region of an atomic structure [321]. It is also not too computationally costly to combine the information from multiple local descriptor outputs to form a global prediction [334]. However, local descriptors require a specified radial cut-off region to describe atoms with respect to their local chemical and structural environments. While this ensures short-range interactions are learned by the MLIP, challenges can arise when attempting to learn long-range interactions [318], as it quickly becomes computationally tedious and infeasible to train MLIPs for cut-off radii beyond a few angstroms.

2.6 *Ab Initio* Thermodynamics

Ab initio thermodynamics is a methodology that can be used to determine the equilibrium composition and geometry of a surface in contact with a gaseous atmosphere under finite temperature, T, and pressure, p, conditions [335]. This is done by calculating the Gibbs free energy, $G(T, p)$, using total energies, typically evaluated by some *ab initio* method such as DFT [65; 66]. This approach allows for a system to be split into subsystems which can be connected by assuming equilibrium between the subsystems, while also allowing for finite temperature effects to be taken into account [335]. However, no temporal or kinetic effects are considered, and the assumption of equilibrium is not always a realistic one.

Consider a solid surface of element X in contact with an atmosphere of gas Y. This environment can be thought of as a reservoir, as any amount of Y particles can be given to or taken away from the surface without any change in T or p [335]. The Gibbs free energy of this entire system can be split into contributions that arise from the bulk solid, G_{solid}, the homogeneous gas phase, G_{gas}, and an additional term introduced through the surface, ΔG_{surf}, [335] as shown in Equation (2.50).

$$G = G_{\text{solid}} + G_{\text{gas}} + \Delta G_{\text{surf}} \tag{2.50}$$

If surface homogeneity is assumed, as is the case for the surfaces of ideal single crystals, ΔG_{surf} will scale linearly with the surface area, A [335]. Then, the surface free energy, γ, can be expressed using Equation (2.51).

$$\gamma = \frac{1}{A}(G - G_{\text{solid}} - G_{\text{gas}}) \tag{2.51}$$

It should be noted that γ is well-defined for a finite part of the total, infinite system.

However, as distance from the interface increases, the solid and gas phases will not be affected by the surface interface. [335]. Therefore, only the finite part of the system affected by the surface needs to be considered [335]. If this part comprises N_X solid atoms and N_Y gas atoms per surface area, Equation (2.51) can be rewritten as Equation (2.52).

$$\gamma(T,p) = \frac{1}{A}\big(G(T,p,N_\mathrm{X},N_\mathrm{Y}) - N_\mathrm{X}\mathrm{X}(T,p) - N_\mathrm{Y}\mu_\mathrm{Y}(T,p)\big) \quad (2.52)$$

where X is the Gibbs free energy per X atom in the bulk, and μ_Y is the chemical potential of Y in the gas phase. Using Equation (2.52), the surface free energy of the clean surface can be defined as:

$$\gamma_\mathrm{clean}(T,p) = \frac{1}{A}\big(G(T,p,N_\mathrm{X},0) - N_\mathrm{X}\mathrm{X}(T,p)\big) \quad (2.53)$$

Then, the Gibbs free energy of adsorption, ΔG_ads, can be expressed as the difference between γ_clean and γ, as shown in Equation (??).

$$\Delta G_\mathrm{ads} = \quad (2.54)$$

$$=$$

eq:DeltaG$_a$ds

For any given (T,p) couple in the gas phase, the most stable surface structure is the one that γ or ΔG_ads at the corresponding μ_Y [335]. The key quantities that determine γ or ΔG_ads are therefore the Gibbs free energies of the solid surface and the bulk solid, as well as the chemical potential of the gaseous environment [335].

The solid-phase Gibbs free energies can be computed by decomposing G into four contributing terms, as shown in Equation (2.55):

$$G = E_\mathrm{total} + F_\mathrm{vib} + F_\mathrm{conf} + pV \quad (2.55)$$

where E_total is the energy , F_vib is the vibrational free energy, F_conf is the configurational free energy and V is the volume [335]. The dominant term in Equation (2.55) is E_total, which can be calculated using an *ab initio* method such as DFT [65; 66]. For the purposes of the work herein, the vibrational contributions to ΔG_ads are not considered, which only leaves the pV term. Using dimensional analysis, it can

be seen that the contribution of the pV term to γ, normalised with respect to A, is atm·Å^3/Å$^2 \sim$ μeV/Å^2, which is negligible for a pressure of even hundreds of standard atmospheres [335]. Therefore, G can be equated to an *ab initio* energy, and the size of the error can be attributed to the accuracy of the specific *ab initio* method [335].

Assuming gas Y exists as a diatomic molecule, Y$_2$, at standard conditions, μ_Y can be expressed as Equation (2.58):

$$\mu_Y(T, p) = \frac{1}{2}\mu_{Y_2}(T, p) \tag{2.56}$$

$$= \frac{1}{2} \cdot \frac{\Delta G_Y}{N_Y} \tag{2.57}$$

$$= -\frac{1}{2} \cdot \frac{k_B T \ln(Q_Y) + pV}{N_Y} \tag{2.58}$$

where Q is the partition function and k_B is the Boltzmann constant. Using statistical mechanics (and once again, excluding vibrational contributions), Equation (2.58) can be reduced to Equation (2.59).

$$\Delta\mu_Y(T, p) \approx \frac{1}{2}E_{Y_2} + \Delta\mu_Y(T, p^{\ominus}) + \frac{1}{2}k_B T \ln\left(\frac{p}{p^{\ominus}}\right) \tag{2.59}$$

where E_{Y_2} can be calculated using an *ab initio* method and $p^{\ominus}$ is standard pressure. $\Delta\mu_Y(T, p^{\ominus})$ can be calculated from tabulated enthalpy and entropy values available in thermochemical tables [336].

Bringing it all together, Equations (2.55) and (2.59) can be used to rewrite Equation (??) as a function of μ_Y, as shown in Equation (2.61).

$$\Delta G_{ads}(\Delta\mu) = \frac{1}{A}\left(E_{total} - E_{clean} - \frac{N_Y}{2}E_{Y_2} - N_Y\Delta\mu_Y\right) \tag{2.60}$$

$$= -\frac{N_Y}{A}\left(E_{ads} + \Delta\mu_Y\right) \tag{2.61}$$

where E_{ads} is the adsorption energy of Y, which can be calculated using *ab initio* methods.

2.7 Software Packages

A variety of software packages were used to conduct the work herein, and they are briefly discussed below.

2.7.1 DFTB+

DFTB+ [79] (website: https://dftbplus.org/) is a software package that can be used to run DFTB calculations. DFTB+ provides several default parameter sets based on the DFTB2 and DFTB3 methods. All parameter sets are constructed using electronic and repulsive components. The former comprises atomic and diatomic contributions which are typically computed from LDA or GGA calculations. The repulsive energy, which is approximated as the sum of pair potentials as shown in Equation (2.39), is represented using splines or polynomials. Parameter sets are stored as Slater-Koster files, which contain electronic and repulsive parameters between atoms within the parameter set. Parameter sets are usually limited to a subset of elements within the periodic table due to the $O(N^2)$ effort required to acquire parameters for N elements. To run a calculation, the DFTB+ software requires two files: `dftb_in.hsd`, which contains runtime information, and an input geometry file. An example `dftb_in.hsd` file for a geometry optimisation, as used herein, can be found in Listing A.4, which loads a `geometry.gen` file containing information pertaining to the atomic structure of the system, and includes a path to the auorg [86] parameter set, which is a extension of the mio [78] parameterisation to include elemental interactions with gold. Dispersion interactions can be accounted for by adding a `Dispersion{}` subblock within the `Hamiltonian{}` block.

2.7.2 FHI-aims

FHI-aims [70] (website: https://fhi-aims.org/) is an all-electron electronic structure theory software package based on NAOs that can be used to run DFT calculations on both periodic and aperiodic systems. FHI-aims provides pre-constructed basis sets associated with different elements, and these basis sets are defined on four levels: 'light', 'intermediate', 'tight', and 'really_tight' [259]. These element-dependent basis sets are highly compact and retain a high level of accuracy for a range of systems, achieving up to millielectronvolt-level energy convergence [70] and small basis set superposition errors for standard DFAs. Two input files are required to run calculations: `geometry.in`, which contains information pertaining to the atomic structure for a given calculation, and `control.in`, which contains all other runtime-specific information. An example `control.in` file using the PBE [229] GGA and a 'tight' basis set for carbon can be seen in Listing A.5. Keywords for various dispersion corrections can also be added.

The following self-consistency convergence criteria were set for all FHI-aims

calculations: 1×10^{-6} eV for the total energy, 1×10^{-2} eV for the sum of eigenvalues, 1×10^{-5} e/$a_0{}^3$ for the charge density, 1×10^{-4} eV Å^{-1} for the energy derivatives and 1×10^{-2} eV Å^{-1} for the maximum residual force component per atom (for structure optimisation calculations).

2.7.3 GULP

GULP [337; 338] (website: https://gulp.curtin.edu.au/gulp/) is a software package that can be used to perform a variety of MM simulations on materials, and includes a variety of forcefields parameterised for different sets of elements as library files.

2.7.4 Atomic Simulation Environment

The Atomic Simulation Environment [339] (website: https://wiki.fysik.dtu.dk/ase/) is a software package that comprises a set of Python modules for the setting up, manipulation, running, analysing and visualisation of atomistic simulations. Interfaces to other software packages are also provided, which can be called as a `Calculator()` object. Both local and global optimisation algorithms, such as limited-memory Broyden–Fletcher–Goldfarb–Shanno and basin-hopping [98; 99], are also available within this software. The Graphical User Interface of this software was used extensively in this work to manually build and visualise structural geometries.

2.7.5 Py-ChemShell

ChemShell [301; 303; 310] (website: https://www.chemshell.org/) is a modular software package that can be used to run hybrid QM/MM calculations, where a variety of QM and MM software packages (either compiled in as libraries or called through external interfaces) can be utilised to evaluate energies and gradients of the QM and MM regions. ChemShell couples the results of the QM and MM regions to obtain the combined QM/MM energy and gradient, whilst appropriately treating the boundary region that couples the two subsystems [301; 310]. The newer Python-based redevelopment of the code, Py-ChemShell [301], is used herein. Listing 2.1 details a Python script that was used to cut a hemispherical cluster from a periodic surface. The `connect_toler` keyword, a scalar for vdW radii to determine bonding interactions, was set to a value of 1.3 for all QM/MM calculations to ensure correct hydrogen saturation of the QM region for the QM calculator.

```python
1   from ase.io import read
2   from chemsh import *
3   from chemsh.io.tools import *
4   import numpy as np
5
6   # Read periodic surface geometry file
7   slab = read('periodic_slab.in')
8   frag = convert_atoms_to_frag(slab, connect_mode='covalent', dim='2D')
9
10  # Cut cluster
11  cluster = frag.construct_cluster(crystal_type='covalent', origin=8,
    ↪  radius_cluster=20.0, radius_active=10.0)
12
13  # Specify QM region
14  qm_region = np.arange(0,90,1)
15
16  # Partition cluster
17  regions = cluster.partition(qm_region=qm_region, origin=0, cutoff_boundary=0.0,
    ↪  radius_active=10.0)
18
19  # Save partitioned cluster
20  regions.save('cluster_partitioned.pun', 'pun')
```

Listing 2.1. An example Python script used to cut a hemispherical cluster from a periodic surface model. Using this script, a cluster of radius $20\,a_0$ (and active radius $10\,a_0$) is cut and partitioned into a system with a QM region size of 90 atoms.

Listing 2.2 details a Python script used to add a gold atom $1.5\,\text{Å}$ above the saved partitioned cluster from Listing 2.1.

```python
1   from ase import Atoms
2   from chemsh import *
3   from chemsh.io.tools import *
4
5   # Load partitioned cluster
6   frag = Fragment(coords="cluster_partitioned.pun", connect_mode='covalent',
    ↪   connect_toler=1.3)
7
8   # Create gold atom 1.5 angstroms above atom 7
9   coords = frag.coords[7].copy()
10  coords /= 1.88973    # Convert from Bohr to angstrom
11  coords[2] += 1.5     # Define gold atom position
12
13  # Add gold atom
14  gold_atom = Atoms('Au', position=[coords])
15  gold_frag = convert_atoms_to_frag(gold_atom, connect_mode=None)
16  frag.append(gold_frag)
17
18  # Rename gold atom to reflect QM region (region 1) status
19  frag.names[frag.natoms-1] = 'Au1'
20
21  # Save fragment
22  frag.save('cluster_with_gold_partitioned.pun', 'pun')
```

Listing 2.2. Python script used to add a gold atom 1.5 Å above a specified atom (atom 7) of the saved partitioned cluster from Listing 2.1.

In this thesis, Py-ChemShell is used to couple the FHI-aims and GULP software packages. Listing 2.3 details a Python script that was used to optimise partitioned clusters using QM/MM. FHI-aims settings can be provided via an `fhiaims.settings` file, while the `fhiaims.basis` file includes basis sets for all atoms.

```
1   from chemsh import *
2
3   # Load partitioned cluster
4   frag = Fragment(coords="cluster_with_gold_partitioned.pun", connect_mode='covalent',
    ↪   connect_toler=1.3)
5
6   # QM settings (FHI-aims)
7   qm_region = frag.getRegion(1)
8   qm = FHIaims(settings="fhiaims.settings", basis="fhiaims.basis",
    ↪   ghost_species='Emptium')
9
10  # MM settings (GULP)
11  ff = 'brenner'  # Specify forcefield
12  mm = GULP(ff=ff)
13
14  # QM/MM settings
15  qmmm = QMMM(frag=frag, qm_region=qm_region, qm=qm, mm=mm, embedding='mechanical',
    ↪   coupling='covalent')
16
17  # Run optimisation
18  opt = Opt(theory=qmmm, algorithm='lbfgs', maxcycle=1000,
    ↪   active=frag.getRegion(1,2,3))
19  opt.run()
```

Listing 2.3. Python script used to optimise a partitioned cluster using QM/MM.

Before the 21.0 release, Py-ChemShell users were able to only parse `control.in`-specific keywords into an `FHIaims()` object, but there was no infrastructure to parse in `geometry.in`-specific keywords. Examples of such keywords include `initial_moment` and `initial_charge`, which allow for an initial spin moment or charge, respectively, to be placed on a particular atom. To allow for the parsing of these keywords, two new variables were introduced into the `FHIaims()` class: `initial_moment` and `initial_charge`. These keywords allow for users to enter a dictionary such as `initial_moment = {A: x,...}` or `initial_charge = {A: x,...}`, into an `FHIaims()` object, where `A` is either an integer that corresponds to the atom number (e.g. `1`), or is a string that corresponds to the atom label (e.g. 'Au'), and `x` is a real number denoting the initial spin moment or initial charge on atom `A`. This implementation of initial spin moments and initial charges by the author can be found within the 21.0 release of Py-ChemShell, and the changes to the source file can be found in Listing A.6.

Similarly, before the 21.0 release, users were not able to run constrained optimisations using Py-ChemShell. A new variable named "`frozen_partial`" was introduced within the `chemsh/dl_find/opt.py` file of the software, which allows users to enter a list of tuples e.g. `frozen_partial = [(A, ax),...]` within the `Opt()` class, where `A` is either an integer that corresponds to the atom number (e.g. `1`), or is a string that corresponds to the atom label (e.g. 'C'), and `ax` is a string containing the axes that should be frozen (e.g. 'xy') for atom `A`. For a fragment with n atoms, an $n \times 3$ array, F, was subsequently written, where every row comprised a three-dimensional vector of binary numbers, with `1` pertaining to unfrozen axes and `0` corresponding to frozen axes. Amendments to the `chemsh/dl_find/opt.py` by the author can be found in Listing A.7. To implement constrained optimisations, the gradients along the frozen axes (for every atom) had to be equated to zero. Gradients, G, are calculated by calculators and invoked as single-point energy evaluations by the `chemsh/dl_find/callback.py` file within Py-ChemShell. Gradients were therefore zeroed after these single-point calculations in the `dlf_get_gradient()` function within the `chemsh/dl_find/dlf_routines.f90` file in Py-ChemShell, which passes the gradient values back to DL-FIND [340]. To obtain the new gradients, G', that accounted for the frozen axes, a Hadamard product between G and F was taken i.e. $G' = F \odot G$. The amended Fortran subroutine can be found in Listing A.8.

2.7.6 PyMOL

The PyMOL [341] software package (website: https://pymol.org/) was used throughout this work to visualise and render structures.

2.7.7 Python Libraries

Various Python libraries were used throughout this work to analyse and visualise data. The NumPy [342] library was used extensively for data processing and analysis, while the Scikit-learn [343] and SciPy [344] libraries were used for statistical analysis. The Matplotlib [345] library was used throughout to plot graphs and figures, while Python scripts from the `logsdail/carmm` [346] GitHub repository (website: https://github.com/logsdail/carmm) were also used to analyse FHI-aims output files.

Chapter 3

Oxygenation State of the Diamond (110) Surface

The following chapter presents work published in the paper referenced below. Section 3.3.2 includes XPS simulation results which were conducted by Samuel J. Hall (University of Warwick), and results from XPS experiments which were conducted by Dr. Marc Walker (University of Warwick) and in collaboration with Dr. Benedikt P. Klein (University of Warwick, Diamond Light Source). Section 3.3.3 includes results from X-ray diffraction experiments, which were conducted by Dr. David Walker (University of Warwick), low-energy electron diffraction experiments, which were conducted by Dr. Marc Walker, and atomic force microscopy experiments, which were conducted by Dr. Julie Macpherson (University of Warwick).

S. Chaudhuri, S. J. Hall, B. P. Klein, M. Walker, A. J. Logsdail, J. V. Macpherson, R. J. Maurer (2022) 'Coexistence of carbonyl and ether groups on oxygen-terminated (110)-oriented diamond surfaces' *Commun. Mater.* **3**: 6 [347]

3.1   Introduction

When grown via CVD [144], the more popular synthesis method for technological applications , diamond surfaces usually leave the growth chamber terminated with hydrogen species [145]. However, processes such as mechanical polishing or chemical cleaning, which typically use strong oxidising agents, render the surface oxygen-terminated. Even if untreated, the hydrogen-terminated surface will gradually oxidise over time if exposed to air [348; 349]. Oxygen-terminated diamond surfaces have been shown to exhibit hydrophilicity [350; 351] and a positive electron

affinity [352]. Oxygen terminations can also cause a reduction in surface electrical conductivity [348], an increase in capacitance [353], and impact the electron-transfer kinetics of inner-sphere electrochemical reactions on BDD electrodes [131]. As explained in Section 1.2.2, polycrystalline BDD surfaces can be treated as a textured surface with a dominant (110) orientation [22; 156], which means a detailed atomic-scale understanding of the oxygen termination and elementary composition of diamond (110) surfaces is of vital importance. Knowledge of the diamond (110) surface oxygenation state would allow for the growth of gold NCs on BDD surfaces to be studied, as is the aim of this thesis.

The oxygenation state of the diamond (110) surface has not been extensively characterised, with previous studies primarily focusing on the (111) and (100) surfaces [152], most likely due to the challenges associated with growing and preparing a (110)-oriented single-crystal diamond surface with a large-enough area at a high-enough quality [139]. Computational studies have identified the presence of both carbonyl (C=O) and ether (C–O–C) functional groups on the diamond (111) and (100) surfaces [354; 355; 356; 357; 358; 359; 360; 361; 362; 363; 364], while experimental techniques, such as infrared and X-ray photoelectron spectroscopy (XPS), also provide an indication of the coexistence of different oxygenic moieties on diamond surfaces. Experimental studies on the (111) and (100) surfaces indicate the presence of both carbonyl and ether groups [365; 366; 367; 368] after the removal of loosely-bound compounds and organic contaminants. The few experimental studies that exist for the (110) surface are also interpreted in the context of a coexistence of carbonyl and ether groups [152; 369; 370; 371]. Peroxide (C–O–O–C) groups have also been found to exist on the (110) surface at low temperatures using infrared spectroscopy [152]. However, in two of these studies, the (110) surfaces were subjected to a hydrogen plasma treatment before analysis [152; 371], which is known to induce the formation of (111) microfacets [152; 371; 372] and compromise the (110) crystallinity of the surface. Makau and Derry only presented XPS data recorded at room temperature, where the physisorbed contaminants may still be present and no details about the quality of the studied crystal were provided [369]. Baldwin *et al.* [370] assumed the crystal contained a mixture of (110) and (111) facets and a large number of step edges, with a measured surface roughness of 1 nm. These studies highlight the challenges of preparing high-quality (110) surfaces.

This chapter will present a systematic characterisation of the structure and stability of various oxygenic terminations of the diamond (110) surface. This is

achieved through a combination of DFT [65; 66] calculations and XPS experiments (at elevated temperatures) on a homoepitaxial (110)-oriented BDD substrate grown via CVD. Using *ab initio* atomistic thermodynamics [335; 373; 374] based on DFT [65; 66], as described in Section 2.6, a phase diagram of a large number of surface structures is established and the most stable oxygenation state of the (110) surface is identified.

3.2 Methods

3.2.1 Computational Structure Search

DFT [65; 66] calculations were conducted using the FHI-aims [70] software package and a 'tight' basis set [259]. Unless otherwise stated, all calculations were conducted using the PBE [229] GGA, though a few calculations were also conducted using the revPBE [231], RPBE [233] and BLYP [375; 376] GGAs, and the TPSS [234] MGGA ; these instances are clearly stated within the text. The pairwise TS [262] dispersion correction, as described in Section 2.2, was used to account for long-range effects in GGA calculations. A few calculations were also conducted using the MBD@rsSCS [287] and MBD-NL [288] schemes ; these instances are clearly stated within the text.

Structures were symmetric with oxygenic species terminating both the top and bottom of a seven-layer dipole-cancelling diamond (110) slab. To account for cases where only partial oxidation might occur, the following surface unit cells were optimised and studied: $p(1 \times 1)$, $p(2 \times 1)$, $p(1 \times 2)$ and $p(2 \times 2)$. A reciprocal space grid of size $16 \times 16 \times 1$ centred around the Γ point of the first Brillouin zone was used for $p(1 \times 1)$ surface calculations. Input and output files for all calculations have been uploaded as a dataset [377] to the NOMAD electronic structure data repository [378] and are freely available under https://doi.org/10.17172/NOMAD/2021.03.01-1.
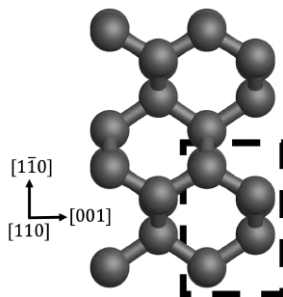
Figure 3.1.

Equation (3.1) was used to calculate the oxygen adsorption energy, E_{ads}, of different surface phases:

$$E_{\text{ads}} = \frac{1}{N}(E_{\text{total}} - E_{\text{slab}} - NE_{\text{O}}), \tag{3.1}$$

where N is the number of oxygen atoms, E_{total} is the total energy of the terminated surface and E_{slab} is the energy of the surface onto which the oxygenic species is being adsorbed. The oxygen adsorbate energy, E_{O}, was calculated as half the total energy of a free oxygen molecule (O_2), while for hydroxyl adsorbates, E_{O} was calculated as the difference between the energy of a free water molecule and half the energy of a free hydrogen molecule. The adsorption energy per unit cell surface area, E_{ads}/A, was calculated by multiplying E_{ads} by N/A.

Using the calculated E_{ads} values, the Gibbs free energy of adsorption, ΔG_{ads}, for every termination was calculated as a function of the gas-phase chemical potential of oxygen, $\Delta\mu_{\text{O}}$, which represents the ease of oxygen adsorption onto the surface [335; 374], using Equation (2.61). Here, the surface was assumed to be in equilibrium with an oxygen atmosphere that behaves as an ideal gas. $\Delta\mu_{\text{O}}$ was then further expressed as a function of both temperature, T, and pressure, p, using Equation (2.59) [335]. $\Delta\mu_{\text{O}}(T, p^{\ominus})$ was calculated using a quartic polynomial fit, as expressed in Equation (3.2), to tabulated enthalpy and entropy data for molecular oxygen at standard pressure from thermochemical tables [336]. A qualitative comparison of Equation (3.2) to the thermochemical data can be seen in Figure 3.2; quantitatively, the coefficient of determination value of Equation (3.2) was evaluated to be 1.00, showcasing the high quality of the fit.

$$\Delta\mu_{\text{O}}(T, p^{\ominus}) = \beta T^4 + \gamma T^3 + \epsilon T^2 + \eta T + \zeta, \tag{3.2}$$

where $\beta = -3.062 \times 10^{-13}$, $\gamma = 8.957 \times 10^{-10}$, $\epsilon = -1.156 \times 10^{-6}$, $\eta = -1.332 \times 10^{-3}$ and $\zeta = -1.552 \times 10^{-1}$.
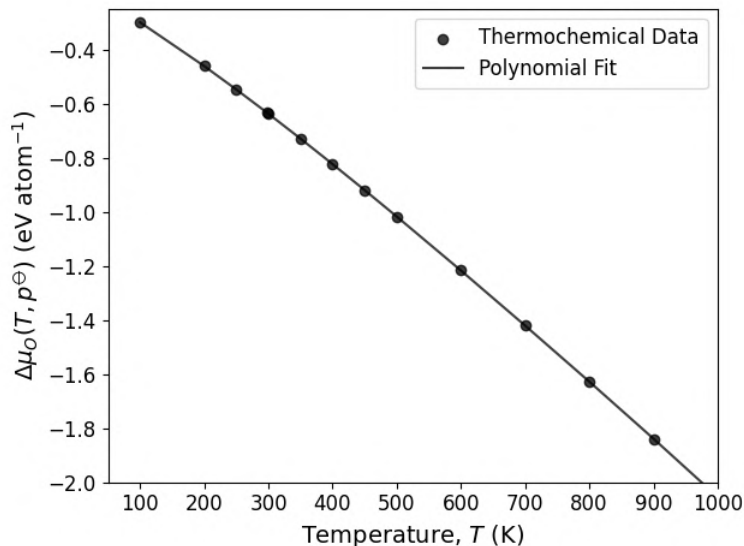
Figure 3.2. Graphical comparison of the fit expressed in Equation (3.2) to $\Delta\mu_O\left(T, p^{\ominus}\right)$ values calculated from tabulated entropy and enthalpy values for molecular oxygen [336].

3.2.2 Core-Level Binding Energy Simulations

This subsection details the methods used by Samuel J. Hall to conduct core-level binding energy simulations, and is included here for completion. The cutting and hydrogen-saturation of clusters were conducted by the author.

XPS core-level binding energies were calculated by Samuel J. Hall using the Delta Self-Consistent Field [379; 380] method by taking the energy difference between two self-consistent Kohn-Sham DFT calculations: the ground-state calculation and the excited core-hole-constrained calculation, where the population of an atom (either carbon or oxygen) was constrained electron. Both periodic and aperiodic core-hole calculations were performed using the PBE [229] GGA. Periodic calculations were performed using the plane-wave pseudopotential CASTEP [71] software package, with default on-the-fly generated ultrasoft pseudopotentials, an electronic energy tolerance of $1 \times 10^{-6}\,\mathrm{eV\,atom^{-1}}$ and a plane-wave cut-off of $650\,\mathrm{eV}$. Here, the core-hole was localised through the generation of a modified pseudopotential with a core-hole included on the target atom. The energetic contribution of core electrons was accounted for via a correction term when calculating binding energies between the ground and excited states [381]. Supercells of various sizes were constructed to systematically assess binding energy convergence and the role of finite size effects.

Aperiodic calculations were also performed by Samuel J. Hall using the FHI-aims [70]. The `force_occupation_basis` keyword was used to constrain the core-hole, along with additional basis functions to better describe the core-states, as laid out by Kahk and Lischner [382]. Here the constraint was defined in terms of a localised atomic orbital basis function, with the Kohn-Sham eigenstate with the highest contribution also being constrained. A variation of the maximum overlap method [383] ensured the constraint remained satisfied. The aperiodic calculations were performed on hemispherical cluster models of various phases, cut from their parent periodic models using the Py-ChemShell [301] software package, as explained in Listing 2.1. Dangling bonds for all atoms were terminated with hydrogen species, though the XPS calculations were conducted on the central carbon and oxygen atoms. Visualisations of the hydrogen-saturated clusters can be found in Figures B.2. Using a Mulliken analysis [384], the density of states of each cluster was compared against its parent periodic model, which are shown in Figure B.3. Cluster models, unlike periodic models, allow for the prediction of absolute binding energies [382; 385] and can therefore be used to validate the relative binding energy shifts calculated using periodic models. Once convergence with respect to model size was established [385], the majority of calculations were conducted with periodic models via CASTEP.

3.2.3 Surface Science Experiments

This subsection details the experimental methods that were used to prepare and analyse a diamond sample provided by Element Six Ltd., and is included here for completion. XPS and low-energy electron diffraction experiments were conducted by Dr. Marc Walker, X-ray diffraction experiments were conducted by Dr. David Walker, and atomic force microscopy experiments were conducted by Dr. Julie Macpherson.

Sample Preparation

A (110)-oriented single-crystal BDD sample (Element Six Ltd.) of size $5.00 \times 5.00 \times 0.47$ mm^3 was homoepitaxially grown via microwave-plasma CVD from a (110)-oriented single-crystal diamond substrate. The growth chamber comprised a $\sim 5\%$ methane and hydrogen gas atmosphere at temperatures of 700–950 °C and pressures of 100–200 Torr, with microwave powers of 1-3 kW at 2.45 GHz [145]. The sample was highly boron-doped [131] to low 10^{20} atoms per cubic centimetre, and boron doping was achieved with the addition of up to 0.02% diborane(6) to the gas phase

during CVD growth. A highly boron-doped sample was used to reduce charging effects within the XPS experiments, the dopant concentration was low enough the surface termination of the diamond sample [386]. The sample was removed from the growth substrate via laser micromachining. A small amount of twinning was observed at the edge of the crystal, which compromises the surface quality and can impact polishing. The twinning was removed by reducing the crystal size to $3.25 \times 3.25 \times 0.47$ mm^3 using laser micromachining. To smooth the growth surface, the (110) surface was first mechanically lapped using a resin bonded wheel and then scaife cast iron polished. Prior to any analysis, the sample was acid-cleaned for 30 min at $\sim$200 °C in concentrated sulphuric acid saturated with potassium nitrate, rinsed with deionised water and cleaned again for 30 min at $\sim$200 °C in concentrated sulphuric acid [348]. The sample was finally ultrasonically cleaned for 15 min in deionised water and left to air-dry. All solutions were prepared from Milli-Q water (Millipore Corp.) with a resistivity of 18.2 MΩ cm at 25 °C.

X-Ray Photoelectron Spectroscopy

An Omicron Multiprobe instrument was used to acquire XPS data, with the sample mounted onto stainless steel Omicron flag plates using tantalum foil and loaded into a fast-entry chamber. Once a pressure of less than 1×10^{-6} mbar was achieved (which took $\sim$1 h), the sample was transferred to a 12-stage storage carousel located between the preparation and main analysis chambers for storage at pressures of less than 2×10^{-10} mbar. XPS measurements were conducted in the main analysis chamber (base pressure 2×10^{-11} mbar) with the sample being illuminated using an XM1000 monochromatic aluminium K$_\alpha$ X-ray source (Omicron NanoTechnology), which corresponds to a photon energy of 1486.7 eV. Photoelectrons were detected using a Sphera electron analyser with survey scans acquired using a 50 eV pass energy and a 0.5 eV step size. Selected high resolution core-level spectra were recorded using a 10 eV pass energy (resolution approximately 0.47 eV) with a 0.1 eV step size. Measurements were conducted at two different electron emission angles, 0° and 60° with respect to the surface normal (90° and 30° with respect to the sample surface respectively). The spectrometer work function was calibrated using a clean polycrystalline silver sample immediately prior to experiments commencing. The analyser transmission function was calibrated using clean silver, gold and copper foils. In order to ensure the removal of any surface-adsorbed contaminants, XPS measurements were taken after annealing the sample to different temperatures. Sample heating was performed in the main analysis chamber using a resistive heater and monitored by a K-type thermocouple attached to the manipulator head. Spectra

were fitted using symmetric pseudo-Voigt profiles [387; 388], and if several peak components were present in a spectrum, they were restricted to have the same widths and Gaussian-Lorentzian ratios.

X-Ray Diffraction

X-ray diffraction experiments were conducted by Dr. David Walker using a PANalytical X'Pert Pro MRD instrument with a 4-bounce germanium (220) hybrid monochromatorwhich gave out pure copper $K_{\alpha 1}$ radiation, and a Pixcel detector was set as a point detector using the whole 14 mm detector width. A pole figure was recorded at a 75.3° 2θ reflection position (where θ is the Bragg angle) with a wide open detector, and by rotating the sample at different azimuthal angles and a number of different tilt angles.

Low-Energy Electron Diffraction

Low-energy electron diffraction experiments were conducted by Dr. Marc Walker in ultra high vacuum conditions using a SPECTALEED optics (Omicron NanoTechnology). Diffraction patterns were taken at various electron energies after degassing the sample to 300 °C.

Atomic Force Microscopy

Atomic force microscopy experiments were conducted by Dr. Julie Macpherson on different locations of the surface using a Bruker Dimension Icon microscope operating in ScanAsyst$^{\text{TM}}$ mode. The resulting data was processed using the Gwyddion [389] software.

3.3 Results and Discussion

3.3.1 Structure and Stability of Oxygen-Terminated Phases

An extensive computational structure search was performed to identify the atomically-flat oxygen-terminated (110) surfaces with the highest stabilities, varying both the termination and surface coverage of oxygen. The search involved placing surface-terminal carbonyl, ether, peroxide, and hydroxy (C–OH) groups, as well as combinations thereof, on a diamond (110) surface slab and performing geometry optimisations. 22 different optimised surface terminations were considered, which are visualised along with their calculated adsorption energies and fractional oxygen surface coverages in Table B.1. Table 3.1 summarises the adsorption energies of the

most stable phases, which are also depicted in Figure 3.3; in both, three- and five-ring ether groups are denoted as 'Ether$_3$' and 'Ether$_5$', respectively, and carbonyl groups are denoted as 'Keto' groups. The most stable phases are found to contain carbonyl, ether, and peroxide functional groups, which agrees with existing experimental literature [152; 369; 370; 371]. In particular, phases containing coexistent carbonyl and ether groups appear to have the largest adsorption energy, with the $p(2 \times 1)$Keto-Ether$_5$ phase having by far the largest E_{ads} and E_{ads}/A values of $-3.03 \, \text{eV atom}^{-1}$ and $-0.34 \, \text{eV Å}^{-2}$, respectively. The $p(2 \times 2)$Keto-Ether$_5$ phase, a half-coverage form of the $p(2 \times 1)$Keto-Ether$_5$ phase, is the most stable phase at lower coverage. To ensure results with the 7-layered slab were converged, the adsorption energies of the $p(2 \times 1)$Keto-Ether$_5$ and $p(1 \times 1)$Peroxide surface phases were recalculated using a 9-layered diamond (110) slab and evaluated to be $-3.03 \, \text{eV atom}^{-1}$ and $-1.42 \, \text{eV atom}^{-1}$ respectively, which are identical to the 7-layered E_{ads} values in Table 3.1.

Phase	E_{ads}/A (eV Å^{-2})	E_{ads} (eV atom^{-1})	Coverage (atoms nm^{-2})
$p(1 \times 1)$Peroxide	−0.31	−1.42	22.2
$p(2 \times 1)$Keto-Ether$_5$	−0.34	3.03	11.1
$p(2 \times 1)$Peroxide	−0.23	−2.04	11.1
$p(1 \times 1)$Ether$_3$	−0.20	−1.77	11.1
$p(1 \times 1)$Ether$_5$	−0.17	−1.52	11.1
$p(2 \times 2)$Keto-Ether$_5$	−0.25	−2.25	5.5
$p(2 \times 2)$Peroxide	−0.10	−1.78	5.5

Table 3.1. Adsorption energies for the most stable oxygen-terminated diamond (110) surface phases. Energies are given as a function of surface area and per atomic adsorbate, as well as the associated surface coverage, for the most stable oxygen-terminated (110) surface phases. Entries are ordered from high to low oxygen surface coverage. A full table for all investigated surface phases can be found in Table B.1.

Figure 3.3 shows the convex hull of the calculated ΔG_{ads} as a function of the gas-phase $\Delta \mu_{\text{O}}$; a full free energy plot featuring all the phases can be found in Figure B.1. Here, the oxygenated surface is considered to be in equilibrium with a gas-phase molecular oxygen at temperature and pressure conditions, (T, p), that correspond to a certain chemical potential (for details see Section 3.2.1). The structure with the largest ΔG_{ads} value (note the inverted vertical axis) is the most stable at the given $\Delta \mu_{\text{O}}$ value, and therefore (T, p) couple, and will be thermodynamically realised under these conditions (notwithstanding any kinetic barriers) [335; 374]. Over

the shown range of $\Delta\mu_O$ values, the following three surface structures constitute the convex hull, i.e. they are the most thermodynamically stable under these conditions: the clean (110) surface, the $p(2 \times 1)$Keto–Ether$_5$ and the $p(1 \times 1)$Peroxide phases. The unterminated surface is the most stable configuration below $\Delta\mu_O = -3.03$ eV atom^{-1}. As $\Delta\mu_O$ increases further, the $p(2 \times 1)$Keto–Ether$_5$ phase is the most stable until $\Delta\mu_O = 0.29$ eV atom^{-1}. Above this , the $p(1 \times 1)$Peroxide phase becomes the most stable configuration, which is to be expected as the free energy of oxygen adsorption increases with $\Delta\mu_O$. This results in an increase in the stability of oxygen-rich phases relative to oxygen-poor surfacesthe $p(1 \times 1)$Peroxide phase possesses the maximal oxygen coverage. Visualisations of the clean surface, $p(2 \times 1)$Keto–Ether$_5$and $p(1 \times 1)$Peroxide phases are given in Figure 3.4.
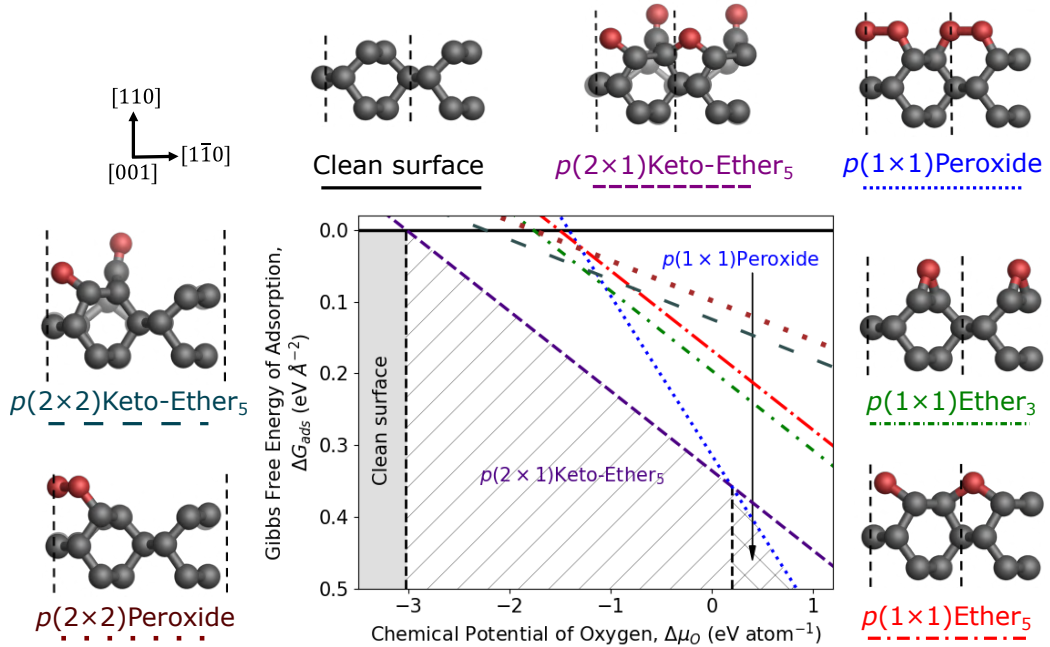
Figure 3.3. Free energy plot for the most stable oxygen-terminated diamond (110) surface phases, as listed in Table 3.1. The shaded region represents the convex hull of ΔG_{ads}, which is presented as a function of $\Delta\mu_O$ in the gas phase. A full free energy plot of all phases can be seen in Figure B.1. Orthographic ball-and-stick visualisations of the first three carbon layers of the phases are shown as viewed 'front on' from the [001] direction, with carbon and oxygen atoms shown in grey and red respectively. Unit cell outlines are shown in black dashed lines with surface axes also presented.

The range of temperatures and pressures at which the most stable phases are realised can be seen in Figure 3.5. Over the temperature range of 0–1000 K, at standard pressure, the $p(2 \times 1)$Keto–Ether$_5$ phase is the most stable configura-
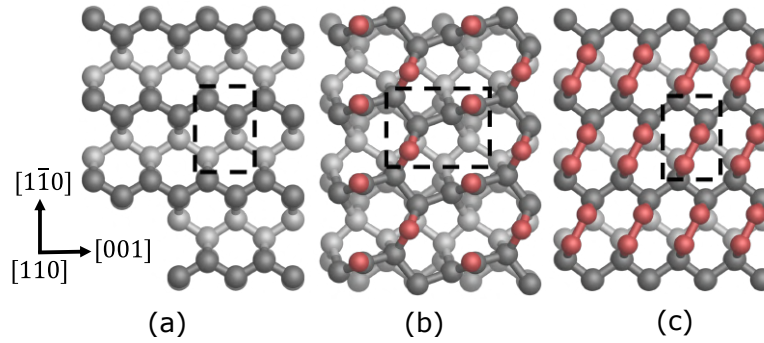
Figure 3.4. Orthographic ball-and-stick visualisations of the three most stable oxygen-terminated diamond (110) surface phases. (a) Clean surface, (b) $p(2 \times 1)$Keto-Ether$_5$ and (c) $p(1 \times 1)$Peroxide, shown as a 'top-down' view from the [110] direction. Carbon and oxygen atoms are shown in grey and red respectively. Unit cell outlines are shown in black dashed lines with surface axes also presented.

tion, and expected at standard atmospheric conditions. The unterminated surface is the most stable phase at higher temperatures ($T > 1000\,\mathrm{K}$) and lower pressures ($p < 10^{-6}$ atm), which is expected as the stability of an oxygen ideal gas increases with increasing T and decreasing p. The $p(1 \times 1)$Peroxide phase is the most stable phase at lower temperatures, with suprema ranging from around $60\,\mathrm{K}$ to around $100\,\mathrm{K}$ over the whole pressure range.
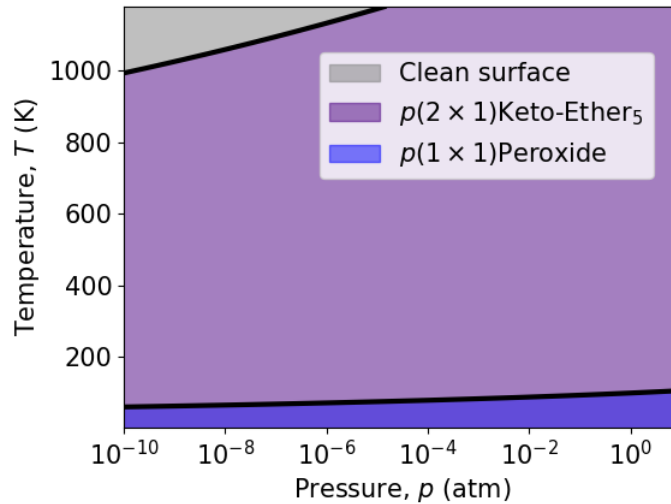
Figure 3.5. Phase diagram of the oxygen-terminated diamond (110) surface. The diagram is presented as a function of both temperature, T, and pressure, p, and shows the three most stable oxygen-terminated diamond (110) surface phases.

Experimental studies on the oxygenation state of the (110) surface provide

56

evidence of oxygen species that broadly supports the results presented above. However, the aforementioned caveat concerning the quality of the surfaces used in these studies [152; 369; 370; 371] should be noted. These studies report carbonyl and ether groups on the surface [152; 369; 370; 371], while surface peroxide species are reported to exist at low temperatures (90 K) [152]. However, the *ab initio* atomistic thermodynamics method used herein does not consider any kinetic information, such as reaction barriers, that are likely to affect which phases are experimentally observed under conditions. Vibrational contributions to ΔG_{ads} were also not considered here, comprising the vibrational entropy and enthalpy [335; 373], which can shift the stability boundaries to some extent. These contributions are unlikely to affect the convex hull though as the Keto-Ether$_5$ phases are significantly more stable than all other considered phases.

As an additional note, the PBE [229] GGA was employed to optimise surface phases, and is known to underestimate adsorption energies and overestimate adsorption distances in many cases [68; 69; 88]. To assess the effect of the choice of DFA, surface structures were reoptimised, and adsorption energies recalculated with different DFAs for the $p(2 \times 1)$Keto–Ether$_5$ and $p(1 \times 1)$Peroxide phases. Table 3.2 details the different E_{ads} values with various DFAs. The DFAs that were compared against PBE [229] were the revPBE [231], RPBE [233] and BLYP [375; 376] GGAs, and the TPSS [234] MGGA. It should be noted that the TS [262] dispersion correction scheme was not used alongside TPSS due to the MGGA already accounting for a certain level of mid-range interactions within its formulation [234]. As can be seen, the absolute adsorption energy values strongly depend on the DFA, with E_{ads} varying within a range of $0.4\,\mathrm{eV\,atom}^{-1}$ for both phases. The DFAs fall into two classes with respect to the relative stability of the two phases that they predict. Nevertheless, the ordering of the two phases and the significantly higher stability of the $p(2 \times 1)$Keto-Ether$_5$ phase are independent of the DFA.

DFA	$p(2 \times 1)$ Keto-Ether$_5$	$p(1 \times 1)$ Peroxide	ΔE_{ads}
PBE [229]	3.03	1.42	1.61
revPBE [231]	2.61	1.55	1.06
RPBE [233]	2.83	1.15	1.68
BLYP [375; 376]	2.62	1.57	1.05
TPSS [234]	3.04	1.41	1.63

Table 3.2. Table showing the E_{ads} values, in electronvolts per atom, for the $p(2 \times 1)$Keto-Ether$_5$ and $p(1 \times 1)$Peroxide phases as calculated using different DFAs, along with the difference in E_{ads} values, ΔE_{ads}.

The choice of dispersion correction scheme was also assessed by reoptimising surface structures with two MBD [285; 286] schemes and recalculating the adsorption energies. Table 3.3 details the different E_{ads} values with the TS [262], MBD@rsSCS [287] and MBD-NL [288] schemes based on the PBE [229] GGA. As can be seen, the absolute adsorption energy values have a very minor dependence on the dispersion correction, but the ordering of the two phases and the significantly higher stability of the $p(2 \times 1)$Keto-Ether$_5$ phase are also independent of the dispersion correction.

Dispersion Correction	$p(2 \times 1)$ Keto-Ether$_5$	$p(1 \times 1)$ Peroxide	ΔE_{ads}
TS [262]	3.03	1.42	1.61
MBD@rsSCS [287]	2.98	1.43	1.55
MBD-NL [288]	3.01	1.45	1.56

Table 3.3. Table showing the E_{ads} values, in electronvolts per atom, for the $p(2 \times 1)$Keto-Ether$_5$ and $p(1 \times 1)$Peroxide phases as calculated using different dispersion correction schemes based on the PBE [229] GGA, along with the difference in E_{ads} values, ΔE_{ads}.

In the following, the origin of the high stability of the Keto-Ether$_5$ phases is addressed. During the structure optimisations, the Keto-Ether$_5$ phases form spontaneously from manually-prepared peroxide groups aligned along the [1$\bar{1}$0] direction. A schematic of the optimisation mechanism is shown in Figure 3.6. The initially-adsorbed peroxide homolytically dissociates to form two oxygen radicals (O·), one of which is saturated by forming a carbonyl group with a surface carbon atom at the expense of a carbon-carbon bond between the first and the second layer, while the other is stabilised by bridging the ridge to form a five-ring ether, resulting in

the $p(2 \times 1)$Keto-Ether$_5$ phase. The $p(2 \times 1)$Keto-Ether$_5$ structure has previously been verified to be the most stable structure via a global structure search [22].
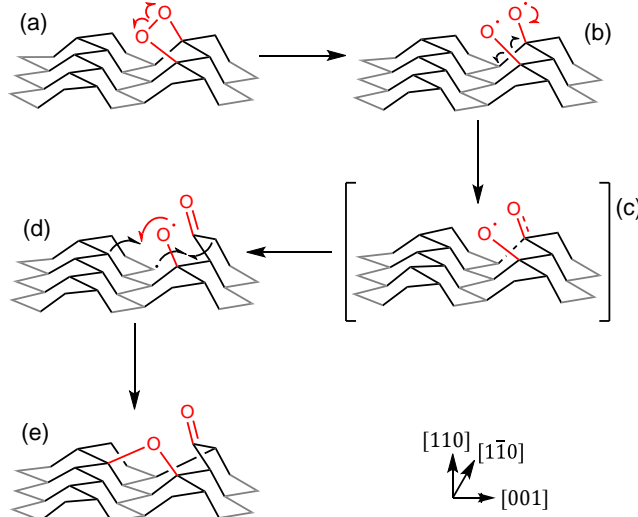
Figure 3.6. Structural rearrangement during the formation of the $p(2 \times 1)$Keto-Ether$_5$ phase. The initial peroxide termination is along the C–C ridge in the [1$\bar{1}$0] direction. First, (a) the O–O bond breaks to form two oxygen radicals (O·), one of which then proceeds to form a double bond with a surface carbon atom (b). Due to the propensity of carbonyl groups to form trigonal planar structures, one of the C–C bonds in the tetrahedral diamond structure also breaks. Both processes are underway in the transition state (c). In order to form a stable structure, the remaining O· bridges over the surface ridge to form an ether group, allowing the carbon radical (C·) to bond with another carbon atom behind the carbonyl (d). The resulting structure is the $p(2 \times 1)$Keto-Ether$_5$ phase (e). Surface axes are also presented.

To further understand the large adsorption energy of $-3.03 \, \mathrm{eV \, atom^{-1}}$, the structure of the $p(2 \times 1)$Keto-Ether$_5$ phase is examined against other surface phases by analysing bond lengths and bond angles. From calculations, carbonyl groups are found to lead to some lattice distortion, with C–C=O bond angles of 123.2° and 125.0° in the $p(2 \times 1)$Keto-Ether$_5$ phase, which indicates a more trigonal planar surface structure. Based on valence shell electron pair repulsion theory, partial sp^2 hybridisation has occurred in the surface carbon layers, which is consistent with partial graphitisation and sp^2 hybridisation in diamond [137; 390; 391; 392]. Further geometrical distortions can be seen by comparing carbon-oxygen bond lengths. Other surface terminations that feature ether, peroxide or hydroxy surface groups seem to preserve the sp^3 hybridisation of the carbon atoms, with a C–O bond length of typically around 1.45 Å, with the exception of 1.55 Å in Ether$_5$ phases; in com-

parison, the C=O bond length in the Keto-Ether$_5$ phases is 1.20 Å, shorter than the observed C–O bonds. The formation of the C=O bond pulls the respective carbon atom above the diamond surface plane, as can be seen in the Keto-Ether$_5$ structures shown in Figure 3.3.

While C=O bonds are typically stronger than C–O bonds, individual surface carbonyl groups cannot be realised on a pristine (110) surface as they would require the breaking of C–C bonds, creating carbon radicals and leading to a substantial deformation of the surface. To circumvent that, another surface moiety such as an ether group must be present to saturate the carbon radicals and minimise the necessary deformation, as shown in Figure 3.6. The adjacent carbonyl and ether functionalities in the Keto-Ether$_5$ phases possess such a high stability because their local pairing is able to balance the high adsorption energy from the carbonyl with low structural deformation in the surface. Taking the distorted surface slab upon adsorption of oxygen in the $p(2 \times 1)$Keto-Ether$_5$ phase to be E_{slab} in Equation (3.1), the E_{ads} value is recalculated to be $-5.26\,\mathrm{eV\,atom^{-1}}$. Therefore, the structural deformation penalty associated with adsorption is $2.23\,\mathrm{eV\,atom^{-1}}$. Without the ether group, this deformation penalty would be much higher. The fact that the geometry optimisation of an adsorbed peroxide leads to the formation of the Keto-Ether$_5$ phase, suggests that the pathway shown in Figure 3.6 can be a viable oxidation mechanism of the diamond (110) surface.

3.3.2 Comparison to X-Ray Photoelectron Spectroscopy

To connect the predicted structures with experimental data, the core-level binding energies of the central carbon and oxygen species in the DFT-based structures were simulated by Samuel J. Hall. Convergence tests of the XPS binding energies were carried out on three increasing cluster sizes and unit cells for the $p(2 \times 1)$Keto-Ether$_5$ phase. Figure 3.7 shows convergence behaviour for the $p(2 \times 1)$Keto-Ether$_5$ phase, while Table 3.4 details the numerical values for this phase. Figure 3.7 shows that both the cluster and periodic calculations converge to the same value (0.8 eV). Finite-size and cell-charging effects can therefore be ruled out from affecting the relative binding energy shifts obtained via periodic calculations. This agreement between periodic and aperiodic models also verifies that the saturating hydrogen species in the aperiodic models do not influence the XPS behaviour of the central carbon and oxygen atoms. All binding energies reported hereafter were obtained from the converged periodic models.
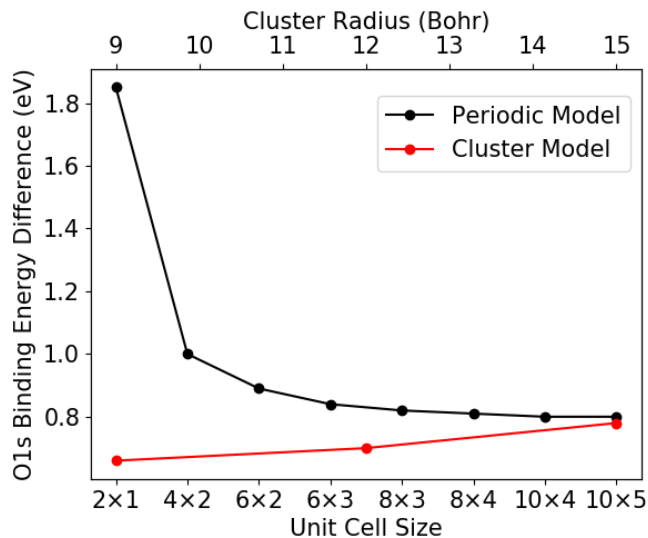
Figure 3.7. Graph showing the convergence of the simulated oxygen $1s$ (O $1s$) binding energy difference between the carbonyl and ether groups in the $p(2 \times 1)$Keto–Ether$_5$ phase using periodic and cluster models.

Unit Cell Size	ΔC $1s$ Binding Energy (eV)			O $1s$ Binding Energy (eV)		
	$\underline{C}$=O	$\underline{C}$–O–C	C–O–$\underline{C}$	C=$\underline{O}$	C–$\underline{O}$–C	Difference
2×1	6.59	7.52	7.15	547.33	549.18	1.85
4×2	2.33	2.41	2.16	543.24	544.24	1.00
6×2	1.88	1.85	1.61	542.80	543.69	0.89
6×3	1.59	1.52	1.29	542.53	543.37	0.84
8×3	1.44	1.35	1.12	542.39	543.21	0.82
8×4	1.34	1.25	1.02	542.30	543.11	0.81
10×4	1.28	1.18	0.96	542.25	543.05	0.80
10×5	1.24	1.15	0.92	542.22	543.01	0.79
Cluster Radius (a_0)	$\underline{C}$=O	$\underline{C}$–O–C	C–O–$\underline{C}$	C=$\underline{O}$	C–$\underline{O}$–C	Difference
9.0	291.16	290.75	290.70	535.72	536.38	0.66
12.0	291.06	290.71	290.78	535.37	536.07	0.70
15.0	–	290.64	290.41	535.18	535.96	0.78

Table 3.4. Table showing the convergence of the carbon $1s$ (C $1s$) binding energies with respect to the diamond bulk (ΔC $1s$) and O $1s$ absolute binding energies for supercells and clusters of various sizes of the $p(2 \times 1)$Keto-Ether$_5$ phase. For cluster models, where a diamond bulk value was not calculated, the absolute binding energy is shown. The entry marked '–' is a result of the calculation failing to converge.

XPS simulations were compared to experimental XPS measurements con-

ducted by Dr. Marc Walker, as well as to literature [369; 370]. The simulation results for the core-electron binding energies of the Keto–Ether$_5$ phase in two coverages, $p(2 \times 1)$ and $p(2 \times 2)$, are detailed in Tables 3.5 and 3.6. The carbon 1s (C 1s) binding energies in Table 3.5 are presented as shifts with respect to the bulk carbon signal of the diamond surface, both for experiment and simulation. In the case of the oxygen 1s (O 1s) data in Table 3.6, only the binding energy difference between the two oxygen species present in the different functional groups are discussed.

Species	Simulation		Experiment		
	$p(2 \times 1)$	$p(2 \times 2)$	Warwick	Makau and Derry [369]	Baldwin *et al.* [370]
C=O	1.16	0.74	0.7	4.5	2.2
C–O–C	1.46	0.80	0.7	1.9	1.1
C–O–C	1.23	0.30	0.7	1.9	1.1

Table 3.5. Comparison of the C 1s binding energy shifts for the Keto-Ether$_5$ phases with respect to the diamond bulk signal. For the simulation of the ether moiety, two carbon binding energies are given, one for each oxygen-bonded carbon atom. The Warwick experiments only show one peak in addition to the bulk diamond signal, while literature reports two additional peaks at a much higher binding energy.

Simulation		Experiment		
$p(2 \times 1)$	$p(2 \times 2)$	Warwick	Makau and Derry [369]	Baldwin *et al.* [370]
1.00	1.31	1.5	2.1	1.7

Table 3.6. Comparison of the O 1s binding energy difference for the Keto-Ether$_5$ phases. The binding energy difference is calculated as the difference between the binding energies of the oxygen atoms involved in the carbonyl (C=O) and ether (C–O–C) functional groups. The experimental data reports the binding energy difference between the two peaks in the corresponding O 1s spectra.

As can be seen in Table 3.5, the calculated C 1s binding energies for carbonyl and ether carbons are close together; in the lower-coverage $p(2 \times 2)$ phase, the C=O binding energy sits evenly between the two C–O–C carbons present in the structure. This result is in agreement with the experiments conducted by Dr. Marc Walker, presented in Figure 3.8, which shows the experimental C 1s and O 1s spectra recorded after annealing the sample surface to 500 °C. The C 1s spectrum in Figure 3.8(a) is dominated by a main peak with a small shoulder at a higher binding energy, caused by a small peak partly underneath the main peak.

The binding energy difference between the two peaks is 0.7 eV (0.6-0.9 eV). As can be seen in Figure 3.8(b), two species are present in the O 1s spectrum: the high binding energy peak corresponds to an ether moiety, while the low binding energy peak can be attributed to a carbonyl group. The relative shift between the two species is 1.5 eV (1.4-1.6 eV). The lower binding energy peak of the carbonyl has a higher intensity, with an intensity ratio of about 2:1. Therefore, both spectra can be fitted sufficiently using only two sub-peaks. The resulting experimental peak shifts between the two respective peaks are also included in Tables 3.5 and 3.6.
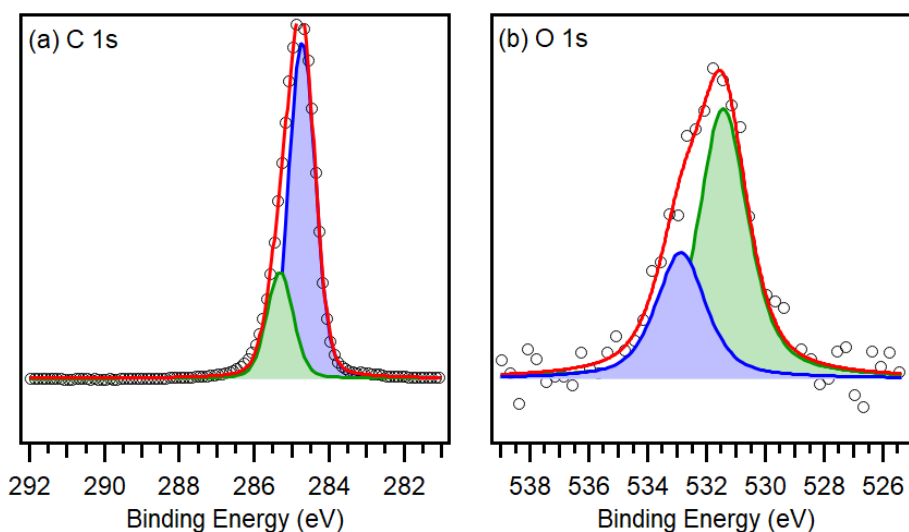
Figure 3.8. Experimental X-ray photoelectron spectra of the sample surface. (a) Carbon 1s (C 1s) spectrum and (b) oxygen 1s (O 1s) spectrum. Spectra were taken at normal emission after annealing the sample to 500 °C. The raw data, fit functions and fitted peaks are shown using open circles, red lines and shaded areas respectively.

Contrary to both DFT calculations and Warwick experiments, data published in literature [369; 370] shows two additional oxygen-related peaks at much higher binding energy. As discussed above, this discrepancy is hypothesised to be due to both crystal quality and the lack of sample annealing prior to running XPS experiments. contaminants in the atmosphere that the sample is exposed to prior to measurement can lead to a high coverage of contaminants adsorbed on the surface; these contaminants will still be present after the sample is introduced into the vacuum chamber of the photoelectron spectrometer [393]. Contaminants can cause surface reactions, introduce additional signals into the O 1s region, and also show oxygen-related peaks in the C 1s region. As can be seen in Figure 3.9(a), the XPS

survey spectrum for the non-annealed sample showed the expected C $1s$ and O $1s$ signals, but small nitrogen $1s$ (N $1s$) and sodium $1s$ (Na $1s$) signals were also observed. No other elements contaminated the sample surface. In order to remove any carbon-, nitrogen-, sodium- and oxygen-containing surface-adsorbed contaminants, the surface was annealed. The annealing temperature had to be chosen carefully; the surface oxide would have decomposed if the temperature was too high. Oxide decomposition on the (110) surface has maximum efficiency in the 760-890 °C temperature range, as seen in temperature-programmed desorption experiments [371]. The annealing temperature was therefore chosen to be lower than 650 °C to exclude surface oxide decomposition.
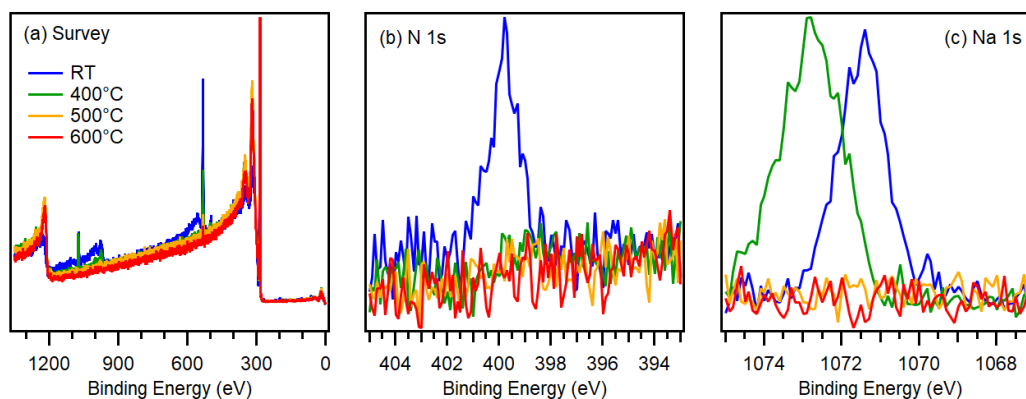
Figure 3.9. (a) Survey, (b) nitrogen $1s$ (N $1s$), and (c) sodium $1s$ (Na $1s$) XPS spectra of the diamond sample at room temperature (RT) i.e. prior to annealing, and after annealing to 400 °C, 500 °C, and 600 °C. Spectra were taken at an emission angle of 0°. The survey spectra are truncated and do not show the full height of the C $1s$ signal.

While absolute coverage determination from XPS is difficult without a trustworthy reference, some approximate results for the oxygen coverage prior to annealing were obtained. When comparing the carbon signals from the unannealed and annealed surfaces, the signal attenuation due to desorbed contaminants can be determined. While this method is tainted by several inaccuracies, it can be used to distinguish a thick layer of contaminants on the surface from a single oxide layer. As can be seen in Figure 3.10(a), the C $1s$ signal of the unannealed surface is significantly attenuated when compared to the annealed surface; it is estimated that a layer with a thickness of 5–40 nm is desorbed due to the annealing. Correspondingly, the O $1s$ signal is massively reduced, as can be seen in Figure 3.10(b). The small N $1s$ and Na $1s$ signals in Figures 3.9(b) and (c) that were detected prior

to annealing vanished post-annealing to 400 °C and 500 °C respectively. A similar desorption/change in attenuation was observed for different annealing temperatures of 500 °C and 650 °C, well below the threshold for surface oxide decomposition.

From these results, it must be concluded that the majority of the carbon and oxygen signals that contribute to the unannealed sample's spectra due to volatile contaminants, which can be desorbed. Therefore, only the spectra relate to the actual oxygenated (110) surface, and the spectra are not usable due to the significant contribution of contaminants to both the C 1s and O 1s XPS signals. It should be noted that some previously published XPS experiments have experienced the same problem, with C 1s and O 1s spectra from literature resembling data from the Warwick experiments for the unannealed surface in Figure 3.10 [369]. The O 1s spectrum of the unannealed surface shows a broad peak which can be fitted by two subpeaks, as shown using dotted lines in Figure 3.10(b), while the C 1s region shows a main peak with one pronounced satellite and a broad shoulder showing a second satellite signal, as can be seen in Figure 3.10(c). When fitted, the binding energy differences from the main peak to the first and second satellite peak are 1.5 eV and 4.1 eV respectively.
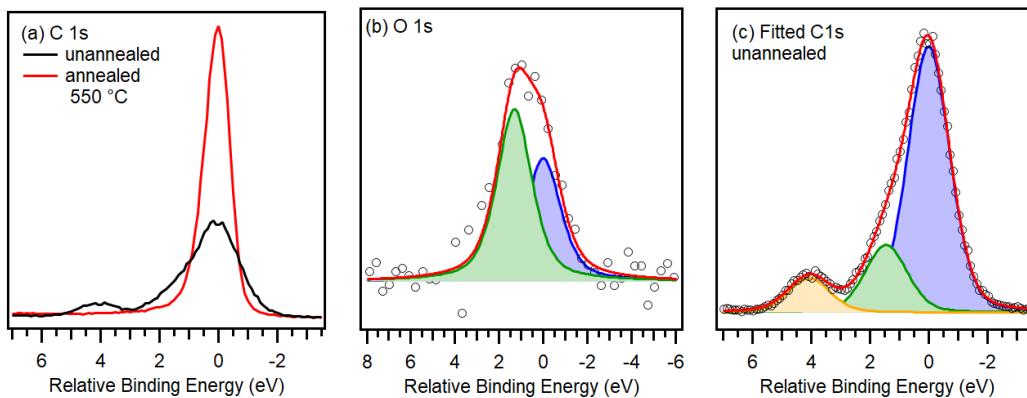
Figure 3.10. (a) Carbon 1s (C 1s) and (b) oxygen 1s (O 1s) XPS spectra of the diamond sample prior to annealing and after annealing to 550 °C, where fitted peak components for the O 1s spectrum prior to annealing are shown as dotted lines. (c) Fitted C 1s spectrum prior to annealing. Open circles: data, red line: fit function, shaded areas: fitted peaks. All spectra taken at an emission angle of 60°. C 1s and O 1s binding energies are reported with respect to the bulk C 1s and the dominant O 1s signals respectively.

3.3.3 Sample Surface Characterisation

The simulated binding energies of the lower coverage $p(2 \times 2)$Keto–Ether$_5$ phase are close to the Warwick XPS data, while the higher coverage $p(2 \times 1)$Keto–Ether$_5$ phase shows a larger deviation from experiment. Good agreement is also observed between simulated O $1s$ binding energy differences for the low coverage structure and the Warwick experiments, as shown in Table 3.6. On the other hand, the simulations of the high coverage structure show deviations that cannot be reconciled with XPS experiments once the obvious surface contaminants are removed.

X-ray diffraction experiments were performed by Dr. David Walker to gather further information about the quality of the sample surface. X-ray diffraction is bulk-sensitive and can therefore provide the orientation of the whole single crystal relative to the sample holder. Figure 3.11 shows a pole figure for the sample (prior to the removal of twinned edge regions) in the [220] direction, proving that the sample is predominantly (110)-oriented along the sample normal with a clear peak at a tilt angle of 1.5°, which corresponds to an azimuthal angle of 64.4°.

Low-energy electron diffraction experiments were also conducted by Dr. Marc Walker as the technique is very surface-sensitive and can therefore provide information on the top few surface layers of the sample. Figure 3.12 shows the diffraction pattern for the sample surface. As can be seen, there is only one set of diffraction spots, which means that the surface is not polycrystalline on the length scale of the spot size $(1 \times 1 \, \text{mm}^2)$, but rather only has one domain (or one domain type combined with amorphous material). The diffraction pattern also has a two-fold symmetry, which is in line with a predominant (110)-oriented face-centred cubic diamond structure [372; 394; 395; 396].
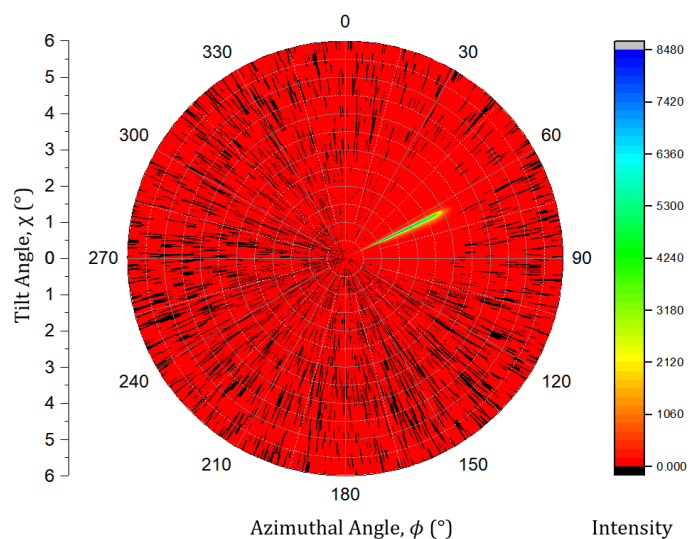
Figure 3.11. Pole figure of the diamond sample prior to the removal of twinned edge regions, recorded in the [220] direction at a 75.3° 2θ reflection position, where θ is the Bragg angle.
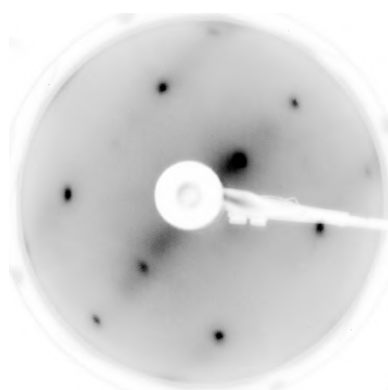
Figure 3.12. Low-energy electron diffraction pattern of the diamond sample surface, taken at an electron energy of 122 eV.

To obtain information about the surface roughness and morphology of the post-growth and polished crystal surface, atomic force microscopy measurements were also performed by Dr. Julie Macpherson on different locations of the surface. Figure 3.13 shows a typical $1.0 \times 1.0\,\mu\text{m}^2$ image of the sample surface topography. The average surface roughness and a root-mean-square roughness were found to be only 0.18 nm and 0.23 nm respectively, which is close to atomic smoothness. The small bright spots are associated with impurities ('dirt') on the surface. Repeated sonication of the crystal during the cleaning stage was found to decrease the small particle density and size. The (110) surface used in the Warwick experiments is

67

therefore of higher quality than all other diamond (110) surfaces currently reported in literature [152; 369; 370; 371; 372; 394; 395; 396; 397].
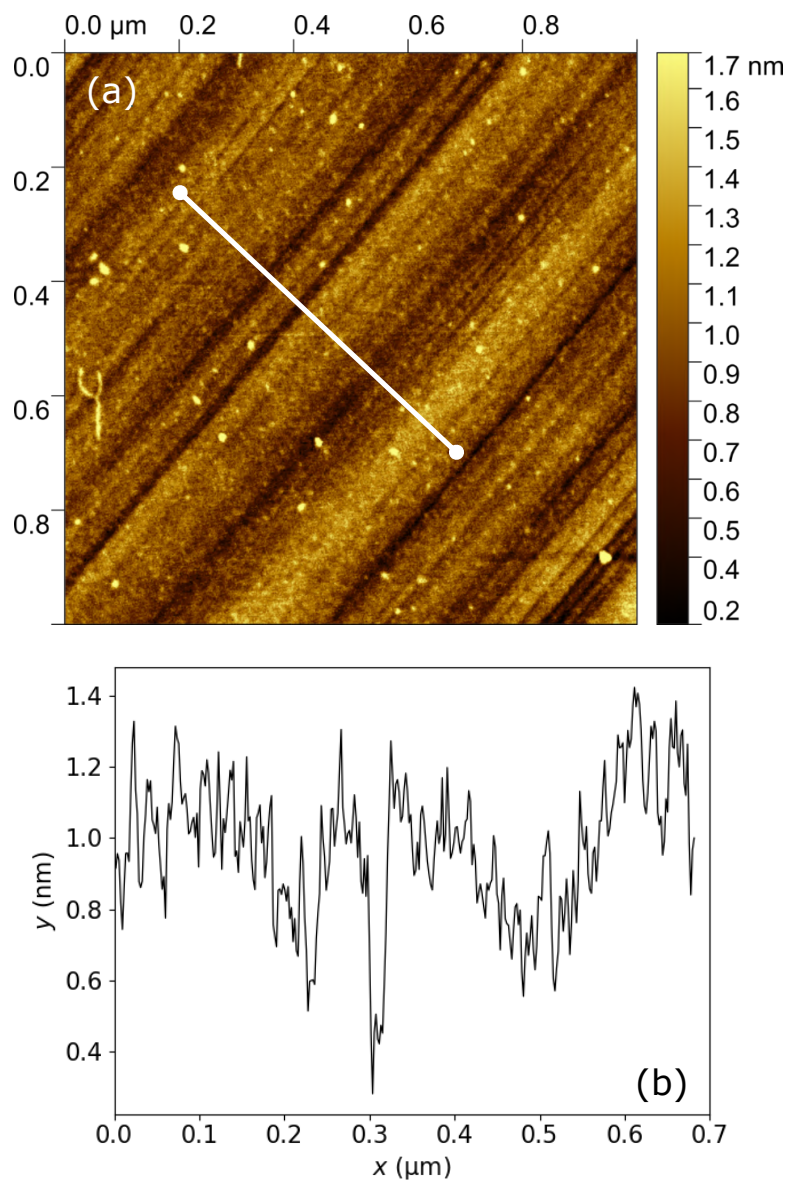
Figure 3.13. Atomic force microscopy of the sample surface topography. (a) $1.0 \times 1.0\,\mu m^2$ image and (b) the roughness profile of the white line in (a). The average surface roughness and the root mean square roughness were found to be $0.18\,nm$ and $0.23\,nm$ respectively.

The observed larger deviations for the $p(2 \times 1)$ system are therefore hypothesised to be due to the higher coverage structure being difficult to realise for realistic surfaces under the experimental conditions employed herein.

3.4   Conclusions

Using *ab initio* atomistic thermodynamics based on DFT, this chapter has established a phase diagram of the most stable oxygen terminations of the diamond (110) surface. Within the 0–1000 K temperature range, the phase diagram is dominated by a highly stable phase of coexistent carbonyl and ether functional groups, while peroxide groups become more stable at low temperatures and high oxygen pressures, which is in agreement with existing experimental literature. Adjacent carbonyl and ether groups are observed which, as suggested from DFT calculations, can form via a transition from an adsorbed peroxide structure. The high stability of the adjacent carbonyl and ether groups arises from cooperative effects that mitigate surface deformation and satisfy all valencies. These findings are robust with respect to the DFA and dispersion correction scheme. From the simulation of core-level binding energies, a lower coverage of this phase was found to produce core-level shifts in agreement with experimental XPS data. While the experimental data does not agree with the simulation results for the most stable highest coverage, this could indicate that this coverage may not be able to be realised due to kinetic hindrance and coverage limitations on realistic surfaces. The combined theoretical and experimental analyses outlined in this chapter provide a much improved understanding of the oxygen-terminated diamond (110) surface, which has been lacking to-date. A better characterisation of the chemical terminations of diamond-based materials is crucial to understanding elementary diamond growth and electrochemical surface processes.

Chapter 4

Growth of Au$_n$ Nanoclusters on Diamond, $1 \leq n \leq 4$

4.1 Introduction

Chapter 3 characterised the oxygenation state of the diamond (110) surface and showed that the most stable surface phase is comprised of coexistent and adjacent carbonyl and ether groups. In this chapter, this surface will be used to computationally explore the initial stages of gold NC formation on BDD surfaces, as is the aim of this thesis. The formation of all NCs and NPs start from single atoms, which are the smallest possible entity that can be deposited [398; 399; 400; 401]. Single metal atoms have been shown to have unique magnetic properties [402] and excellent (electro)catalytic applications [203; 204; 205; 206], often outperforming NCs in the latter regard [183; 184; 185]. This makes it essential to investigate a variety of possible stabilisation mechanisms that can not only promote the deposition of single gold atoms onto BDD surfaces provide some key insights into the initial stages of atomistic nucleation.

Using STEM, single gold atoms were shown to be stable atop polycrystalline BDD surfaces, while DFT calculations showed that these single atoms possess low diffusion barriers on an idealised oxygen-terminated surface [22]. However, the DFT calculations reported in Hussein *et al.* [22] did not account for local surface defects and dopants. This is because modelling isolated defects and dopants at extended surfaces is challenging, as the majority of computational studies typically employ periodic boundary conditions [68]. These define an infinitely repeating unit cell and have been used to model extended surfaces and interfaces [68] in numerous studies,

including within the previous chapter. However, periodic models become problematic when modelling local defects or dopants as they acquire a periodicity under such boundary conditions, when they should in fact be aperiodic. For example, when it comes to modelling BDD, boron dopants typically have a concentration of one boron atom within a thousand carbon atoms [131], and computational studies either do not include the dopant [22; 386], whilst assuming that this low concentration means it is safe to exclude boron, or they model the dopant within periodic boundary conditions anyway [403; 404] and are not concerned with the periodicity of the dopant. Embedded cluster approaches are an alternative that can be used to appropriately model surface defects [68]. However, there are certain challenges that arise when attempting to model structures using embedded cluster approaches. One such challenge is the creation of additional surfaces at the edges of the cluster when it is cut from a substrate. For semiconducting materials such as BDD, this involves the cleaving of several covalent bonds, and these dangling bonds can give rise to states within the band gap of the material and thus artificially increase its reactivity [68]. This issue can be resolved by either saturating the dangling bonds using hydrogen atoms for example, or by somehow embedding the cluster within a surrounding material that is significantly more extended [68]. However, both of these solutions require the cluster itself to be large enough to ensure finite-size effects are avoided, which can occur due to the spatial confinement of the wavefunction within the cluster and increase the band gap of the material [68; 405; 406], whilst not being so large as to be computationally infeasible [68]. It is therefore essential to develop a computational framework that can be used to study the adsorption energetics and kinetic barriers of gold adatoms and NCs on extended diamond surfaces whilst also accounting for local surface defects and dopants.

Ab initio methods such as DFT [65; 66] can provide detailed insights into the structural and electronic properties of supported metal atoms [68; 69; 407]. However, as a standalone method, its computational scalability [408] generally means that higher-rung DFAs, which are theoretically more accurate [409] and generally more computationally expensive [68; 69], cannot be used to study large, periodic models without incurring high, and often infeasible, computational costs [410]. Due to this, many large-scale studies are often limited to using GGAs or MGGAs as the DFA to the true XC functional in the calculation of the Kohn-Sham ground-state energy [68; 69], as explained in Section 2.1.4. These DFAs typically either correctly estimate the adsorption energy or the reaction barriers, but never both [68; 69]. They also suffer from the lack of long-range dispersion interactions within their

formulations, though this has been successfully addressed recently with the development of *a posteriori* dispersion correction schemes, as discussed in Section 2.2.

To circumvent the limitations of DFT with periodic boundary conditions and to address the fact that surface chemistry is intrinsically local, as well as potentially aperiodic in nature in the presence of surface defects on semiconductors, embedded cluster approaches based on a hybrid quantum mechanics/molecular mechanics (QM/MM) [300; 301] methodology, as explained in Section 4.2.1, are a viable alternative. Cluster approaches are generally computationally cheaper than periodic approaches and can therefore permit calculations with higher-rung DFAs such as HGGAs, which can allow for a systematic assessment of the accuracy of lower-rung DFAs along Jacob's ladder [219] against higher-rung DFAs without changing the model setup [406]. This is particularly important in the case of metal adsorption on insulating and semiconducting materials as there are very little experimental data that exists about adsorption structures and energetics that computational approaches can be compared against.

This chapter will explain the construction of an embedded cluster model that can be used to study the adsorption of single gold atoms on oxygen-terminated diamond (110) surface. This model will be used to analyse the effect of surface defects and substitutional boron dopants within the diamond substrate on gold atom adsorption. By using DFT as the QM method within this hybrid QM/MM framework, the embedded cluster model will be used to benchmark the performance of various QM density-dependent methods, including DFAs and *a posteriori* dispersion correction schemes. Finally, after the identification of suitable QM methods and stabilisation mechanisms for single gold atom adsorption on diamond, the growth of small gold NCs will be briefly analysed, with a focus on gold dimers, trimers and tetramers.

4.2 Methods

Henceforth, the notation '$\chi^{+\psi}/\phi$' is used to denote specific hybrid QM/MM methods, where χ is the DFA, ψ is the dispersion correction and ϕ is the forcefield used.

4.2.1 QM/MM Model Setup

The Py-ChemShell [301; 340] software package was used to cut hemispherical clusters of radius $20.0\,a_0$ (and active radius $10.0\,a_0$) from the PBE$^{+\text{TS}}$-optimised $p(2\times$

1)Keto-Ether$_5$ periodic model and run hybrid QM/MM calculations, as shown in Listing 2.1. The surface oxygenation state was initially modelled to comprise coexistent and adjacent carbonyl and ether groups at full oxygen coverage, as was shown to be the most thermodynamically stable state in Chapter 3. Figure 4.1 details the cutting and partitioning processes of a cluster into QM and MM regions from the periodic surface model, which was done using a Python script that can be found in Listing 2.1. The FHI-aims [70] and GULP [337; 338] software packages were used to treat the QM and MM regions respectively; QM- and MM-specific calculation settings are detailed in Section 4.2.3. QM/MM energies were calculated using an additive scheme [304], as described in Equation (2.46) and implemented within the Py-ChemShell [301] software, and the hydrogen link-atom approach [309] was used to treat cleaved covalent interactions across the QM-MM interface, with the `connect_toler` keyword (a rescaling coefficient for vdW radii to determine bonding interactions) set to a value of 1.3 for all QM/MM calculations to ensure correct hydrogen saturation of the QM region for the FHI-aims calculator, as shown in Listing 2.3.
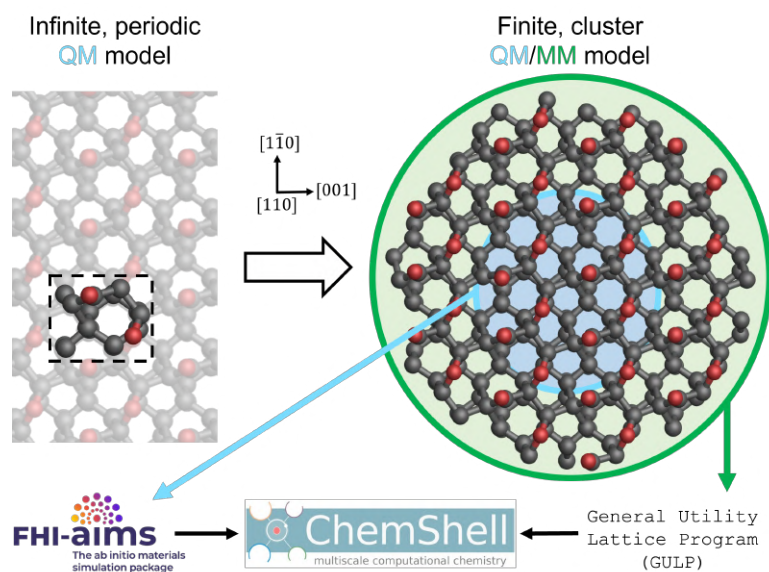
Figure 4.1. Process of cutting an infinite, periodic surface model into a finite, embedded cluster surface model partitioned into QM and MM regions. Atoms within the blue circle represent the QM region of the cluster, while the green annulus represents atoms within the MM region. Also shown are the software packages used to treat the different regions. The surface is visualised from the [110] direction, with surface axes presented, and the unit cell outlines are shown with black dashed lines. Carbon and oxygen atoms are shown in grey and red respectively.

4.2.2 Construction of Structures

The fully oxygen-terminated surface model shown in Figure 4.2(a) is an idealised one and unlikely to be realistic due to kinetic hindrance and coverage limitations. Oxygen vacancies in oxides have been shown to influence the catalytic properties of small gold clusters [407; 411; 412; 413]. For this reason, a point defect at the surface was modelled by removing a carbonyl oxygen, as shown in Figure 4.2(b). The embedded cluster model eliminates any finite size effects that would have been caused by interactions of the defect site with its periodic image. To ensure this defect was modelled correctly, PBE^{+TS}/REBO optimisation was conducted after the removal of the carbonyl oxygen atom. Because diamond surfaces are usually hydrogen-terminated after CVD growth [145], uncoordinated carbon atoms were subsequently saturated with hydrogen species and the surface was reoptimised using PBE^{+TS}/REBO.

Boron-doped surfaces, as are commonly used in electrochemical applications [21; 22], were also investigated. Single substitutional boron dopants introduce a charge into the diamond, regardless of whether it is uncompensated (neutral) or compensated (ionised). The uncompensated instance was modelled as only neutral boron gives rise to the characteristic one-phonon absorption and luminescence features observed in BDDs [414]. Boron dopants can be situated within the top few surface layers or deep within the bulk material. For the former case, where the effects of the boron are localised, a single boron atom was explicitly placed within the second and third carbon layers of the QM region by replacing a carbon atom, and an initial 1+ charge was placed on the neighbouring uncoordinated carbon atom, as shown in Figures 4.2(c) and (d) respectively. The initial conditions of the localised dopant calculations therefore represent uncompensated boron. An assumption was made that the explicit presence of the boron atom in the top surface layers will not affect the surface oxygenation state of the substrate.

To model boron dopants located deep within the bulk material, where the effects of the dopant are delocalised, a +1 charge was placed on the entire QM region to account for the one fewer electron, as shown in Figure 4.2(e). It should be noted that the delocalised dopant model is not boron-specific, as boron is not explicitly included, and is applicable for any delocalised single substitutional triel such as aluminium, gallium or indium. While non-boron triels are not common diamond dopants, some have still found use. For example, aluminium dopants have been shown to induce superconductivity [415; 416], though boron was deemed to

74

be a more suitable dopant to attain superconductivity [416]. Gallium dopants have been shown to suppress the graphitisation of diamond tools by increasing their wear resistance [417; 418], while indium dopants have been shown to improve the wettability of diamond [419]. In both cases, only (the effect of) one dopant atom was included within the QM region to match common boron dopant densities [131]. The finite, cluster-based QM/MM approach allows for the single dopant/charge defect to be accurately modelled in a much more pertinent manner as opposed to a periodically repeating dopant and/or charge.
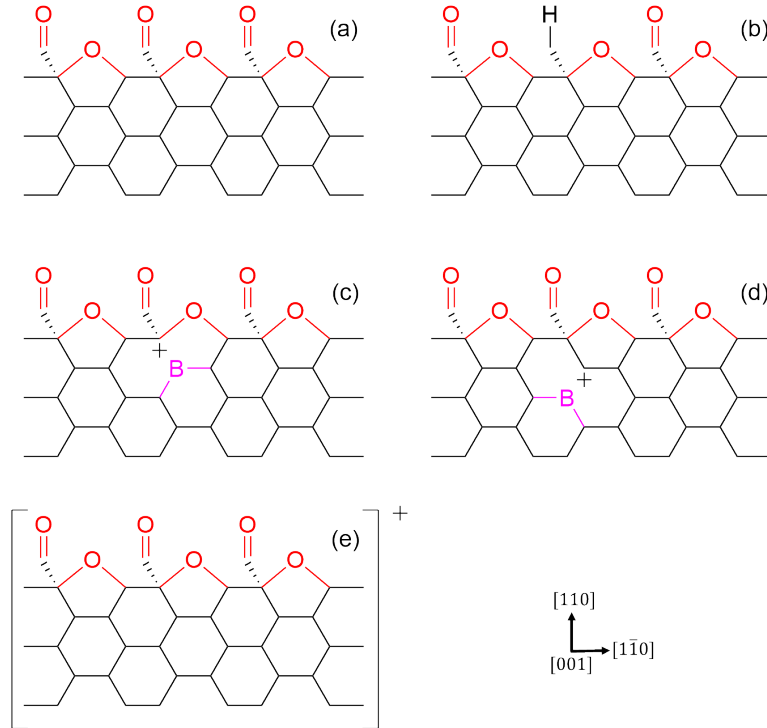
Figure 4.2. Skeletal visualisations of the five substrate models investigated. Substrates are (a) an idealised oxygen-terminated diamond (110) surface (b) a SCOV-defective surface (c) a boron-doped surface with the dopant in the second layer (d) a boron-doped surface with the dopant in the third layer and (e) a delocalised triel-doped surface. Visualisations are shown from the [001] direction.

To construct Au_n NCs, $n \in \{2, 3, 4\}$, the most stable Au_{n-1} structures of a particular shape on the surface were identified, and the additional n^{th} gold atom was subsequently manually placed at several positions to generate Au_n NCs. While this procedure will not explore the entire space of possible NCs of a given atom count on a specified surface, it can provide insights into preferred shapes and their energetics. Ideally, a global optimisation would have been conducted via basin-hopping [98; 99] for example to identify the global minimum for every Au_n NC

75

on a given surface. However, as of the 21.0 release of the Py-ChemShell [301] software, global optimisation algorithms could not be conducted, which necessitated the manual preparation of NC structures.

4.2.3 Computational Settings

QM DFT [65; 66] calculations were performed using the FHI-aims [70; 259] software package, and NAOs were represented using a 'tight' basis set. Unless otherwise specified, the pairwise, long-range TS [262] dispersion correction method was used to account for vdW interactions in calculations with GGAs and HGGAs. The TS [262] method was not used alongside MGGAs, which already account for a certain level of mid-range interactions [234], or LDAs, which exhibit an artificial energy minimum between subsystems that is taken to unintentionally mimic vdW interactions [68], as explained in Section 2.1.4. For periodic calculations, the interaction between the gold adatom and its periodic images were excluded for the TS dispersion correction. The Atomic Simulation Environment [339] was used to generate structures. Density of states graphs were plotted using the `logsdail/carmm` [346] GitHub repository, with a Gaussian broadening value of $0.02\,\text{eV}$ used for smoothing. For the single gold atom, spin polarisation was accounted for by setting an initial spin moment of 1 for the single unpaired electron; the implementation of this feature within the Py-ChemShell [301] software is explained in Section 2.7.5, with the amended source file provided in Listing A.6.

The PBE [229] GGA was the primary DFA used herein, though a variety of other DFAs were also used to run DFT calculations; DFAs are clearly stated throughout for clarity. DFAs whose diamond lattice constant values were within $\pm 0.02\,\text{Å}$ of the PBE [229] value were chosen for this study. This was done to ensure interatomic distances within the diamond substrate were not too dissimilar from a specific DFA-optimised value and allow for fair comparisons to be made between DFAs. The lattice constant for each DFA was picked by either optimising the lattice vectors of the primitive diamond unit cell with a two-atom *motif*, as outlined in Section 1.2.2, or by taking values from the Materials Science and Engineering dataset [420]. DFAs, either implemented within FHI-aims or from the Libxc [218] library, from different rungs of Jacob's ladder [219] were investigated and compared against the PBE [229] GGA. LDAs investigated were: GDSMFB [421], KSDT [422] and PZ-LDA [423; 424]. The following GGAs were studied: PBEsol [232], revPBE [231] and RPBE [233]; and the following MGGAs were examined: SCAN [237], rSCAN [425], M06-L [235], TPSS [234], TPSSloc [236] and revTPSS [426]. The following HGGAs were consid-

ered: HSE03 [241], HSE06 [242], PBE0 [243] and PBEsol0 [244]. The `dfauto` [427] implementation within the FHI-aims [70] software package was used to run calculations with the SCAN [237] and rSCAN [425] MGGAs, and a screening parameter of $0.11\,a_0{}^{-1}$ was set for the HSE06 [242] HGGA.

MM calculations were performed with the GULP [337; 338] software package. The reactive empirical bond order (REBO) potential specified in Brenner *et al.* [428; 429] was used to run the majority of MM calculations, as it has been shown to accurately describe interactions between carbon, oxygen and hydrogen atoms [429], as well as accurately predict carbon-carbon bond lengths and angles within diamond [428]. Some calculations were performed using the Tersoff [430] forcefield; forcefields are clearly stated throughout for clarity.

4.2.4   Energy Calculations

Equation (4.1) was used to calculate the adsorption energy per atom, E_{ads}, of different gold NCs comprising n atoms (Au_n):

$$E_{\text{ads}} = \frac{E_{\text{total}} - E_{\text{substrate}} - E_{\text{Au}_n}}{n} \tag{4.1}$$

where n is the number of gold atoms, E_{total} is the total energy of the gold-diamond complex, E_{surface} is the energy of the clean surface onto which the gold cluster was adsorbed, and E_{Au_n} is the gas-phase energy of the Au_n NC in the frozen surface-adsorbed geometry. Equation (4.2) was used to calculate the cohesion energy, E_{coh}, for any Au_n NC:

$$E_{\text{coh}} = \frac{E_{\text{Au}_n} - nE_{\text{Au}_1}}{n} \tag{4.2}$$

For structure optimisations with any QM/MM method, the PBE$^{+\text{TS}}$/REBO-optimised oxygen-terminated diamond substrate was reoptimised using the new QM/MM method. A single gold atom was then placed $1.5\,\text{Å}$ above the adsorption site, and another optimisation was conducted using the specified QM/MM method. For the construction of binding energy curves for a specified QM/MM method, single-point calculations were run on the specified QM/MM-optimised surface substrate, with the gold atom being placed at various heights above the surface.

To calculate the kinetic stabilities of a gold adatom with a specified QM/MM method, the gold atom was first translated to a new site along either the [001] or the [1$\bar{1}$0] directions, and placed $1.5\,\text{Å}$ above the specified QM/MM-optimised surface. A

constrained optimisation was then conducted, where the position of the gold atom was only allowed to relax along the [110] direction, with motion along the [001] and [1$\bar{1}$0] directions frozen. The kinetic stability of the gold atom with any specified QM/MM method was then calculated as the energy difference between the initial optimised structure and the constrained-optimised structure. An explanation of how constrained optimisations were implemented within the Py-ChemShell [301] software is explained in Section 2.7.5, with the amended source files provided in Listings A.7 and A.8.

4.3 Results and Discussion

4.3.1 QM Region Size Optimisation

As explained in Section 4.1, it is important to ensure that the size of the QM region embedded within the MM region is large enough to avoid any finite-size effects. The appropriate QM region size of the substrate was chosen by comparing various properties of PBE^{+TS}/REBO-optimised embedded cluster models with varying QM region size against the parent PBE^{+TS}-optimised periodic model. All convergence tests were conducted on an idealised surface model. First, the structural deviations of each PBE^{+TS}/REBO-optimised embedded cluster were compared against the initially cut cluster from the PBE^{+TS}-optimised periodic model, which can therefore be taken to be an appropriate representation of the periodic model. As can be seen from Figure 4.3(a), PBE^{+TS}/REBO-optimised embedded clusters with QM region sizes of 10, 20, 60, 70 and 90 atoms have the lowest root-mean-square deviation (RMSD), 0.037 Å, with respect to the PBE^{+TS}-optimised periodic model, while the RMSDs for the 30-, 40-, 50- and 80-atom QM regions were at least 0.7 eV, showing a greater disparity in optimised structures.

Following this structural comparison, the energetics of the PBE^{+TS}/REBO-optimised embedded cluster models were compared against the PBE^{+TS} periodic model. Figure 4.3(b) shows a graph of the band gap, which is the energy between the highest occupied molecular orbital and the lowest unoccupied molecular orbital, of the embedded clusters as a function of QM region size. As can be seen, the band gap generally decreases as the QM region size increases and tends towards the QM periodic value (2.3 eV). This shows that clusters with smaller QM region sizes do experience some finite-size effects. In contrast, as the QM region size increases, the embedded cluster becomes more structurally and energetically similar to the periodic QM structure. Despite this overall trend, over the range of QM region sizes

explored, the 30-atom QM region was found to possess the closest band gap to the periodic value, while the 90-atom QM region had the next closest value (2.6 eV). The band gaps of the 10-, 20-, 60- and 70-atom QM regions, which were found to be structurally similar to the periodic model, were found to be at least 0.5 eV higher than the QM periodic value.

Finally, the adsorption energy of a single gold atom was evaluated as a function of the QM region size using Equation (4.1), as shown in Figure 4.3(c). The 70- and 90-atom QM regions (71- and 91-atoms respectively including the gold atom) resulted in the closest adsorption energy value, -0.30 eV to the QM periodic value (-0.31 eV). The 20- and 30-atom QM regions significantly overestimate the adsorption energy, while the 10-, 40-, 50-, 60- and 80-atom QM regions report similar adsorption energetics to the QM periodic model, but are not as close as the 70- and 90-atom QM regions.

Taking all three properties into consideration, it can be seen that the 90-atom QM region results in an optimised structure, band gap and gold adsorption energy most similar to the QM periodic model, and was thus chosen as the optimal QM region size within the QM/MM cluster. While a larger QM region would most likely result in a cluster with a final geometry and energetics more similar to the periodic model, convergence problems were encountered with larger QM region sizes (100 and 110 atoms). Regardless, the embedded cluster with a 90-atom QM region was found to have similar structural and energetic properties to the periodic QM cluster and was thus deemed an appropriate size.

To further confirm the higher computational efficiency of the hybrid QM/MM approach, scaling graphs were constructed after conducting single-point calculations on the PBE^{+TS}-optimised periodic model and the PBE^{+TS}/REBO-optimised model, with a 90-atom QM region, after a gold atom was adsorbed onto the model surfaces. Figure 4.4 shows the computational cost of these single-point calculations as a function of number of cores, which were run on Lenovo NeXtScale nx360 M5 servers with dual Intel Xeon E5-2680 v4 (Broadwell) 14-core processors at 2.4 GHz, as available within the Orac high performance computing cluster provided by the Scientific Computing Research Technology Platform of the University of Warwick. All calculations used the Eigenvalue SoLvers for Petaflop-Applications [431] library and the ELectronic Structure Infrastructure [432]. As can be seen in Figure 4.4, QM/MM calculations are vastly cheaper than the periodic QM calculations. Furthermore,

periodic QM calculations failed when using 16 cores or fewer due to memory issues. This confirms the superior computational efficiency of the hybrid QM/MM approach and that it can be used to access more computationally-costly methods such as MGGAs and HGGAs.
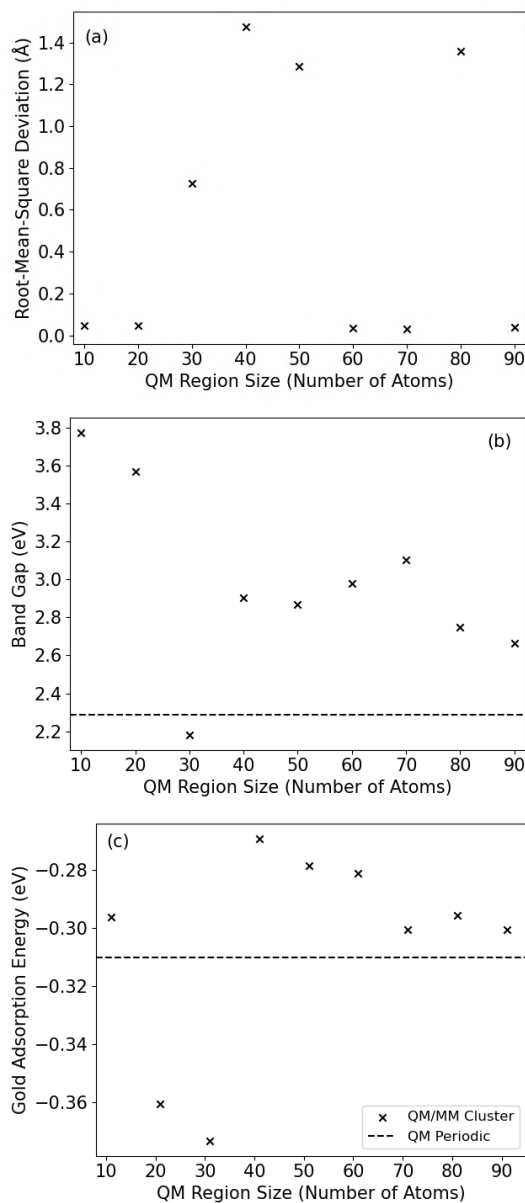
Figure 4.3. Scatter graphs showing the (a) RMSDs and (b) band gaps of a single gold atom atop PBE^{+TS}/REBO-optimised cluster models against the initial cluster cut from the PBE^{+TS}-optimised periodic model; and (c) adsorption energies of a single gold atom atop PBE^{+TS}/REBO-optimised cluster models against the PBE^{+TS}-optimised periodic model, all as a function of QM region size.
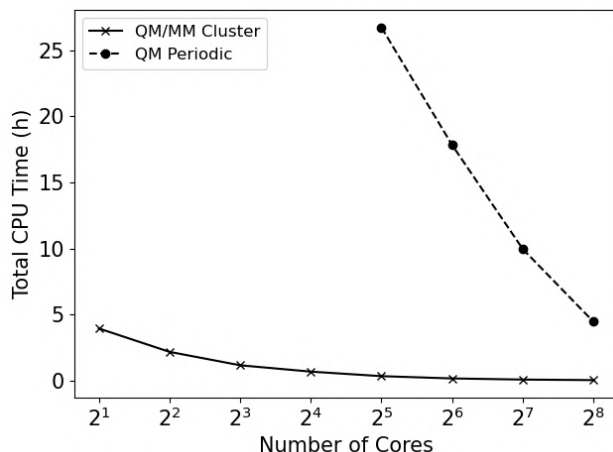
Figure 4.4. Scaling graphs of single-point PBE^{+TS}/REBO and PBE^{+TS} calculations of the idealised oxygen-terminated diamond (110) surface. The embedded cluster model comprised 527 atoms with a 90-atom QM region, while the periodic model comprised 92-atom unit cell.

4.3.2   Effect of Defects and Dopants

The idealised, fully oxygen-terminated diamond (110) surface was used as the starting point for all QM/MM models. Other systems based on this were also studied, where defects and dopants were introduced into the surface model, namely a defective surface with a saturated carbonyl oxygen vacancy (SCOV), and boron-doped surfaces with the dopant modelled explicitly and implicitly, as explained in Section 4.2.1 and visualised in Figure 4.2. The final adsorption structures are all visualised in Figure 4.5, while Table 4.1 summarises the adsorption energetics and the computed Mulliken charges [384] of the single gold atom. The interactions between the gold atom and each surface are discussed in more detail below.

As can be seen in Table 4.1, the introduction of defects or dopants into the idealised surface seems to increase the adsorption energy of the gold atom, which is reflected in the lower adsorption height, indicating the closer proximity of the adatom to the surface. For all investigated defective and doped surfaces, the sign of the Mulliken charge [384] on the gold atom was found to be positive, which is indicative of charge transfer from the gold atom into the surface and explains the relatively adsorption energies. In contrast, for the idealised , the sign of the Mulliken charge is positive, indicating charge accumulation. It should be noted that the more complete a basis set is, the more ambiguous a Mulliken analysis becomes as it is not *a priori* clear which electrons should be counted towards the basis functions of one

atom rather than another. While other charge analysis schemes do exist [284; 433], these are also not perfect, but the Mulliken analyses used herein can still be used as a qualitative indicator of trends.
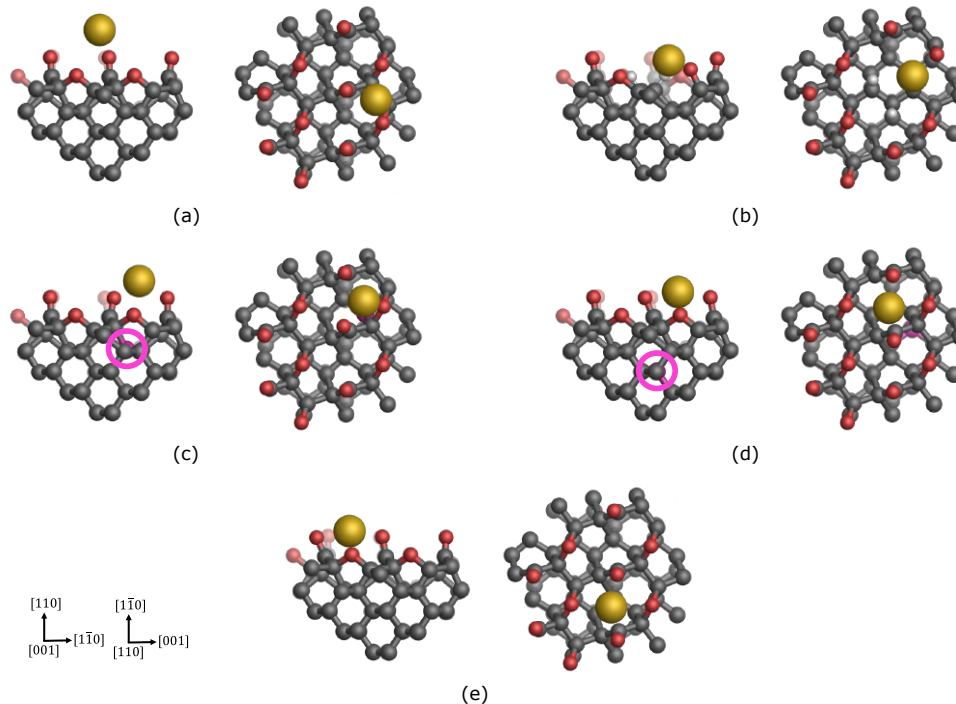
Figure 4.5. Orthographic ball-and-stick visualisations of a gold adatom on different oxygen-terminated diamond (110) surface substrates, as optimised using the PBE^{+TS}/REBO method. Substrates are (a) an idealised oxygen-terminated diamond (110) surface (b) a SCOV-defective surface (c) a boron-doped surface with the dopant in the second layer (d) a boron-doped surface with the dopant in the third layer and (e) a delocalised triel-doped surface. Visualisations of the QM region are shown from the [001] and [110] directions, and surface axes are also shown, with the saturating hydrogen link-atoms at the QM region boundary excluded for clarity. Carbon, oxygen, hydrogen, boron and gold atoms are shown in grey, red, white, pink and gold, respectively. For clarity, pink circles are also included to show which carbon atom the boron atom is situated behind for (c) and (d).

Idealised surface: In the case of the idealised, fully oxygen-terminated surface, the gold adatom was found to weakly adsorb onto a carbonyl oxygen atom, at a height of 1.71 Å above the with an adsorption energy of −0.30 eV, as detailed in Table 4.1. This is a very weak adsorption energy and its associated diffusion barriers are also likely very weak, as has been shown previously [22]. Such weak adsorption of the gold adatom is to be expected for this idealised system due to the high stability

System	Adsorption Energy (eV)	Adsorption Height (Å)	Mulliken Charge ($|e|$)
Idealised	−0.30	1.71	−0.14
SCOV-defective	−2.31	−0.12	+0.07
Boron dopant (2nd layer)	−1.66	1.03	+0.28
Boron dopant (3rd layer)	−1.75	0.35	+0.16
Delocalised triel dopant	−1.98	0.36	+0.26

Table 4.1. Adsorption energies, adsorption heights and Mulliken charges for a single gold adatom on various oxygen-terminated diamond (110) surface substrates. Adsorption energies were calculated using the PBE^{+TS}/REBO method. Adsorption heights are given with respect to the plane of carbonyl oxygen atoms.

of the coexistent carbonyl and ether functional groups on the diamond surface, as shown in Chapter 3. The valencies of all surface atoms are satisfied and the surface is therefore 'closed-shelled'. There are no unpaired electrons for the gold atom to interact with, which means the interaction between the adatom and the surface is governed by weak long-range effects such as vdW forces and electrostatics.

SCOV-defective surface: Compared to the idealised system, the gold adatom was found to adsorb significantly stronger to the SCOV-defective diamond surface with an adsorption energy of −2.31 eV, indicating that this is a much more adsorption complex. This is reflected by the negative adsorption height in Table 4.1, which indicates that the gold atom sits below the plane of carbonyl oxygen atoms, and is thus much closer to the surface carbon atoms than in the idealised surface. The addition of a gold atom first causes one of the C–O bonds within a surface ether group to dissociate, and the gold subsequently atom sits in between the former ether group.

To elucidate the nature of the bond between the gold adatom and the diamond surface, the projected density of states of the gold atom and its neighbouring former-ether oxygen atom was computed based on a Mulliken analysis [384] and are shown in Figure 4.6(a). The highest occupied orbital, as shown by the peak centred at an eigenenergy of −4.1 eV, includes contributions from oxygen p-states, as well as gold s-, p- and d-states. In contrast, the lowest unoccupied orbital, as shown by the peak centred at −2.4 eV, is dominated by gold s-states but also includes

contributions from both oxygen and gold p-states, with a small contribution from gold d-states. Furthermore, in both peaks, the contributions from oxygen s- and gold f-states are near-zero and negligible. The presence of the single gold atom can therefore be seen to form both bonding and antibonding orbitals, and is indicative of a bonding interaction between spd-hybridised orbitals of the gold atom and the oxygen p orbitals that has been observed between gold and oxygen atoms in literature [434].
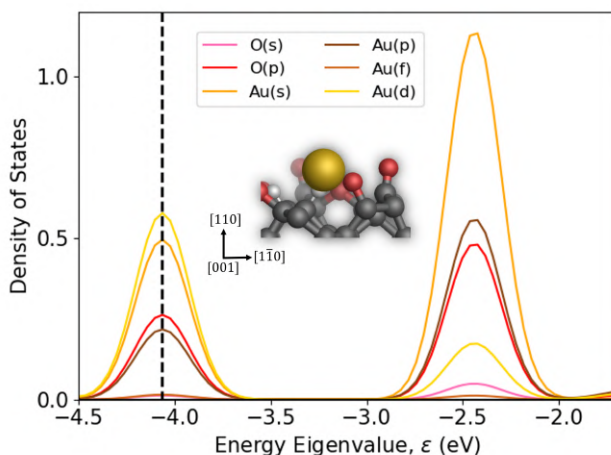
Figure 4.6. Projected density of states of the orbital contributions from a single gold (Au) atom and its neighbouring formerly-ether oxygen (O) atom on a SCOV-defective oxygen-terminated diamond (110) surface after optimisation with the PBE$^{+\text{TS}}$/REBO method. The black dashed vertical line indicates the position of the highest occupied orbital. Also shown is an orthographic ball-and-stick visualisation of a single gold adsorbed onto the SCOV-defective surface along the [001] direction. Carbon, oxygen, hydrogen, and gold atoms are shown in grey, red, white, and gold respectively.

The Au–O bond length was found to be 2.09 Å and the angle between the gold, oxygen and the other ether carbon atoms is 118.0°. The observed bond length is only 0.07 Å longer than the sum (2.02 Å) of the covalent radii for gold (1.36 Å) and oxygen (0.66 Å) [435], while a similar bond length (2.06 Å) has been observed in gold-based trifluoromethoxy complexes [436]. As shown in Table 4.1, the positive sign of the Mulliken charge [384] on the gold atom is indicative of charge depletion and a loss of electron density from the gold atom to the surface. In contrast, the formerly-ether oxygen atom was found to have a Mulliken charge [384] of $-0.30\,|e|$, which indicates charge accumulation. The effective ionic valence, which is the difference between the formal and Mulliken charges of the anion, can be used as a measure of ionicity/covalency; an effective ionic valence of $0\,|e|$ would indicate full ionicity

while larger values would indicate increasing levels of covalency [437]. If the Au–O bond is assumed to be ionic, then the formal charge number of the oxygen anion would be −1, which would result in an effective ionic valence of 0.70 $|e|$. Ionic compounds such as sodium fluoride, sodium chloride and magnesium oxide were evaluated to have effective ionic valences less than 1.5 $|e|$ [437], which would suggest that the interaction between the gold and the former-ether oxygen atoms is more ionic than covalent. As mentioned, an assumption was made by treating the Au–O bond as ionic for the calculation of the effective ionic valence, while Mulliken charge decompositions have inherent issues of their own, as discussed above. However, this analysis can still give an indication that the interaction is somewhere between a polar covalent bond and an ionic bond rather than a non-polar covalent bond, which is to be expected due to oxygen being more electronegative than gold [438].

Boron-doped surface: As can be seen from Table 4.1, the boron-doped systems result in a much stronger adsorption energy of the single gold adatom as compared to the idealised system, though not as strong as the SCOV-defective system. This increased stability compared to the idealised system is not unexpected, as boron dopants within other materials have been reported to increase the adsorption energy of hydrogen [439; 440; 441; 442] and metal atoms such as calcium [443; 444] and sodium [445; 446]. The stronger adsorption observed for boron-doped surfaces as opposed to the idealised, undoped surface occurs as boron dopants possess one fewer valence electron than the carbon atoms in diamond. Such p-type dopants form an electron-deficient region that the metal adatom is attracted to [445]. While the difference between the adsorption energies for the localised cases is only 0.09 eV, the 0.68 Å disparity in adsorption height is much more noticeable. This is due to the location of the boron dopant within the surface layers. In the surface where the dopant is in the second layer, the boron atom lies below an ether oxygen atom, whereas the boron dopant within the third layer lies below a carbonyl oxygen atom, as visualised in Figure 4.2. As explained above, the gold atom is attracted to the electron-deficient regions caused by p-type dopants such as boron [445] and thus moves closer to them. This is achieved by the gold atom adsorbing above the ether and carbonyl oxygen atoms that lie atop the second- and third-layer dopants, respectively, as are visualised in Figures 4.5(c) and (d) respectively.

In contrast, the adsorption energy and height calculated from the delocalised model do not differ significantly from the localised model with the boron atom in the third layer, with the adsorption energy and height differing by only 0.23 eV and

0.01 Å respectively. This similarity is to be expected, because as the dopant moves from the surface layers into the bulk, the localised results from lower layers should converge to the delocalised result. Unlike the idealised case, the introduction of a charge within the delocalised model causes the structure of the surface atoms to change and accommodate the gold atom. This surface rearrangement allows the gold atom to get closer to an ether oxygen atom and sit in between two carbonyl oxygen atoms, resulting in a smaller adsorption height and a larger adsorption energy than the idealised surface.

In general, it can be seen that the idealised, fully-oxygenated diamond (110) surface results in very weak adsorption of the gold atom, which is likely to have very low kinetic barriers. However, as Table 4.1 shows, the introduction of defects or dopants into the surface significantly increase the adsorption energy of the gold atom. In particular, the SCOV defect was found to result in a very strong adsorption energy of -2.31 eV, and after projecting the density of states of the gold atom and its neighbouring carbon and oxygen atoms, it can be deduced that this strong adsorption is due to the formation of a polar covalent bond between the gold adsorbate and the diamond surface. Despite the interaction not being as strong as the defect studied here, the introduction of boron dopants, both localised and delocalised, was also found to increase the stability of the single gold atom on the surface.

4.3.3 Benchmarking Computational Methods

Having established that defects and dopants seem to promote the adsorption of single gold atoms, it is important to confirm that these trends are retained irrespective of what forcefield, dispersion correctionand DFA is chosen within the QM/MM method. This section will benchmark the embedding REBO [428; 429] forcefield against another forcefield, the TS [262] against other *a posteriori* dispersion correction schemesand the PBE [229] GGA against other DFAs.

Benchmarking MM Forcefields

As mentioned above, the QM region was embedded within an MM region that was treated using the REBO [428; 429] forcefield. It is important to ensure that the adsorption energetics of a single gold atom are not significantly affected by the choice of the embedding forcefield environment. To investigate this further, the adsorption energy of a single gold atom on an idealised surface as calculated using PBE^{+TS}/REBO was benchmarked against the Tersoff [430] forcefield, where

PBE$^{+\text{TS}}$ was used as the complementary QM method. Similar to REBO, the Tersoff forcefield was developed specifically for carbon, with applications to amorphous carbon [430], and is thus an appropriate forcefield to benchmark the REBO forcefield against.

Table 4.2 benchmarks the REBO forcefield against the Tersoff [430] forcefield. As can be seen, both the REBO and Tersoff forcefields result in the same adsorption energy for a single gold atom on an idealised surface. Furthermore, there is a very small disparity in adsorption height (0.05 Å), showing that both forcefield methods predict virtually identical adsorption energetics for the single gold atom, and that the REBO forcefield is appropriate to embed the QM region within.

MM Forcefield	Adsorption Height (Å)	Adsorption Energy (eV)
REBO [428; 429]	1.71	−0.30
Tersoff [430]	1.76	−0.30

Table 4.2. Adsorption energetics of a single gold atom adsorbed onto an idealised oxygen-terminated diamond (110) surface after a PBE$^{+\text{TS}}$/MM optimisation, using various MM forcefield methods.

Benchmarking *A Posteriori* Dispersion Correction Schemes

As discussed in Section 2.2, it is important to ensure long-range dispersion effectssuch as vdW forcesare treated appropriately as they can have a significant effect on the adsorption structure. Thus far in this chapter, only the pairwise additive TS scheme was been used alongside the PBE GGA, which does not explicitly account for beyond-pairwise vdW interactions. For this reason, the TS scheme was benchmarked against some *a posteriori* MBD [285; 286] correction schemes, namely the MBD@rsSCS [287] and MBD-NL [288] variants, which are described in Section 2.2.2. PBE$^{+\text{TS}}$/REBO was also benchmarked against non-dispersion-corrected PBE i.e. PBE/REBO calculations. The performances of various dispersion corrections were benchmarked by comparing the final adsorption energy and adsorption height after a full dispersion-corrected PBE/REBO geometry optimisation, and by constructing binding energy curves by running a series of dispersion-corrected PBE/REBO single-point calculations with the gold adatom placed at various heights above the surface, which represent the variation of the adsorption energy as a function of adsorption height.

Dispersion corrections were benchmarked on the idealised, SCOV-defective and the delocalised triel-doped systems. The idealised and SCOV-defective systems were chosen as they would permit dispersion corrections to be benchmarked on both 'more physisorbed' and 'more chemisorbed' systems respectively. The delocalised-triel doped system was chosen for three reasons: firstly, with common boron dopant densities, the probability of the dopant atom being within the bulk material is much higher than it being in the top surface layers. Secondly, the delocalised model is applicable to any triel dopantand thirdly, the predicted adsorption height and energy do not differ significantly from the localised case with the boron dopant in the third layer, as can be seen in Table 4.1.

Dispersion Correction	Adsorption Height (Å)	Adsorption Energy (eV)
Idealised Surface		
TS [262]	1.71	−0.30
MBD@rsSCS [287]	1.60	−0.29
MBD-NL [288]	1.63	−0.27
No Dispersion	1.82	−0.15
SCOV-Defective Surface		
TS	−0.17	−2.31
MBD@rsSCS	−	−
MBD-NL	−0.13	−2.29
No Dispersion	−0.11	−2.04
Delocalised Triel-Doped Surface		
TS	0.36	−1.97
MBD@rsSCS	1.06	−1.73
MBD-NL	1.06	−1.71
No Dispersion	1.09	−1.53

Table 4.3. Adsorption energetics of a single gold atom adsorbed onto various oxygen-terminated diamond (110) surfaces after a dispersion-corrected PBE/REBO optimisation, using various *a posteriori* dispersion correction schemes. No data were attained using MBD@rsSCS for the SCOV-defective system due to the negative polarisabilities for some atoms after the initial FHI-aims calculation settings.

Table 4.3 details the adsorption heights and energies of a single gold atom after TS-, MBD@rsSCS-, MBD-NL-, and non-dispersion-corrected PBE/REBO optimisations of the idealised, SCOV-defectiveand delocalised triel-doped surfaces. As can be seen for the idealised surface, there is very little disparity between TS and the MBD approaches with respect to both adsorption heights and energies. Both MBD@rsSCS [287] and MBD-NL [288] perform very similarly to each other, and predict slightly weaker adsorption than TSdespite the gold atom adsorbing closer to the surface; however these differences are minor (0.11 Å and 0.03 eV at most for adsorption heights and energies respectively). The lack of a dispersion correctionhoweverwas found to have a more evident effect, with the gold atom adsorb-

ing 0.11 Å higher than with TS and 0.22 Å higher than with MBD@rsSCS. Furthermore, the adsorption energy of the gold atom was found to be even weaker without a dispersion correction. The non-dispersion-corrected results are in line with literature where non-dispersion-corrected PBE, as well as other GGAs, have been observed to underestimate adsorption energies and overestimate adsorption distances [68; 69; 220; 447], and highlights the importance of including a dispersion correction with such DFAs.

For the SCOV-defective surface, the TS method once again was found to perform quite well with respect to MBD-NL. After a PBE$^{+\text{MBD-NL}}$/REBO optimisation, the gold atom was found to adsorb 0.04 Å higher than with TS, and this small disparity is reflected in the adsorption energywhich is only 0.02 eV weaker than with TS. No data were attained using MBD@rsSCS for this system due to the negative polarisabilities for some atoms after the initial FHI-aims calculation settings. It should be noted that this is a technical limitation of the MBD@rsSCS approach that can occur in some systems under certain conditions, and is not physically meaningful. Once again, without any dispersion correction, the gold atom was found to adsorb higher than dispersion-corrected approaches, and the adsorption energy was evaluated to be at least 0.25 eV weaker, further attesting to the need for a dispersion correction.

Finally, for the delocalised triel-doped system, there does appear to be some dependency on the choice of dispersion correction. The TS correction predicts stronger adsorption than both MBD approaches, and the gold atom optimises to a site far closer to the surface. The MBD approaches perform very similar to each other, with the gold atom predicted to adsorb 1.06 Å above the surface with an adsorption energy just larger than -1.7 eV. The consistency between the MBD approaches and the relatively weaker adsorption energies are indicative of the beyond-pairwise interactions being taken into account, and these effects have a greater influence within this charged system rather than the neutral idealised and SCOV-defective systems. The lack of a dispersion correction yet again resulted in the gold atom adsorbing higher than dispersion-corrected approaches, and the adsorption energy was evaluated to be at least 0.18 eV weaker.

Figure 4.7 shows the binding energy curves as calculated using various dispersion-corrected PBE/REBO calculations on the idealised, SCOV-defective and delocalised triel-doped surfaces. As can be seen from Figure 4.7(a) for the idealised surface, all

dispersion-corrected curves have similar shapes with adsorption energy minima between -0.12 eV and -0.11 eV at an adsorption height of 3.0 Å, showing that there is no major dependency on which dispersion correction is used. Once again, the MBD approaches predict slightly weaker adsorption than the pairwise TS scheme. However, without any dispersion correction, the adsorption energy minimum reduces to -0.05 eV, indicating very little, near-zero adsorption, as has been reported in Table 4.3 and in literature [68; 69; 220; 447].

Moving onto Figure 4.7(b), which shows binding energy curves for the SCOV-defective surface, all dispersion-corrected PBE/REBO [229] curves have similar shapes with adsorption energy minima between -0.27 eV and -0.19 eV at an adsorption height of 3.5 Å with TS [262] and MBD-NL [288], respectively. This shows that there is a slight disparity depending on what dispersion correction is used, but it is not a major difference. Furthermore, the minima of the curves are much deeper for the SCOV-defective surface than the idealised surface. However, without any dispersion correction, the adsorption energy minimum reduces to -0.05 eV, which is very similar to that of the idealised surface, and further showcases the importance of including a dispersion correction for the SCOV defect.

Finally, all the curves for the delocalised triel-doped surface have a minimum at a height of 2.0 Å, as can be seen in Figure 4.7(c). For this surface, both TS and MBD@rsSCS have near-identical results, with an adsorption energy value of -0.76 eV at 2.0 Å. MBD-NL predicts a similar curve to these two dispersion schemes, with an adsorption energy value of -0.73 eV at 2.0 Å, though some differences arise at around 3.0 Å. Yet again, a lack of a dispersion correction results in a shallower curve with an adsorption energy of -0.60 eV at 2.0 Å.
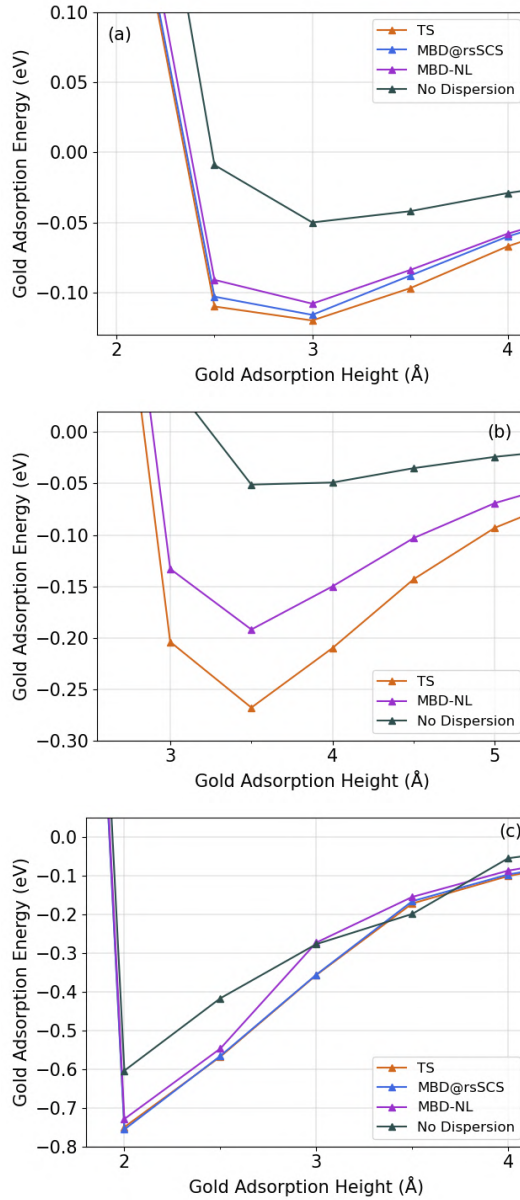
Figure 4.7. Binding energy curves showing the adsorption energy of a single gold adatom on various oxygen-terminated diamond (110) surface substrates as a function of height above the plane of carbonyl oxygen atoms on the substrate surface after dispersion-corrected PBE/REBO calculations. Substrates are (a) an idealised oxygen-terminated diamond (110) surface (b) a SCOV-defective surface and (c) a delocalised triel-doped surface.

Overall, after comparing the TS scheme against the MBD@rsSCS and MBD-NL dispersion schemes, no major dependency on the flavour of dispersion correction can be observed. However, a lack of a dispersion correction was found to result in

a weaker adsorption energy and a larger adsorption height, showing the importance of accounting for vdW effects if the DFA does not include any mid-/long-range dispersion interactions.

Benchmarking Density-Functional Approximations

Now that both the REBO forcefield and the TS dispersion correction have been shown to provide accurate results, the final factor that remains is the choice of DFA. The choice of DFA can have a large effect on the absolute adsorption energy value, as was shown in Table 3.2 in Chapter 3. As was done when benchmarking dispersion corrections, DFAs were benchmarked on the idealised, SCOV-defective and the delocalised triel-doped systems by first comparing the final adsorption energy and adsorption height after a full QM/REBO geometry optimisation. After this, binding energy curves were constructed to see how the adsorption energy of the single gold atom changes as a function of adsorption height with different DFAs.

Idealised surface: Figure 4.8(a) details the performance of various DFAs on an idealised surface after a full QM/REBO optimisation. All DFAs predict fairly weak adsorption of the single gold atom, with adsorption energies ranging from $-0.04\,\text{eV}$ to $-0.67\,\text{eV}$. Overall, there is an inverse relationship between the final adsorption height and the adsorption energy, which is to be expected because a smaller adsorption height is generally reflective of a stronger interaction between the adsorbate and substrate. Different rungs of DFAs (along Jacob's ladder [219]) generally seem to produce results that group together in specific areas. LDAs (GDSMFB [421], KSDT [422], and PZ-LDA [423; 424]) were found to predict the largest adsorption energy (between $-0.66\,\text{eV}$ and $-0.67\,\text{eV}$). This is in line with observations that LDAs typically overestimate binding between metal and organic/carbonaceous interfaces, which results in overestimated adsorption energies and underestimated adsorption heights [69]. With this in mind, the LDA values can be thought of as an upper bound on the true adsorption energy between the single gold atom and the surface. Most GGAs, MGGAs and HGGAs are grouped together, apart from the RPBE [233] GGA and the TPSS [234] MGGA, while the PBEsol0 [244] HGGA sits in between these two groups.

Some differences in adsorption energy (and height) arise depending on the choice of (TS-corrected) GGA. The revPBE [231] GGA predicts stronger adsorption than PBE [229] ($-0.40\,\text{eV}$ as opposed to $-0.30\,\text{eV}$), though their values differ by only $0.1\,\text{eV}$. This is to be expected as both PBE and revPBE possess the same

mathematical form for $\mathcal{F}_x$, as expressed in Equation (2.20). The PBEsol [232] GGA predicts a similar adsorption energy of $-0.42\,\text{eV}$ and also has a similar formulation to PBE , as explained in Section 2.1.4. The similarities between the PBE, revPBE and PBEsol formulations for $\mathcal{F}_x$ indicate why these GGAs predict fairly similar adsorption energetics. The RPBE [233] GGAhoweverpossesses a different mathematical form for $\mathcal{F}_x$, and has been shown to not perform very well for physisorbed systems where vdW effects govern adsorption [289; 448; 449], which could explain the disparity between results attained using RPBE and other PBE-like GGAs.

Most of the MGGAs predict similar adsorption energetics to each other and to most GGAs, with the only exception being the TPSS [234] MGGA, which predicts similar adsorption energetics to the RPBE [233] GGA. The TPSS [234] MGGA also uses Equation (2.20) to calculate $\mathcal{F}_x$, but uses a different parameter to the χs^2 term, which could explain the disparity to the majority of PBE-like GGAs. Studies have sought to correct for this disparity by building TPSS-like MGGAs and 'fitting' to GGA results. The TPSSloc [236] MGGA uses a localised PBE-like DFA for the correlation within a TPSS-like DFA form, while the revTPSS [426] formulation is based on the PBEsol modification to the PBE correlation. These corrections might explain why the TPSSloc [236] and revTPSS [426] results align better with GGA results. The M06-L [235] MGGA also includes a PBE-like component within its formulation, as explained in Section 2.1.4, which might also explain its similar performance to PBE-derived DFAs, while its slightly stronger adsorption energy for the gold atom as compared to PBE has been previously observed [450]. In general, both SCAN-based MGGAs predict slightly stronger adsorption than the TPSS-based MGGAs, though they still predict very similar adsorption energetics to revTPSS and TPSSloc, despite the different nature of $\mathcal{F}_x$ within the SCAN formulation. $\mathcal{F}_x$ in SCAN obeys the conjectured $\mathcal{F}_x \leq 1.174$ bound for all densities [237; 238; 239], while within (rev)TPSS and PBE, $\mathcal{F}_x$ monotonically tends to the general Lieb-Oxford bound (1.804) [240]. The fact that these different formulations with different properties result in very similar adsorption energetics provides confidence in the accuracy of their predictions.

Figure 4.8. Plots benchmarking the performance of various DFAs on an idealised oxygen-terminated diamond (110) surface. Shown are: (a) Scatter graph showing the adsorption energy and adsorption height of a single gold adatom after a full geometry optimisation, and (b) Binding energy curves showing the adsorption energy of a single gold adatom as a function of height above the substrate surface. In (b), DFAs are divided according to (from left to right): LDAs, TS-corrected GGAs, MGGAs and TS-corrected HGGAs.

The HSE03 [241], HSE06 [242] and PBE0 [243] HGGAs predict similar adsorption energetics to all GGAs apart from RPBE [233]. This is to be expected as 75% of the PBE0 exchange energy is from PBE [243], while 75% of the exchange energy for both HSE DFAs is from PBE, and the remaining 25% is split between a long-ranged PBE-like component and a short-ranged, screened Hartree-Fock component [241; 242], as explained in Equation (2.22). The PBEsol0 [244] HGGA predicts a much weaker adsorption energy than other HGGAs as well as the PBEsol [232] GGA, which accounts for 75% of the exchange energy within PBEsol0. This is

somewhat surprising given the agreement seen between PBE, PBEsol and other HGGAs but could indicate that mixing exchange energy from PBEsol and Hartree-Fock components, as is done within PBEsol0 [244], can lead to contrasting results (for this system at the very least).

Moving onto the binding energy curves shown in Figure 4.8(b), all DFAs result in a curve with a minimum between 2.5 Å and 3.5 Å, before tending to an adsorption energy of 0 eV as adsorption height increases. LDAs have an adsorption energy minimum of -0.26 eV at an adsorption height of 2.5 Å which is much deeper than other QM methods. A deeper minimum is more indicative of stronger binding, and this is once again in line with observations that LDAs predict stronger adsorption [69]. The PZ-LDA [423; 424] and KSDT [422] results are indistinguishable over the range of adsorption heights investigated, and are identical to the GDSMFB [421] results over 2.5 Å.

With the MGGAs, DFAs within the same families have similar binding energy curves. TPSSloc [236] and revTPSS [426] have adsorption energy minima of -0.03 eV and -0.04 eV respectively at an adsorption height of 3.0 Å. This adsorption height is the same as that of PBE [229], but the adsorption energies are much smaller, which explains why these two MGGAs predict weaker adsorption than PBE [229] in Figure 4.8(a). While TPSS [234] has a similar adsorption energy minimum of -0.03 eV, this occurs at an adsorption height of 3.5 Å, which is the same height as the RPBE [233] GGA, albeit a lower adsorption energy. This same adsorption height explains why the RPBE [233] and TPSS [234] data points are so close together in Figure 4.8(a), and the lower adsorption energy explains why TPSS [234] predicts slightly weaker adsorption than RPBE [233]. The SCAN [237] and rSCAN [425] MGGAs have similar binding energy curves, with minima of -0.10 eV at 2.5 Å, which is a similar adsorption energy minimum to the PBE [229] GGA and the same adsorption height as the revPBE [231] and PBEsol [232] GGAs; this trend is reflected by the positions of the SCAN [237] and rSCAN [425] data points in Figure 4.8(a). M06-L [235] has an adsorption energy minimum at -0.08 eV at an adsorption height of 3.0 Å, similar values to the PBE [229], SCAN [237] and rSCAN [425] DFAs, and this is reflected by the position of the M06-L [235] data point in Figure 4.8(a).

The HSE03 [241], HSE06 [242] and PBE0 [243] HGGAs have very similar binding energy curves, with adsorption energy minima at -0.11 eV at an adsorp-

tion height of 3.0 Å. This close agreement explains why these HGGAs are so close together in Figure 4.8(a), and the similar adsorption energy minima and adsorption height values to other DFAs further explains the proximity of these HGGAs to the majority of GGAs and MGGAs. However, the PBEsol0 [244] HGGA has a much shallower adsorption energy minimum of -0.06 eV at 3.0 Å. Despite the similarity between the PBEsol0 [244] and revTPSS [426] curves, the PBEsol0 [244] data point in Figure 4.8(a) sits quite close to the RPBE [233] and TPSS [234] data points.

SCOV-defective surface: The second substrate of interest was a surface with a SCOV defect. To ensure the SCOV defect was correctly modelled with every DFA, DFA_i, the carbonyl oxygen was first removed and the surface was re-optimised using DFA_i/REBO. After this initial optimisation, the surface structure at the defect site, centred at the former carbonyl carbon atom, should change from bent (originally trigonal planar with the carbonyl oxygen atom in the idealised system) to trigonal pyramidal. Because diamond surfaces are usually hydrogen-terminated after CVD growth [145], uncoordinated carbon atoms were subsequently saturated with hydrogen species and the surface was reoptimised using DFA_i/REBO, after which the surface structure at the defect site should change to tetrahedral. Based on valence shell electron pair repulsion theory, this shows a return of the ·C· atom to an sp^3-hybridised state. This is the correct surface configuration as an oxygen atom is needed to pull the carbon atom above the diamond (110) surface plane to form a carbonyl group at the surface, as shown in Figure 3.6. Without this oxygen atom, the carbon atom would remain in an sp^3-hybridised configuration.

Only the DFAs that returned the former-carbonyl carbon atom to an sp^3-hybridised configuration were considered herein. This was evaluated by studying the conformational isomerism of the structure centred at the former-carbonyl carbon atom. After the removal of the carbonyl oxygen atom and optimisation with a given DFA, the dihedral angle between a surface ether oxygen atom and a surface carbon atom, along the bond between the former-carbonyl and corresponding ether carbon atoms was calculated. Table 4.4 details the calculated dihedral angles after optimisation with various DFAs. As can be seen, most DFAs correctly return the structure to a synclinal conformation, with dihedral angles of approximately 60°, which indicates a more sp^3-hybridised configuration. However, the SCAN [237] MGGA and all investigated HGGAs result in an anticlinal conformation, with dihedral angles of approximately 150°, which indicates the former-carbonyl carbon atom remains in a more sp^2-hybridised state. Even though the higher-rung HGGAs

predict the anticlinal conformation, this is most likely a local energy minimum and as explained above, is not the correct physical conformation for the surface after the removal of a carbonyl oxygen atom. Therefore, the SCAN [237] MGGA and the PBE0 [243], PBEsol0 [244], HSE03 [241], and HSE06 [242] HGGAs were not studied for this system. For clarity, the Newman projections [451] of the synclinal and anticlinal conformations are provided in Figure 4.9.

DFA	Dihedral Angle (°)
PZ-LDA [423; 424]	54.3
KSDT [422]	54.3
PBE [229]	55.7
revPBE [231]	55.2
RPBE [233]	56.1
PBEsol [232]	55.1
TPSS [234]	56.4
TPSSloc [236]	56.0
revTPSS [426]	56.3
SCAN [237]	142.7
PBE0 [243]	141.3
PBEsol0 [244]	140.9
HSE03 [241]	141.1
HSE06 [242]	141.1

Table 4.4. Dihedral angles between a surface ether oxygen atom and a surface carbon atom, along the bond between the formerly-carbonyl and corresponding ether carbon atoms after the removal of a carbonyl oxygen atom from the idealised oxygen-terminated diamond (110) surface and subsequent optimisation with various DFAs. DFAs are ordered from low- to high-rung (along Jacob's ladder [219]), and the TS [262] dispersion correction method was applied to all GGAs and HGGAs.
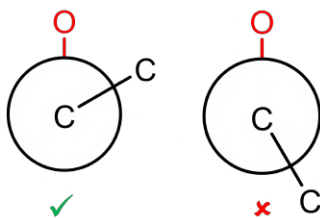
Figure 4.9. Newman projections [451] of the synclinal (left) and anticlinal (right) conformations after optimisation with various DFAs. The projection is along a bond between the former-carbonyl carbon atom and a surface ether carbon atom. The synclinal conformation is the correct model for the SCOV defect (prior to hydrogen saturation).

Figure 4.10(a) details the performance of various DFAs on a SCOV-defective surface. As can be seen from this figure, the introduction of a SCOV defect at the surface significantly increases the range of adsorption energies as compared to the idealised surface, from -0.03 eV to -2.84 eV. This range of adsorption energy values indicate that while some DFAs still predict weak adsorption, some DFAs such as PBE [229], revPBE [231] and revTPSS [426] predict much stronger adsorption and a possible covalent bonding interaction between the gold adatom and the substrate surface, as discussed earlier.

Both LDAs, PZ-LDA [423; 424] and KSDT [422], predict similar adsorption energies to each other: -0.67 eV and -0.63 eV respectively. However, there is quite a large range of adsorption energies predicted amongst TS-corrected GGAs, and all GGAs apart from RPBE [233] predict stronger adsorption than the LDAs. The revPBE [231] and PBE [229] GGAs predict very strong adsorption (-2.84 eV and -2.31 eV respectively), with the former predicting stronger adsorption much like in the idealised case. Their negative adsorption heights means that the gold adatom sits below the plane of carbonyl oxygens i.e. it sits within the 'well' caused by the vacancy. The PBEsol [232] GGA predicts weaker adsorption than PBE but strong adsorption nonetheless with an adsorption energy of -1.35 eV. Much like in the idealised case, RPBE [233] GGAhoweverstill predicts a very weak adsorption energy of -0.18 eV and predicts the gold adatom to adsorb 1.94 Å above the surface. This is quite unusual as RPBE [233] has been reported to typically perform better for chemisorbed systems [448].

Much like with GGAs, there is a vast range of adsorption energies predicted by various MGGAs. The revTPSS [426] MGGA predicts a very strong adsorption

energy of $-2.20\,\text{eV}$, slightly weaker than the PBE [229] GGA as in the idealised case. The negative adsorption height indicates that revTPSS [426] also predicts the gold adatom to sit below the plane of carbonyl oxygens. TPSS [234], however, much like the RPBE [233] GGA, also predicts a very weak adsorption energy of $-0.03\,\text{eV}$, with the gold adatom adsorbing $2.48\,\text{Å}$ above the surface. The performance of TPSSloc [236]howeverdiffers quite a lot from the idealised case, with the MGGA predicting a very weak adsorption energy of $-0.22\,\text{eV}$, though the adatom is adsorbed closer to the surface than predicted by RPBE [233] and TPSS [234], with an adsorption height of $1.02\,\text{Å}$.

Moving onto the binding energies for the SCOV-defective surface that can be seen in Figure 4.10(b), DFAs generally predict stronger adsorption than in the idealised surface. Note that no binding energy curve for the KSDT [422] LDA is reported due to a DFA-specific error within the FHI-aims calculation: 'no classical turning points were found when solving for the radial equation'. LDAs and most GGAs predict similar binding energy curves, with revPBE [231] and PZ-LDA [423; 424] predicting the strongest adsorption at an adsorption height of $3.0\,\text{Å}$. PBE [229] and PBEsol [232] have similar curves to each other, despite the disparity in optimal adsorption structures as seen in Figure 4.10(a). The RPBE [233] GGA, however, once again predicts the weakest adsorption among GGAs and has a binding energy minimum at $4.0\,\text{Å}$, which is a larger adsorption height value than all other LDAs and GGAs, much like in the idealised case.
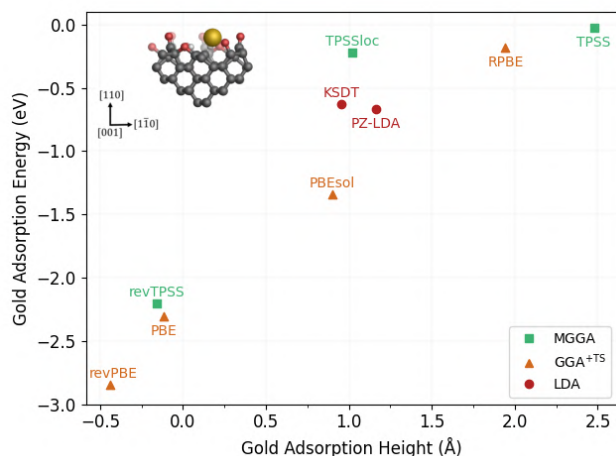
(a)
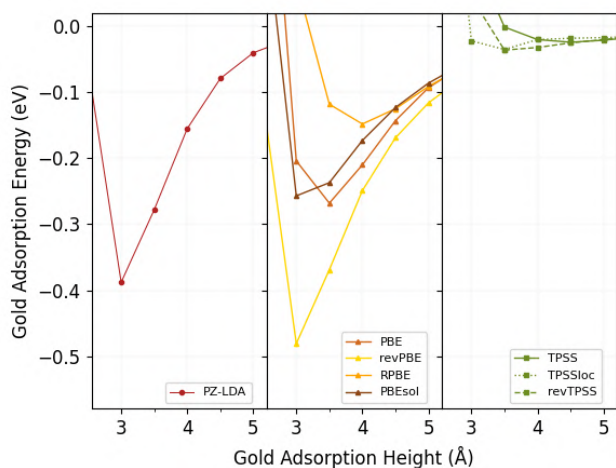
(b)

Figure 4.10. Plots benchmarking the performance of various DFAs on a SCOV-defective oxygen-terminated diamond (110) surface. Shown are: (a) Scatter graph showing the adsorption energy and adsorption height of a single gold adatom after a full geometry optimisation, and (b) Binding energy curves showing the adsorption energy of a single gold adatom as a function of height above the substrate surface. In (b), DFAs are divided according to (from left to right): LDAs, TS-corrected GGAs and MGGAs.

Curiously, all MGGAs result in very shallow binding energy curves for the single adatom. This is to be expected for both TPSS [234] and TPSSloc [236] as they predict very weak adsorption in Figure 4.10(a). Furthermore, the binding energy curve for TPSS [234], much like RPBE [233], has a minimum at a value larger than all other DFAs. However, the revTPSS [426] binding energy curve is very similar to other MGGAs' despite the strong adsorption predicted in Figure 4.10. This reflects

the importance of local surface relaxation upon adsorption of the single adatom.

Delocalised triel-doped surface: Figure 4.11(a) details the performance of various DFAs on the final substrate of interest, which was a delocalised triel-doped surface. Once again, the introduction of a charge into the surface significantly increases the range of adsorption energies as compared to the idealised surface, from $-0.03\,\text{eV}$ to $-2.84\,\text{eV}$. This range of adsorption energy values indicate that all DFAs generally predict fairly strong adsorption between the single gold atom and the surface. There is once again a general inverse relationship between the adsorption heights and energies, though DFAs are generally grouped into two areas of adsorption heights: 0–$0.4\,\text{Å}$ and 0.9–$1.5\,\text{Å}$ above the plane of carbonyl oxygen atoms. This difference in adsorption heights is primarily due to the final optimised adsorption site. In the smaller set of adsorption heights, 0–$0.4\,\text{Å}$, the structure of the surface atoms changes to accommodate the gold atom, and the gold atom is able to get closer to an ether oxygen atom and sit in between two carbonyl oxygen atoms, resulting in a smaller adsorption height and a larger adsorption energy. In contrast, for the larger 0.9–$1.5\,\text{Å}$ set of adsorption heights, the surface does not change as much and sterically hinders the gold atom from getting closer to the ether oxygen atom. The gold atom therefore binds to the carbonyl oxygen atom, resulting in a larger adsorption height and a weaker adsorption energy.

The PZ-LDA [423; 424] predicts a very strong adsorption energy of $-3.06\,\text{eV}$, which is second-strongest interaction of all investigated DFAs and only $0.04\,\text{eV}$ weaker than the HGGA PBEsol0 value. This strong adsorption energy is reflected in the final adsorption height of the gold atom, $0.23\,\text{Å}$, which is very close to the surface. The PBE [229], revPBE [231] and PBEsol [232] GGAs once again predict fairly similar adsorption energies to each other. Much like in the idealised case, revPBE [231] ($-2.11\,\text{eV}$) predicts stronger adsorption than PBE [229], while PBEsol [232] ($-2.49\,\text{eV}$) predicts slightly stronger adsorption than both PBE [229] and revPBE [231]. Overall, the adsorption energies are quite similar to each other, which is to be expected given their similar formulation for $\mathcal{F}_\text{x}$. However, some differences arise in the final adsorption heights of the single gold atom. Both PBE [229] and PBEsol [232] result in the gold atom being adsorbed fairly close to the surface, with adsorption heights of $0.36\,\text{Å}$ and $0.28\,\text{Å}$ respectively. However, the final adsorption height with revPBE [231] is relatively much higher than PBE(sol), with the single gold atom predicted to adsorb $0.91\,\text{Å}$ above the surface. The RPBE [233] GGA once again predicts much weaker adsorption than the other PBE-based GGAs,

and the weakest of all investigated DFAs, though the adsorption energy itself is quite strong, unlike for the idealised and SCOV-defective surfaces. This relatively weak adsorption energy is reflected in its predicted adsorption height, which is the highest among all investigated DFAs.
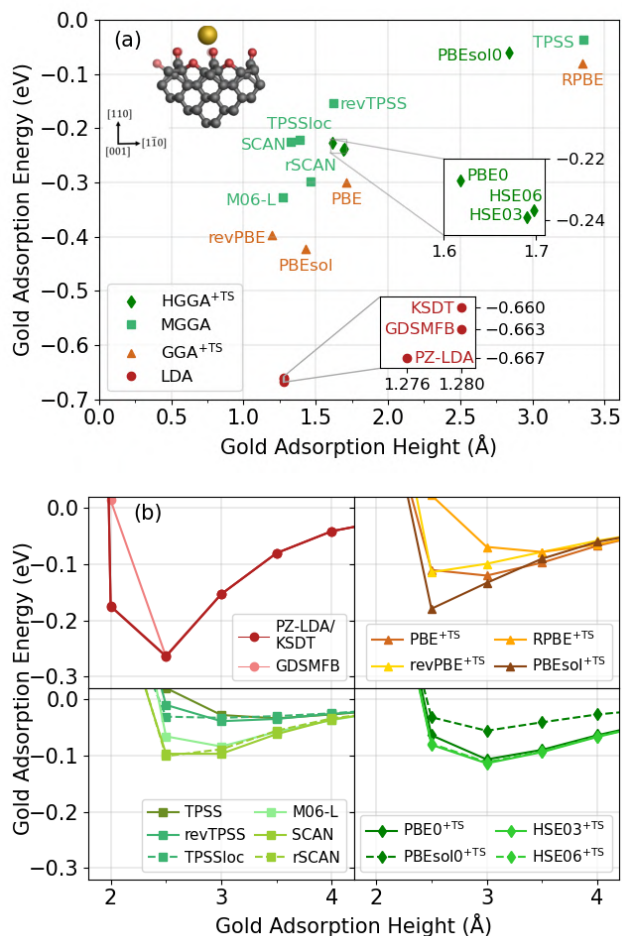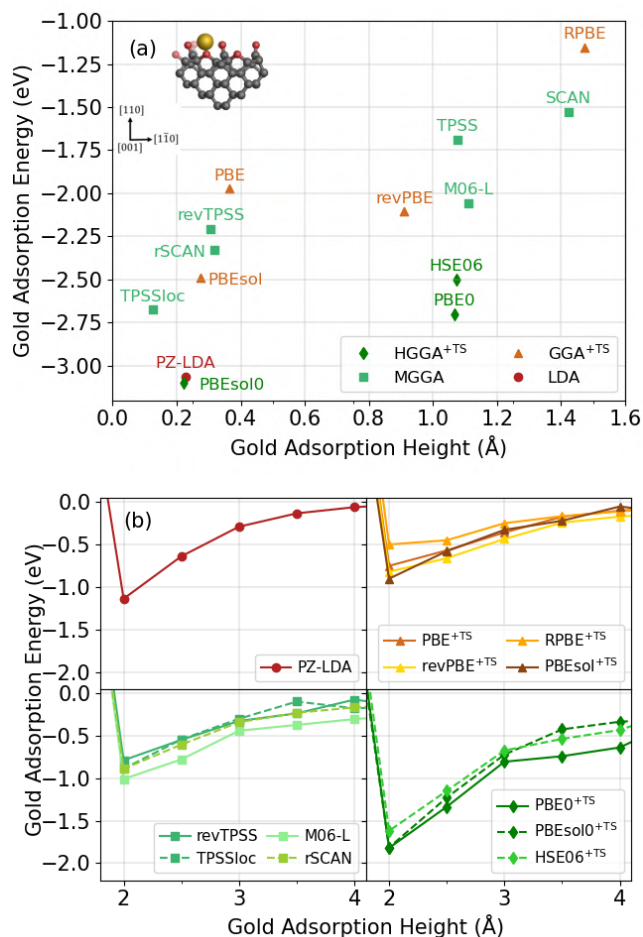

Figure 4.11. Plots benchmarking the performance of various DFAs on a delocalised triel-doped oxygen-terminated diamond (110) surface. Shown are: (a) Scatter graph showing the adsorption energy and adsorption height of a single gold adatom after a full geometry optimisation, and (b) Binding energy curves showing the adsorption energy of a single gold adatom as a function of height above the substrate surface. In (b), DFAs are divided according to (from left to right): LDAs, TS-corrected GGAs, MGGAs and TS-corrected HGGAs.

MGGAs are split within the two aforementioned adsorption height ranges. The SCAN [237] MGGA predicts the second-weakest adsorption $(-1.53\,\text{eV})$ of all investigated DFAs, and this is reflected in the second-largest adsorption height

of 1.43 Å, which is not too dissimilar to the RPBE-predicted adsorption height. The TPSS [234] and M06-L [235] MGGAs predict stronger adsorption than both RPBE [233] and SCAN [237] with adsorption energies of $-1.69\,\mathrm{eV}$ and $-2.06\,\mathrm{eV}$ respectively, and this is reflected in the lower adsorption heights of 1.08 Å and 1.11 Å respectively. Both TPSS [234] and M06-L [235] predict similar adsorption heights to the PBE0 [243] and HSE06 [242] HGGAs, but predict weaker adsorption energies. In contrast, revised versions of TPSS [234] and SCAN [237], namely revTPSS [426], TPSSloc [236] and rSCAN [425], generally predict stronger adsorption energies of $-2.21\,\mathrm{eV}$, $-2.68\,\mathrm{eV}$ and $-2.33\,\mathrm{eV}$ respectively, with the gold atom adsorbing much closer to the surface with adsorption heights of 0.31 Å, 0.13 Å and 0.32 Å respectively. The revTPSS [231] and rSCAN [425] results are right in the middle of the PBE [229] and PBEsol [232] results in Figure 4.11(a), while the TPSSloc [236] result shows only slightly stronger adsorption than these four DFAs. Despite the difference in adsorption heights, these MGGAs also predict a similar adsorption energy to the revPBE [231] GGA. The similarity between the revised MGGAs and the PBE-based GGAs can be attributed to their GGA-based formulation. As discussed before, TPSSloc [236] includes a PBE-like component, while revTPSS [426] is based on the PBEsol modification to PBE. These MGGAs were designed by fitting to GGA results [236; 426] and could explain the level of agreement observed between GGAs and revised MGGAs.

Unlike for the idealised and SCOV-defective surfaces, some disparities between HGGAs and lower-rung DFAs can be seen in Figure 4.11(a), with HGGAs generally predicting stronger adsorption. As mentioned above, the PBEsol0 [244] HGGA predicts very similar adsorption energetics to the PZ-LDA [423; 424], with the strongest adsorption energy of all investigated DFAs ($-3.10\,\mathrm{eV}$) and a very small adsorption height of 0.22 Å, which is only 0.01 Å lower than the PZ-LDA-predicted value. Despite predicting stronger adsorption energies, PBEsol0 [244] predicts a similar adsorption height for the single gold atom as compared to the aforementioned revised MGGAs, PBE [229] and its base PBEsol [232] GGA. In contrast, the PBE0 [243] and HSE06 [242] HGGAs predict similar adsorption energies ($-2.70\,\mathrm{eV}$ and $-2.50\,\mathrm{eV}$ respectively) to PBEsol [232] and the revised MGGAs, but both predict a larger adsorption height of 1.07 Å, which agrees more with the TPSS [234] and M06-L [235] results. The PBE0 [243] and HSE06 [242] results can be seen to differ a fair amount from the PBE [229] result, despite both HGGAs including PBE components within their formulations.

Moving onto the binding energies for the delocalised triel-doped surface that can be seen in Figure 4.11(b), DFAs once again generally predict stronger adsorption than in the idealised surface. The curves for all DFAs have their binding energy minimum at an adsorption height of 2.0 Å. The PZ-LDA [423; 424] predicts fairly similar binding energy curves to GGAs, but has a larger binding energy minimum value of -1.14 eV. Of all investigated GGAs, PBEsol [232] has the deepest curve with a binding energy minimum of -0.90 eV, which is followed by revPBE [231] (-0.82 eV), then PBE [229] (-0.75 eV) and finally RPBE (-0.50 eV). This trend of binding energy minimum values is reflected in Figure 4.11(a), where as discussed above, the adsorption energy trend is PZ-LDA [423; 424] > PBEsol [232] > revPBE [231] > PBE [229] > RPBE [233].

The binding energy curves of MGGAs are quite similar to the GGA curves: the largest MGGA binding energy minimum arises from M06-L [235] (-1.01 eV). The rSCAN [425] binding energy minimum is at -0.89 eV, which is only 0.01 eV away from the PBEsol [232] value (-0.90 eV), and further explains the agreement seen between the two DFAs in Figure 4.11(a). The revTPSS [426] and TPSSloc [236] binding energy minima are at -0.79 eV and -0.89 eV, which are very close to the aforementioned rSCAN [425], PBEsol [232] and PBE [229] values. While the HGGA binding energy curves have a similar shape to lower-rung DFAs, all three investigated HGGAs have much deeper curves. Both PBE0 [243] and PBEsol0 [244] have binding energy minima at 1.82 eV, while the range-separated HSE06 [242] HGGA has a binding energy minimum at -1.62 eV. This could explain the trend in Figure 4.11(a) where HGGAs generally predict stronger adsorption for this system than lower-rung DFAs.

In summary, the PBE [229] GGA was generally found to be able to provide important insights into the nature of single gold atom adsorption on various oxygen-terminated diamond (110) surfaces. For the idealised and SCOV-defective surfaces, PBE [229] was able to accurately capture the physisorbed and chemisorbed natures of the gold adatom, respectively. A high degree of agreement was also observed with most other GGAs, as well as many higher-rung MGGAs and HGGAs. This is important as there is no existing experimental data that describes the adsorption energetics of single gold atoms on such surfaces. Some DFAs such as RPBE [233] GGA and the TPSS [234] MGGA were observed to perform poorly and should not be used for these surfaces. Some disagreement was found between PBE and higher-rung HGGAs for the delocalised triel-doped surface, though the PBE results were quite

similar to some MGGAs. The disparity with HGGAs indicates that PBE is perhaps not the ideal DFA for this surface, though it was still able to capture the stronger adsorption interaction with respect to the idealised surface. For this surface, it is perhaps more appropriate to use either HSE06 [242], PBE0 [243], revPBE [231] or M06-L [235] as the DFA. In general, this section has shown that the trends observed in Section 4.3.2 are robust with respect to the choice of embedding forcefield, *a posteriori* dispersion correction scheme and DFA.

4.3.4 Kinetic Stabilities

Having established how the adsorption energy of a single gold atom varies with different defects and dopants within oxygen-terminated diamond (110) surfaces, it is also important to calculate the kinetic stability of the single gold atom on various surfaces. Using STEM, Hussein *et al.* observed gold atoms to be very stable atop polycrystalline BDD surfaces, even withstanding the momentum transfer from the highly energetic electron beam ($\sim$200 kV), as explained in Section 1.1.1, and thermal baking [22], as is typically done to store samples before they are transferred to the microscope. Both of these factors can lead to the movement of gold atoms along the diamond substrate surface [22] if the kinetic barriers are overcome. However, the barriers associated with the idealised system were calculated to be very low [22] and unlikely to stabilise single gold atoms. This indicates that the high stability of single gold atoms observed by Hussein *et al.* [22] is likely due to surface defects and (boron) dopants that were not seen within their STEM images. For this reason, constrained QM/REBO optimisations were performed to calculate the energy required to move a single gold atom across idealised, SCOV-defective and boron-doped substrate surfaces after adsorption.
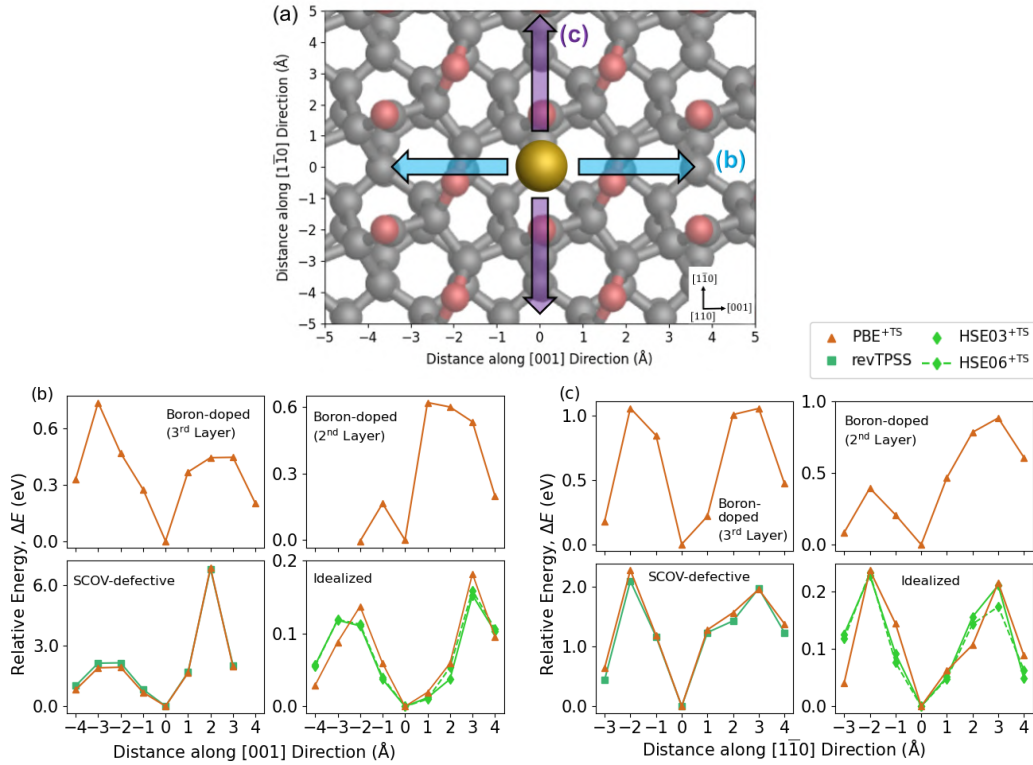
Figure 4.12. Relative energies (ΔE) of translating a single gold atom across various oxygen-terminated diamond (110) surface substrates from its initial adsorption site, which is placed at the origin. (a) Paths of motion along the idealised surface; (b) relative energies along the [001] direction; (c) relative energies along the [1$\bar{1}$0] direction. The Tkatchenko-Scheffler (TS) dispersion correction was used alongside the PBE, HSE03, and HSE06 density-functional approximations, but not alongside revTPSS.

Figure 4.12 details the relative energies of a single gold atom along the [001] and [1$\bar{1}$0] directions with respect to the initial adsorption site. The curves are not symmetrical about the origin as the structure is not symmetrical along the [001] and [1$\bar{1}$0] axes that pass through this point. As can be seen, along both directions, the surfaces that lead to stronger adsorption of the single gold atom have larger energetic barriers. More specifically, the introduction of defects and dopants increases the stability of the single gold atom along both investigated directions. This is to be expected as the gold adsorbate was shown to be more strongly bound to these surfaces, which means more energy would be required to overcome this interaction and translate the gold atom across the diamond surface.

As shown in both Figures 4.12(b) and (c), the on an idealised surface are quite low as compared to defective and doped surfaces. Such a low kinetic bar-

rier is to be expected as the gold atom is not strongly bound to the surface, as was shown in Table 4.1 and Figure 4.8. The energetic barriers to moving the gold atom along the [001] direction were calculated to be 0.14 eV and 0.18 eV for the negative and positive displacements respectively with PBE^{+TS}/REBO, which are in close agreement with the energy barrier of 0.16 eV that was predicted by Hussein *et al.* using a periodic PBE^{+TS}-optimised model of the idealised surface [22]. The relative energies along the [1$\bar{1}$0] direction were generally found to be higher with PBE^{+TS}/REBO, with barriers of 0.24 eV and 0.21 eV for the negative and positive displacements respectively, which are once again comparable to the energy barrier of 0.25 eV predicted by Hussein *et al.* using a periodic surface model [22]. The relatively higher barriers along the [1$\bar{1}$0] direction are to be expected, as the gold atom has to move above the plane of carbonyl oxygen atoms that are along this axis, as shown by the purple arrows in Figure 4.12(a). In contrast, for the [001] direction, the gold atom has to 'only' move above the plane of ether oxygen atoms to move across the surface, as shown by the blue arrows in Figure 4.12(a), which is at a lower height than the plane of carbonyl oxygen atoms and therefore would be expected to have a lower energy barrier.

While the PBE [229] GGA was shown to perform well with respect to other DFAs for the prediction of adsorption energetics in Section 4.3.3, a further comparison of barriers was conducted against the HSE03 [241] and HSE06 [242] HGGAs. Both of these HGGAs predict near-identical relative energies to each other. As can be seen in Figure 4.12(c), there is some difference in the shapes of their curves with respect to the PBE curve along the [1$\bar{1}$0] direction. However, both HSE03 and HSE06 predict energy barriers of 0.15 eV and 0.17 eV respectively for the positive displacement, and a barrier of 0.23 eV for the negative displacement. These values are only slightly lower than the PBE value but are still very similar, showing that the PBE [229] GGA remains an appropriate choice. For the [001] direction, PBE again predicts a similar barrier to the values predicted by both HGGAs (0.12 eV and 0.15 eV along the negative and positive displacements respectively. A thermal baking temperature of 60 °C, as was used by Hussein *et al.*, corresponds to a translational kinetic energy of 0.043 eV atom^{-1} [22], which is lower than the barriers calculated with all three DFAs for all displacements and directions, which means atomic movement along the surface during the aforementioned thermal baking process is unlikely.

Unlike the idealised surface, the SCOV-defective surface has very large kinetic barriers associated with it. Such large barriers are expected as the gold atom

sits within the 'well' caused by the defect, so a lot of energy would be required to overcome the interaction between the gold atom and the surface described earlier, remove the gold atom out of the defect site and move it across the diamond surface. Along the [001] direction, as shown in Figure 4.12(b), the barriers were calculated to be 1.93 eV and 6.86 eV along the negative and positive displacements respectively using PBE^{+TS}/REBO. This disparity between the displacement can once again be explained by the structural asymmetry. Along the negative displacement, as can be seen in Figure 4.5(c), the surface is hydrogen-terminated in the neighbourhood of the gold atom, which means a relatively lower energy would be required to move the atom across the surface. However, along the positive displacement, the gold atom has to move above a carbonyl oxygen atom, which requires much more energy. For the [1$\bar{1}$0] direction, the predicted energy barriers were also very high, with values of 2.28 eV and 1.96 eV along the negative and positive displacements respectively using PBE^{+TS}/REBO. Once again, the accuracy of PBE was benchmarked against the revTPSS [426] MGGA, which was shown to perform similarly to PBE in Section 4.3.3. The calculated curves and energy barriers with revTPSS [426] were found to agree very well with PBE, as can be seen in Figure 4.12(b) and (c).

Substituting a carbon atom with boron in the surface layers of the diamond substrate also increases the stability of the gold atom compared to the idealised system. For the [001] direction, the barrier along the negative displacement is larger when the boron dopant is in the third layer (0.74 eV) rather than the second layer (0.16 eV). However, the barrier along the positive displacement is larger when the boron dopant is in the second layer (0.62 eV) rather than the third layer (0.45 eV). Along the [1$\bar{1}$0] direction, the boron dopant being within the third layer results in a barrier of 1.03 eV along the negative displacement, whereas the second-layer boron results in a barrier of 0.39 eV. Along the positive displacement, the second- and third-layer barriers are 0.89 eV and 1.04 eV respectively. These barriers are lower than the ones associated with the SCOV-defective surface but clearly increase the kinetic stability of the single gold atom.

In general, it can be seen that the idealised, fully-oxygenated diamond (110) surface results in very low energetic barriers for the single gold atom. However, as Figure 4.12 shows, the introduction of defects or dopants into the surface significantly increase the of the gold atom. Similar to the trend observed with adsorption energies, the barriers associated with explicitly-modelled boron dopants were not as large as those associated with the SCOV defect, though they were also found to

109

increase the stability of the single gold atom on the surface. Furthermore, the barriers predicted for the idealised and SCOV-defective surfaces were found to be robust with respect to the choice of DFA. Overall, these barriers suggest that the gold atom should not move across the surface when they are thermally baked before transfer to the microscope. However, the low barriers associated with the idealised surface are unlikely to be large enough to resist the highly energetic electron beam that is used in STEM experiments, as explained in Section 1.1.1, which means the high stability of single gold atoms on BDD observed by Hussein *et al.* [22] is most likely due to surface defects and dopants that were not seen within the STEM images.

4.3.5   Growth of Au_n Nanoclusters, $2 \leq n \leq 4$

Having studied and identified some stabilisation mechanisms for single gold atoms on diamond surfaces, the QM/MM methodology can now be used to study the growth of gold NCs on diamond. Out of all investigated defects and dopants, the SCOV-defective surface was shown to both thermodynamically and kinetically stabilise single gold atoms the most, and was thus used as the support for gold NCs. It should be noted though that the surface would comprise a variety of defects and dopants in reality, but the SCOV-defective surface should still provide a suitable model to study gold nucleation.

The growth of 2-, 3- and 4-atom gold NCs was studied using PBE^{+TS}/REBO, which was shown to perform well in describing the adsorption energetics of gold on SCOV-defective surfaces, with various properties of interest investigated to analyse the interaction between the NC adsorbate and the diamond surface, as well as to elucidate interactions within the NC itself. Figure 4.13 shows graphs detailing three key properties of interest that were used to characterise the nature of NC particle growth: gold-gold distances within NCs, the adsorption energies of NCsand the cohesion energies of NCs.
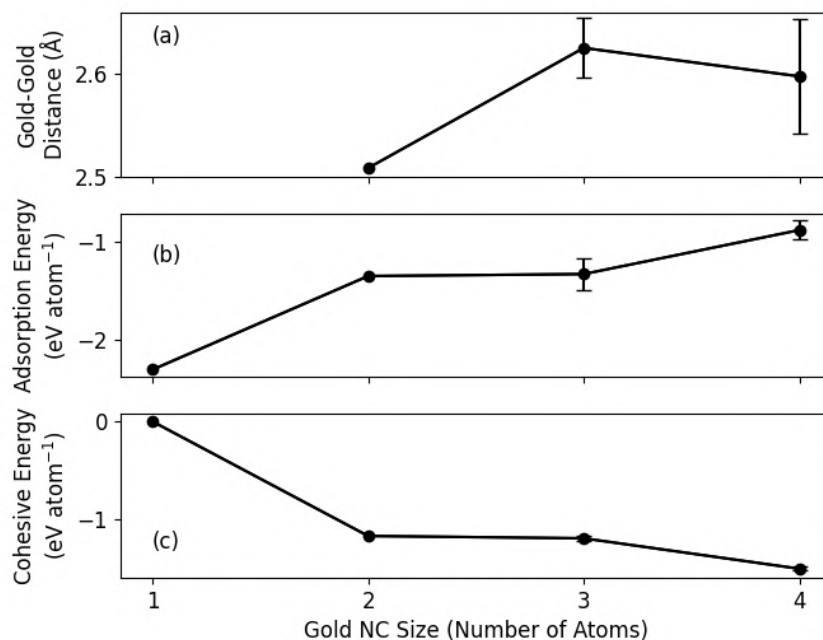
Figure 4.13. Graphs showing how certain properties of interest within gold NCs supported by a SCOV-defective oxygen-terminated diamond (110) surface vary with NC size after PBE^{+TS}/REBO optimisations. Shown are (a) the mean average gold-gold distances (b) mean average adsorption energies and (c) mean average cohesion energies, all provided as a function of gold NC size. Error bars for each property of interest represent one standard deviation of uncertainty.

As can be seen in Figure 4.13(a), the mean average gold-gold distance generally increases as the NC size increases. Similarly, the adsorption energy of the NC becomes less negative as the NC size increases, as shown in Figure 4.13(b). However, the adsorption energy does not include any deformation energy of the gold NC from its most stable structure in the gas phase. For this reason, the cohesion energy of the NCs were also calculated using Equation (4.2) and shown in Figure 4.13(c). Unlike the trend with adsorption energies, the cohesion energy was found to become more negative as the number of gold atoms within the NC increases. This shows that as the size of the supported NC increases, the resemblance to its optimal isolated structure also increases. This trend has also been observed for other supports such as magnesium oxide (001), graphite (0001), and cerium(IV) oxide (111) surfaces [407], and suggests that the interaction between the gold atoms increases as NC size increases, while the interaction strength between the NC and the surface substrate decreases. This hypothesis can be confirmed by evaluating the ratio of the cohesion and adsorption energies, E_{coh}/E_{ads}, which can be used to evaluate which interactions

111

dominate the NC structure. Table 4.5 details $E_{\mathrm{coh}}/E_{\mathrm{ads}}$ ratios for varying NC sizes, and shows that this ratio increases as the NC size increases. For an Au_4 NC, the $E_{\mathrm{coh}}/E_{\mathrm{ads}}$ ratio was calculated to be 1.72, which is fairly similar to values reported for Au_4 NCs supported on graphite and magnesium oxide surfaces by Engel *et al.* [407]. This is a relatively low number, and indicates that, on the SCOV-defective surface, gold NCs have a preference to interact with the surface rather than their structure be dominated by gold-gold interactions within the cluster.

Gold NC Size (Number of Atoms)	$E_{\mathrm{coh}}/E_{\mathrm{ads}}$ Ratio
1	0
2	0.87
3	0.90
4	1.72

Table 4.5. Table showing the ratio of the mean average cohesion and adsorption energies, $E_{\mathrm{coh}}/E_{\mathrm{ads}}$, for various gold NC sizes.

While generic trends have been identified, it is also prudent to study NCs of each atom count individually. The following subsections will delve into how NCs of a particular atom count were generated, as well as compare their different shapes, properties and interactions. It should be noted that for all studied Au_n NCs, the structures found and presented below are not an exclusive list of all possible dimer, trimer and tetramer structures respectively. To properly explore the space of all possible NC structure, for a given atom count, global optimisation algorithms such as basin-hopping [98; 99] would be required, though no current infrastructure exists to conduct this with the Py-ChemShell software. However, the following subsections should provide an indication into which NC geometries are more energetically favourable and stable on SCOV-defective systems.

Growth of Gold Dimers

As shown in Figure 4.14, five initial locations for the second gold atom were identified, each with displacement vectors of: $(0, 0, 1.5)$, $(1.5, 0, 1.5)$, $(-1.5, 0, 1.5)$, $(0, 1.5, 1.5)$ and $(0, -1.5, 1.5)$ (all numbers in angstroms) with respect to the location of the first gold atom, where the abscissa, ordinate and applicate refer to relative distances along the [001], [1$\bar{1}$0] and [110] directions respectively. As the first gold atom sits below the plane of carbonyl oxygen atoms in the SCOV-defective surface, as shown

in Table 4.1, the applicate was set to 1.5 Å in all displacement vectors to ensure the second gold atom would be placed above the carbonyl oxygen atoms rather than occupy the same space as a carbonyl oxygen atom.
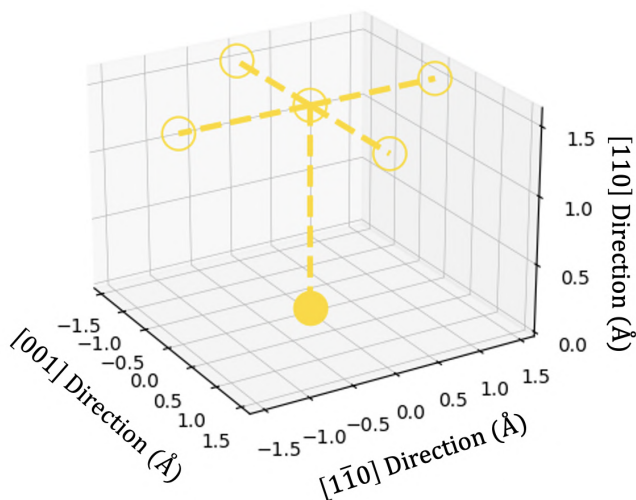
Figure 4.14. Plot showing the five initial locations for the second gold atom, shown as unfilled gold circles, with respect to the position of the first gold atom, shown as a filled gold circle, on a SCOV-defective oxygen-terminated diamond (110) surface.

Despite five different initial locations for the second gold atom, all systems optimised to the same structure, which is visualised in Figure 4.15. As can be seen, the second gold atom preferentially binds to the first gold atom rather than the diamond surface. While gold dimers have been observed to preferentially adopt a perpendicular arrangement with respect to the surface upon addition of a second gold atom, where the gold second atom sits on top of the first gold atom [196; 407], in this case, the dimer is preferentially tilted towards the surface with an Au–Au–O angle of 162.8° and a gold-gold distance of 2.51 Å. Similar gold-gold distances within dimers have been observed on different substrates [407], while a similar Au–Au–O angle has been reported for dimers atop other oxygen-terminated substrates such as cerium(IV) oxide (111) surfaces (163.4°) [192; 407]. The addition of the second gold atom does not have a significant effect on the first gold atom, which still sits in the SCOV defect site. The bond distance between the first gold atom and the former-ether oxygen atom reduces from 2.09 Å to 2.01 Å.
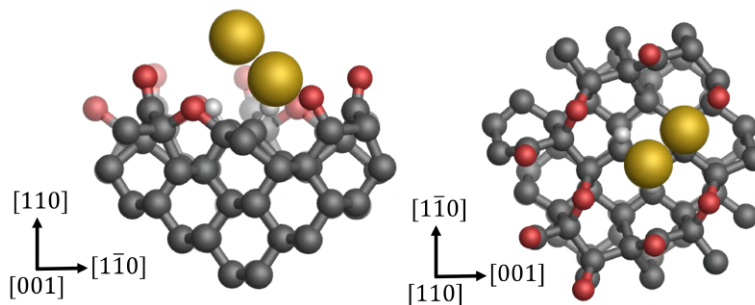
Figure 4.15. Orthographic ball-and-stick visualisations of a gold dimer adsorbed onto a SCOV-defective oxygen-terminated diamond (110) surface after a PBE^{+TS}/REBO optimisation. Visualisations of the QM region are shown from the [001] and [110] directions, and surface axes are also shown, with the saturating hydrogen link-atoms at the QM region boundary excluded for clarity. Carbon, oxygen, hydrogen and gold atoms are shown in grey, red, white and gold respectively.

The cohesion energy of this dimer was evaluated to be -1.17 eV, which is very similar to experimental (-1.16 eV [452]) and computational (-1.15 eV [407] and -1.16 eV [453]) values in literature. The adsorption energy of the gold dimer was calculated to be -1.35 eV atom^{-1}, which indicates a fairly strong interaction between the dimer and the surface. Despite this strong interaction, STEM measurements reported by Hussein *et al.* observed very few gold dimers on BDD surfaces [22]. To understand why, the adsorption and kinetic stabilities of the second gold atom were investigated. Taking the structure shown in Figure 4.5(b), which is the complex of a single gold atom supported by a SCOV-defective surface, to be $E_{\text{substrate}}$ in Equation (4.1), the adsorption energy of the second gold atom was calculated to be -2.73 eV, indicating a strong interaction between the second gold atom and the substrate.

Figure 4.16 shows the kinetic barriers of the second gold atom along the [001] and [1$\bar{1}$0] directions with respect to the initial adsorption site. Along the [001] direction, the energy barriers for the second atom were calculated to be 2.07 eV and 1.36 eV along the negative and positive displacements respectively. The energy barriers along the [1$\bar{1}$0] direction were calculated to be 2.97 eV and 0.56 eV along the negative and positive displacements respectively. The barrier is lower along the positive displacement due to the SCOV defect, which results in there being no oxygen atom within the neighbourhood of the defect site that the second gold atom needs to move above, as shown in Figure 4.16(c).
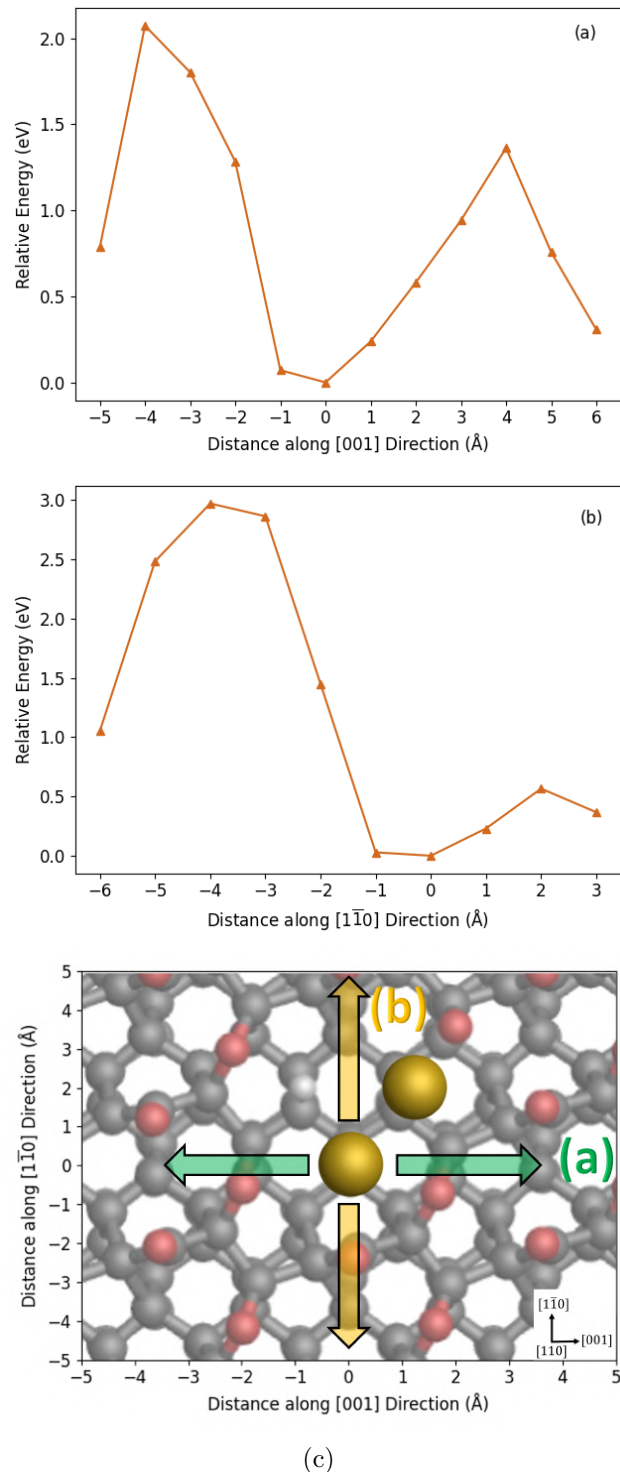
Figure 4.16. Relative energies, calculated using PBE^{+TS}/REBO, of moving the second gold atom within a gold dimer adsorbed onto a SCOV-defective oxygen-terminated diamond (110) surface substrate from its initial adsorption site, which is placed at the origin, along the (a) [001] and (b) [1$\bar{1}$0] directions. The paths along these directions are visualised in (c).

The strong adsorption energy of the second gold atom, and its high barriers, indicate that gold dimers should be stable on BDD surfaces. However, the lack of dimers observed by Hussein *et al.* [22] could be due to factors that were neglected within this chapter. Firstly, while it is clear that the high stability of single atoms observed by Hussein *et al.* [22] is mostly likely due to defects within the BDD surface, there is no experimental evidence that suggests this defect is a SCOV. There are many other defects that could exist, such as a saturated ether oxygen vacancy or even vacancies within the carbon layers of the surface, that could stabilise single atoms and result in different energetics for the second gold atom.

Furthermore, Hussein *et al.* showed that the effect of the highly energetic electron beam on atomic movement was negligible under the conditions they employed for STEM imaging [22]. However, the addition of an electrochemical potential was found to induce atomic movement across the BDD surface [22]. Experimental parameters such as the electrochemical potential, solvation and ions were not included within the QM/MM model setup, and can clearly have an effect on the stability of surface-adsorbed atoms. The potential-induced movement of atoms across the surface was observed to be an important pathway for the formation of BDD-adsorbed NCs [22]. The high stability of the second gold atom observed here could suggest that the formation of dimers is the rate-determining step for nucleation, though this would need to be investigated further to confirm.

Growth of Gold Trimers

Similar to the construction of the gold dimer, five initial locations of the gold atom were identified, each with displacement vectors of: $(0, 0, 1.5)$, $(1.5, 0, 0)$, $(-1.5, 0, 0)$, $(0, 1.5, 0)$ and $(0, -1.5, 0)$ (all numbers in angstroms) with respect to the location of second gold atom, which are shown in Figure 4.17. The applicate was set to $0\,\text{Å}$ in the last four displacement vectors as the second gold atom was already above the plane of carbonyl oxygen atoms. Despite five different initial locations for the third gold atom, four of the systems optimised to form a triangular structure perpendicular to the surface, as shown in Figure 4.18(a), while one optimised to form a triangular structure that bridges over the surface along the [001] direction, as shown in Figure 4.18(b).
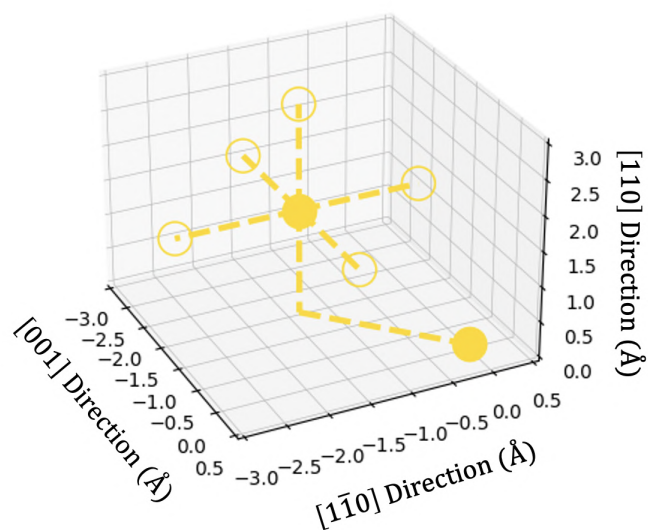
Figure 4.17. Plot showing the five initial locations for the third gold atom, shown as unfilled gold circles, with respect to the position of the first and second gold atoms, shown as filled gold circles with the former at the origin, on a SCOV-defective oxygen-terminated diamond (110) surface.
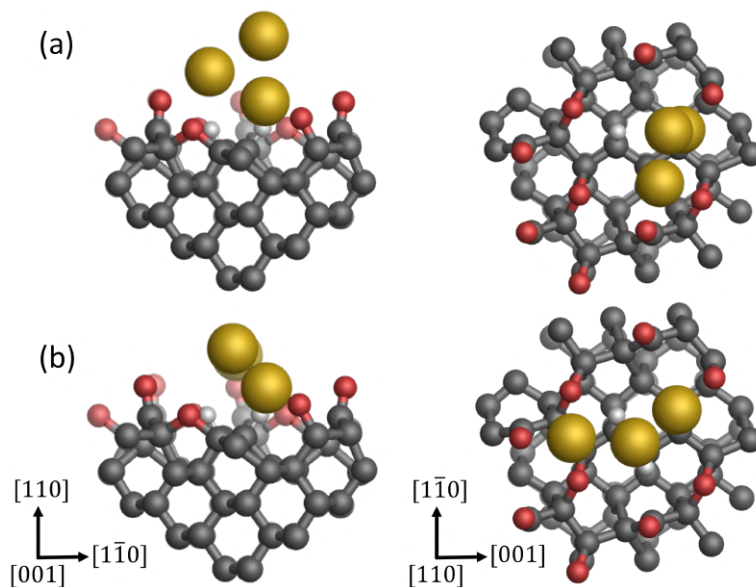
Figure 4.18. Orthographic ball-and-stick visualisations of two gold trimers adsorbed onto a SCOV-defective oxygen-terminated diamond (110) surface obtained after PBE^{+TS}/REBO optimisations. Visualisations of the QM region are shown from the [001] and [110] directions, and surface axes are also shown, with the saturating hydrogen link-atoms at the QM region boundary excluded for clarity. Carbon, oxygen, hydrogen and gold atoms are shown in grey, red, white and gold respectively.

117

In the trimer shown in Figure 4.18(a), the third gold atom adsorbs above the first gold, which forces the second gold atom closer toward the surface and slightly pulls the first gold atom away from the surface. This can be seen by the Au–O bond length slightly elongating from 2.01 Å in the dimer to 2.06 Å. The triangular structure is geometrically near-equilateral, with Au–Au–Au angles of 60.4°, 60.6° and 59.0°, while the mean average Au–Au distance was 2.64 Å. The cohesion energy for the triangular NC was calculated to be the same as for the dimer: -1.17 eV, while the adsorption energy was calculated to be -1.16 eV atom^{-1}, which is a 0.19 eV atom^{-1} weaker than the interaction between the dimer and the surface.

In contrast, the trimer is shown in Figure 4.18(b) is more planar than the other trimer, and the third gold atom adsorbs behind the second gold atom along the [001] direction to form a structure that bridges over the surface. This triangular structure is geometrically obtuse and near-isosceles, with Au–Au–Au angles of 123.0°, 28.8° and 28.3°, and Au–Au distances of length 2.62 Å and 2.58 Å between the first and second gold atoms, and the second and third gold atoms respectively. The Au–Au–O angle was found to be the same as in the dimer (162.8°), though the Au–O length increased to 2.08 Å, similar to the single atom value. This structure was found to be adsorb more strongly to the surface than the near-equilateral NC, with an adsorption energy of -1.49 eV atom^{-1}, and was also found to have a slightly larger cohesion energy of -1.22 eV atom^{-1}.

Growth of Gold Tetramers

Both trimers in Figure 4.18 were used as the substrate for the fourth gold atom. With the first trimer shown in Figure 4.18(a), eight initial sites were identified for the fourth atom, with directional vectors of $(-1.5, 0, 0)$, $(1.5, 0, 0)$, $(0, -1.5, 0)$ and $(0, 0, 1.5)$ with respect to the location of the third gold atom, and $(-1.5, 0, 0)$, $(1.5, 0, 0)$, $(0, 1.5, 0)$ and $(0, 0, 1.5)$ with respect to the second gold atom (all numbers in angstroms). With the second trimer shown in Figure 4.18, seven initial sites for the fourth atom were identified, with directional vectors of $(-1.5, 0, 0)$, $(0, 1.5, 0)$, $(0, -1.5, 0)$ and $(0, 0, 1.5)$ with respect to the third gold atom, and $(0, 1.5, 0)$, $(0, -1.5, 0)$ and $(0, 0, 1.5)$ with respect to the second gold atom (all numbers in angstroms). All initial positions can be seen in Figure 4.19.
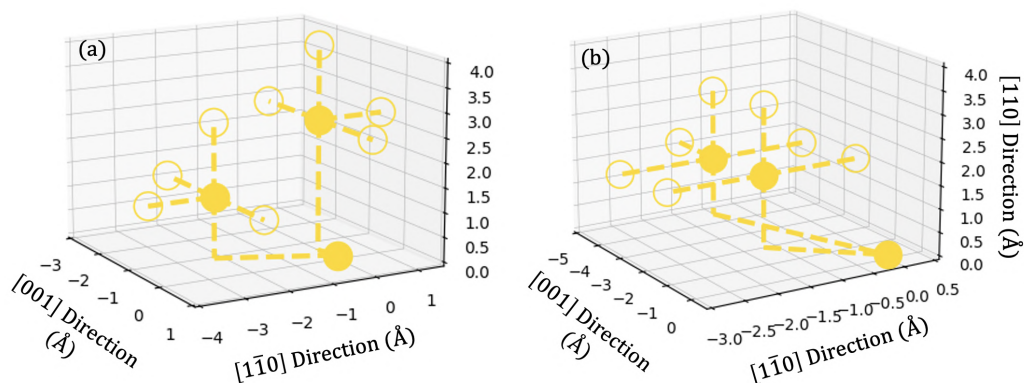
Figure 4.19. Plot showing the initial locations for the fourth gold atom, shown as unfilled gold circles, with respect to the position of the first, second and third gold atoms, shown as filled gold circles with the former at the origin, in the trimers shown in (a) Figure 4.18(a) and (b) Figure 4.18(b).

The resulting Au_4 NCs are shown in Figure 4.20, with (a)-(d), originating from the trimer shown in Figure 4.18(a), and (e)-(g) originating from the trimer shown in Figure 4.18. Table 4.6 details the adsorption and cohesion energies of the tetramers visualised in Figure 4.20. As Figure 4.20 shows, there are significantly more stable structures that arise after the addition of a fourth gold atom. This is because as the NC size increases, the number of possible sites for an additional atom increases in a superlinear fashion.
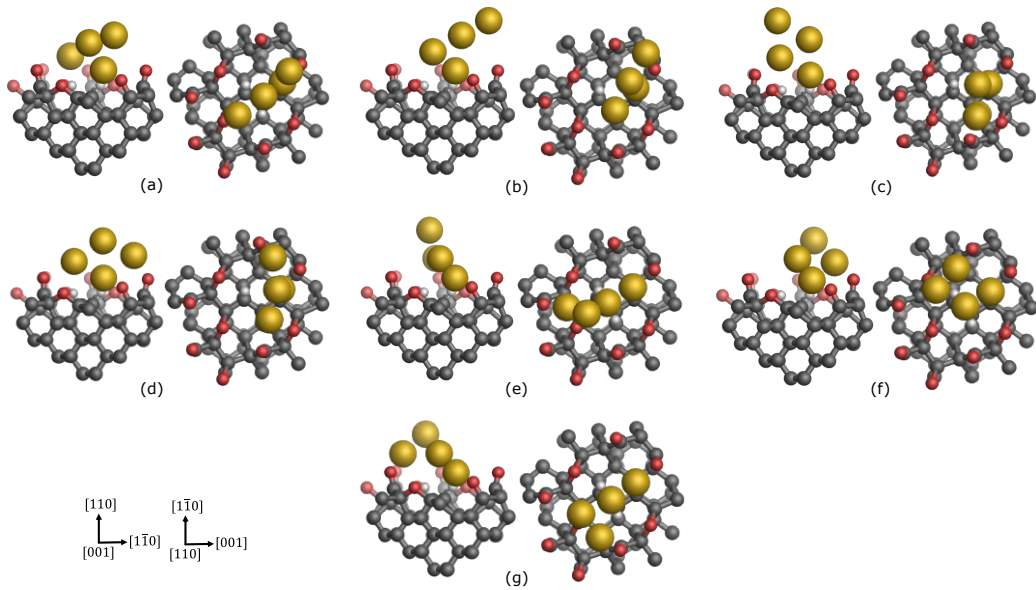
Figure 4.20. Orthographic ball-and-stick visualisations of seven gold tetramers adsorbed onto a SCOV-defective oxygen-terminated diamond (110) surface obtained after PBE^{+TS}/REBO optimisations. Visualisations of the QM region are shown from the [001] and [110] directions, and surface axes are also shown, with the saturating hydrogen link-atoms at the QM region boundary excluded for clarity. Carbon, oxygen, hydrogen and gold atoms are shown in grey, red, white and gold respectively.

Structure	Adsorption Energy (eV atom^{-1})	Cohesion Energy (eV atom^{-1})
(a)	-0.97	-1.50
(b)	-0.99	-1.52
(c)	-0.77	-1.52
(d)	-0.99	-1.48
(e)	-0.85	-1.52
(f)	-0.79	-1.48
(g)	-0.77	-1.53

Table 4.6. Adsorption and cohesion energies of the supported gold tetramers visualised in Figure 4.20.

The first tetramer, as shown in Figure 4.20(a), is a Y-shaped structure that can almost be considered an amalgamation of the two trimers from Figure 4.18. The fourth gold atom adsorbs 2.51 Å behind the third gold atom within the ini-

tial near-equilateral triangle to form a 'tail' that bridges over the surface along the [001] direction. As shown in Table 4.6, this particular structure was found to have an adsorption energy of $-0.97\,\mathrm{eV\,atom^{-1}}$ on the SCOV-defective surface, which is $0.19\,\mathrm{eV\,atom^{-1}}$ weaker than its parent trimer, which itself was $0.19\,\mathrm{eV\,atom^{-1}}$ weaker than its parent dimer. The cohesion energy of this tetramer was calculated to be $-1.50\,\mathrm{eV}$, which is $0.33\,\mathrm{eV}$ larger than its parent trimer. The distances between the first and second atoms, and the second and third atoms were found to slightly elongate to $2.66\,\mathrm{\AA}$ (from $2.62\,\mathrm{\AA}$) and $2.59\,\mathrm{\AA}$ (from $2.58\,\mathrm{\AA}$) respectively. The second structure, as can be seen in Figure 4.20(b), is similar in structure to the first, except the fourth atom preferentially adsorbs to the third gold atom, which is the apex of the near-equilateral triangle from the [001] viewpoint, rather than closer to the surface. As can be seen in Table 4.6, this structure not only adsorbs stronger to the SCOV-defective surface with an adsorption energy $0.02\,\mathrm{eV\,atom^{-1}}$ more negative than structure (a), but also has a cohesion energy $0.02\,\mathrm{eV\,atom^{-1}}$ more negative than (a). This indicates that this structure is more stable on the SCOV-defective surface than structure (a).

The third structure, as shown in Figure 4.20(c), has a slightly different shape than (a) and (b). The fourth gold atom adsorbs above the second and third gold atoms, and away from the surface, resulting in the NC structure being rhomdoidal and an extension of its parent trimer along the [110] direction. In this structure, the distance between the first and second gold atoms, which are closest to the surface, elongates to $2.70\,\mathrm{\AA}$, and this is reflected by its relatively weaker adsorption energy of $-0.77\,\mathrm{eV\,atom^{-1}}$ with respect to tetramers (a) and (b). Such a trend in adsorption energies of gold tetramers has been observed in literature, where Y-shaped structure generally have stronger adsorption energies on substrates compared to rhomboidal or parallelogram-like structures [187; 407; 454]. The fourth structure, as visualised in Figure 4.20(d), is also rhomboidal, with the fourth gold atom adsorbing on the other side of the third atom as compared to (c) and closer to the surface. The closer proximity of the fourth atom to the surface means there are more gold atoms at the NC-surface interface, which results in an adsorption energy $0.22\,\mathrm{eV\,atom^{-1}}$ more negative than structure (c).

Three tetramers were found to form from the more stable parent trimer, which are visualised in Figure 4.20(e)-(g). As can be seen in Table 4.6, these tetramers generally have weaker adsorption energies than structures (a)-(d). In structure (e), the fourth gold atom adsorbs above the second and third gold atoms

to form a near-equilateral triangular structure with Au–Au–Au angles of 58.6°, 66.6° and 54.8°. The adsorption energy of this structure is $-0.85\,\mathrm{eV\,atom^{-1}}$, which is $0.08\,\mathrm{eV\,atom^{-1}}$ more negative than structure (c), but not as strongly adsorbed as the Y-shaped tetramers. Structure (f) is another rhomboidal structure but is preferentially tilted towards the surface rather than perpendicular to it like in (c). As can be seen in Table 4.6, the adsorption energies between (c) and (g) are quite similar, though the cohesion energy of (f) is the same as that of (d). Structure (g) has similar adsorption and cohesion energies to structure (c) despite having a different geometry.

4.4 Conclusions

This chapter has shown that embedded cluster models via a QM/MM framework can be used to study the adsorption and diffusion kinetics of single gold atoms on oxygen-terminated diamond (110) surfaces. After deciding on an appropriate QM region size, the framework was used to analyse the effects of local surface defects and dopants on adsorption energies and barriers. For the idealised, fully-oxygenated surface, the gold atom was found to weakly adsorb onto the surface. The kinetic barriers associated with this surface along both the [001] and [1$\bar{1}$0] directions were also found to be very low, and while they are high enough to counter the translational kinetic energy that arises from thermal baking, the idealised surface is unlikely to stabilise single gold atoms when studied under experimental STEM conditions. The introduction of defects and boron dopants into the surface substrate, however, was found to significantly increase the adsorption energy and barriers of the single gold atom. In the former case, the introduction of a SCOV into the surface was found to result in very strong adsorption between the surface and the adatom, which was due to the formation of a bond between the gold atom and a surface ether oxygen atom. This means that the high stability of single gold atoms observed on polycrystalline BDD surfaces observed by Hussein *et al.* is most likely due to surface defects and dopants that were not seen in their STEM images or accounted for within their DFT calculations.

After the identification of stabilisation mechanisms for the single gold atom, the validity of the trends observed using PBE$^{+\mathrm{TS}}$/REBO method was evaluated by benchmarking the method against other forcefields, dispersion correction schemes and DFAs. The REBO forcefield was shown to be an appropriate embedding environment for the QM region, while very little dependency was found on the flavour

of dispersion correction, though a dispersion correction was found to be necessary to accurately capture the adsorption energetics of the single gold adatom. The PBE GGA was also found to generally perform very well with respect to other GGAs, as well as higher-rung MGGAs and HGGAs for calculating adsorption energies and barriers.

Finally, the PBE^{+TS} method was subsequently used to study, albeit somewhat inefficiently, the formation of small gold NCs on SCOV-defective surfaces. Overall, the mean average gold-gold distance was found to increase as the size of the NC increased. Furthermore, the mean average adsorption energy of all structures was found to become less negative as the NC size increased, while the mean average cohesion energy was found to become more negative. This indicates that the NC structure gets closer to its gas-phase optimum as the NC size increases, while the interaction between the NC adsorbate and the surface weakens. By comparing the cohesion energy:adsorption energy ratios, the growth of gold NCs was found to be preferentially determined by their interaction with the SCOV-defective surface rather than by gold-gold interactions within the clusters. For dimers, the second gold atom was found to preferentially bind to the first gold atom irrespective of its initial placement, but the structure was tilted towards the surface rather than orthogonal to it. The second gold atom was found to strongly adsorb to the substrate and also possess large barriers. This indicates that the lack of gold dimers observed by Hussein *et al.* [22] is likely due to factors not considered herein, such as the electrochemical potential. For trimers, two stable structures were found depending on the initial site of the third atom: a near-equilateral triangular arrangement perpendicular to the surface, and a near-isosceles triangular arrangement that bridged over the surface. Finally, for tetramers, seven structures were found to form, mostly due to the larger variety of possible initial sites for the fourth atom. Of these tetramers, Y-shaped structures were found to be more stable than rhomboidal ones.

While some aspects such as the electrochemical potential were neglected within the QM/MM models, the framework presented in this chapter can be used to explore the effects of defects and dopants, which have been observed to increase the stability of adsorbates, as well as study the initial stages of metal NC formation on defective and doped surfaces.

Chapter 5

Structures of Gold Nanoclusters on Diamond

The following chapter presents work published in the paper referenced below. The au_full parameter set used herein was provided by Francis He (Carnegie Mellon University), and the validation dataset used for this parameter set, as visualised in Figure 5.2, was created with Dr. Adam McSloy (University of Warwick). The training of MLIPs, as detailed in Section 5.2.2, was conducted by Dr. Julia Westermayr (University of Warwick).

J. Westermayr, <u>S. Chaudhuri</u>, A. Jeindl, O. T. Hofmann, R. J. Maurer (2022) 'Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic–inorganic interfaces' *Digital Discovery* **1**: 463– 475 [455]

5.1 Introduction

The size and geometry of surface-deposited NCs are known to play an integral role in determining their catalytic ability [456]. The optimisation of such structures is particularly challenging due to the inherent structural complexity and the large number of degrees of freedom that exist, such as the geometric shape and surface coverage [68]. As discussed in Chapter 4, the manual method used to build NCs is not efficient, and while it can be used to study NCs comprising only a few atoms, the number of possible sites for an additional gold atom increases at a superlinear rate as the NC size increases. Furthermore, while *ab initio* methods like DFT offer a high degree of accuracy, they require a high computational cost to run cal-

culations [408] and it becomes infeasible to use them as a calculator in structure exploration methods. Global structure searches using DFT for gas-phase gold NCs have typically been limited to small structures with diameters below 1.2 nm [457] and containing 9–13 [458], 12–14 [459], 15–19 [460] or 40 [461] atoms. Thus, there is a great need for efficient methods that can be used to conduct accurate structure searches to explore the large configuration space of possible NC structures.

An additional point to note when exploring the configurational space of possible NC structures is that long-range effects such as vdW forces will have a large impact on the final geometry as well as the properties exhibited by the NC. Furthermore, for large systems comprising numerous atoms, beyond-pairwise interactions will have a greater impact on the final NC shape and size. Therefore, any calculator that is used to conduct structure searches for gold NCs on diamond must be able to correctly account for beyond-pairwise, long-range effects. Methods such as DFTB2 [78], as outlined in Section 2.38, have been shown to provide a good compromise between computational cost and accuracy, and can also be coupled with numerous dispersion corrections such as DFT-D3 [282], TS [462; 463] and MBD [289] schemes via the DFTB+ [79] software package. DFTB has been used to explore the structures of gold NCs gold [464; 465], though these studies have been limited to isolated NCs not supported by a substrate. This is partly due to the existence of few reliable parameterisations that can accurately model hybrid organic-inorganic interfaces [86; 466; 467] as parameterisations are typically developed for a particular subset of elements for a specific purpose, which means they possess a low transferability.

MLIPs, as discussed in Section 2.5.2, are another methodology that can not only offer high computational efficiency, but can also be used to perform calculations at an *ab initio*-level of accuracy, if the appropriate training data is supplied, making it possible to rapidly search chemical space for optimal structures [468; 469; 470]. In particular, MLIPs have been shown to be capable of advancing structure searches [471; 472; 473] and geometry optimisations [474; 475] of highly complex and large-scale systems comprising many thousands of atoms [476]. MLIPs have been used to explore the structures of metallic alloys [477] and various metals [478], including gold [457; 479; 480; 481; 482], but these are also mostly limited to isolated structures, and very few investigate the structures of supported NCs [483]. However, most established MLIP approaches learn short-range interactions between atoms via a radial cut-off term, within which the atomic interactions are captured. This can

lead to challenges when attempting to capture long-range interactions [318]. Recent attempts to account for long-range interactions have explicitly treated them as separate additive contributions to the MLIP [317; 318; 484; 485], such as the third and higher generation NN potentials of Behler and co-workers [486; 487], where a charge-equilibration scheme was introduced. Earlier work by Behler and co-workers [485; 488] has also shown that the simulation of liquid water can be facilitated with NNs trained on energies and atomic charges, where the latter was used to correct for electrostatic interactions. This scheme was later complemented with long-range dispersion interactions based on the DFT-D3 [282] correction scheme [488]. Atomic charges were further used in TensorMol-0.1 [484] to augment the total energy with Coulomb and vdW corrections. A similar approach was applied by Unke and Meuwly in PhysNet, where the total energy was corrected with additive terms that include electrostatic corrections obtained from partial atomic charges and a DFT-D3 [282] dispersion correction term [318]. Recently, this description was extended in SpookyNet, where the total energy was corrected using empirical terms for the nuclear repulsion based on an analytical short-range term, a term for electrostatics and a term for dispersion interactions [317]. The aforementioned approaches have been shown to accurately describe spectroscopic signatures [484], small clusters on surfaces [487], water dimers [489] and clusters [488], crystals [489], and phase diagrams [490]. However, none of these approaches account for density-dependent dispersion schemes such as the TS [262] method, nor do they explicitly account for beyond-pairwise dispersion interactions that are included within the MBD schemes [285; 286], as explained in Section 2.2.

This chapter will explain the development of an ML method trained on sparse dispersion-corrected DFT data that can be used to facilitate fast and accurate structure searches. First, the viability of dispersion-corrected DFTB2 methods as an alternative to dispersion-corrected DFT will be investigated. This will be done by benchmarking elemental interactions with gold by comparing the adsorption energetics and structural differences of small organic molecules on gold NCs and gold NCs on diamond (110) surfaces. After this, the development of a deep learning approach that combines SchNet-based [315; 333] MLIPs with an established long-range dispersion correction method will be presented. The accuracy of the MLIPs will be benchmarked against dispersion-corrected DFT for a variety of supported gold NCs, and its superior performance with respect to DFTB2 as well as its significantly cheaper computational cost with respect to both DFTB2 and DFT will be demonstrated. Finally, the MLIPs will be used to analyse the adsorption of large

gold NCs on diamond that would otherwise be computationally intractable with DFT.

5.2 Methods

Similar to Chapter 4, the notation '$\chi^{+\psi}$' is used henceforth to denote specific QM methods, where χ is the QM method, which is either a DFA, a DFTB2 parameter set or an MLIP, and ψ is the dispersion correction.

5.2.1 QM Calculation Settings

DFT Calculations

DFT [65; 66] calculations were performed using the FHI-aims [70] software package. The PBE [229] GGA, which was shown in Chapter 4 to perform well for gold adsorbates on diamond, was used in all calculations along with the MBD@rsSCS [287] scheme to account for long-range dispersion effects.

DFTB2 Calculation Settings

DFTB2 [78] calculations were performed using the DFTB+ [79] software package. Calculations were performed using either the publicly-available auorg [86] parameterisation or the auorg-derived au_full parameterisation that was provided by Francis He. The latter parameterisation was generated using an ML-based approach similar to the one outlined by Li *et al.* [468]. The au_full parameterisation retained the same electronic components as auorg but its repulsive component, as detailed in Equation (2.39), was instead fitted to energies and forces from DFT calculations, that were conducted on organic molecules on gold surfaces, using a natural cubic spline that was constrained to have a strictly positive second derivative. For au_full, the repulsive potential was treated as a function of the sum of interatomic distances and was therefore defined for every pairwise combination of elements, and the radial cut-off intervals for all pairwise interactions were set to (0 Å, 10 Å). It should be noted that at the time of use, the au_full parameter set was still in development and not finalised. The MBD@rsSCS scheme [285; 286; 287] was used to account for long-range effects within DFTB2 calculations, though some calculations were performed using the DFT-D3 [282] correction scheme, which are clearly stated within the text.

5.2.2 Training of MLIPs

Training Dataset

DFT [65; 66] calculations were conducted using the FHI-aims [70] software package and the PBE$^{+\text{MBD@rsSCS}}$ method [229; 287]. Using the Atomic Simulation Environment [339], gold NCs of various sizes were placed onto the centre of an unterminated diamond (110) surface and optimised, with all carbon atoms being fully frozen during optimisations . 62 geometry optimisations were used as the starting point for the training dataset, distributed as 5, 4, 8, 8, 9, 10 and 18 geometry relaxations on gold NCs of size 15, 20, 30, 35, 40, 45 and 50 atoms respectively. Training data points were collated using every relaxation step of optimisations, which therefore comprised both optimised and not fully-optimised structures. This resulted in 5368 data points, of which 4500 were used to train four MLIPs on vdW-free energies (E) and forces (F), 500 were used for validation and 368 were used for testing. Furthermore, four separate geometry optimisations of Au$_{20}$, Au$_{30}$, Au$_{40}$ and Au$_{50}$ NCs were kept as a hold-out set to test the performance of the models.

Input and output files for the structure optimisations conducted to generate the training data have been uploaded as a dataset to the NOMAD electronic structure data repository [378] and are freely available under https://doi.org/10.17172/NOMAD/2021.10.28-1.

SchNet-Based MLIPs

The SchNet [315; 316] NN was used by Dr. Julia Westermayr to create MLIPs by learning the representation of atomic environments along with its relation to the targeted output.

Two types of SchNet-based MLIPs were generated, which were combined to create an overall MBD@rsSCS-corrected SchNet calculator, SchNet$^{+\text{MBD@rsSCS}}$. The first type, SchNet$_{\text{initial}}(E, F)$, accounted for short-ranged interactions and was trained on E and F for each atom from PBE$^{+\text{MBD@rsSCS}}$ optimisations by subtracting the long-range vdW contributions to the total energies and forces for each atom. Energies and forces for the whole system were obtained by summing up the respective atomic contributions, and F was treated as the derivative of E with respect to atomic positions $\mathbf{r}$. Four SchNet$_{\text{initial}}(E, F)$ models were trained, each differing in the choice of trade-off value, t, that was used within the loss function to weight E and F during training, with $t = 0.03, 0.04, 0.04$ and 0.05 being used for each model,

while the weights within the NN were initialised randomly. The reason for multiple models was to make use of the query-by-committee approach [491], which takes the mean average of the predictions (in this case E) of $\tilde{n}$ ML models, and can be used to improve the accuracy and robustness of predictions by a factor $\sqrt{\tilde{n}}$.

Equation (5.1) expresses the squared loss function, $\ell^2(E, F)$, that was used to train the $\text{SchNet}_{\text{initial}}(E, F)$ and $\text{SchNet}_{\text{adapt1.}}(E, F)$ MLIPs for a system with N atoms. This type of loss function is very sensitive to any outliers as the difference between the predicted and actual values is squared, which therefore penalises any large errors. However, for the $\text{SchNet}_{\text{adapt2.}}(E, F)$ and $\text{SchNet}_{\text{adapt3.}}(E, F)$ MLIPs, a smooth $\ell^1(E, F)$ loss function, as expressed in Equation (5.2), was used instead. This was because the training datasets for these two MLIPs deliberately included data points which possessed larger forces and energies than most of the data points.

$$\ell^2(E, F) = t \left| E - \sum_{i=1}^{N} E_{\text{ML}}(i) \right|^2 + (1 - t) \left| F - \frac{\partial}{\partial \mathbf{r}} \left(\sum_{i=1}^{N} E_{\text{ML}}(i) \right) \right|^2 \tag{5.1}$$

$$\ell^1(E, F) = t \left| E - \sum_{i=1}^{N} E_{\text{ML}}(i) \right| + (1 - t) \left| F - \frac{\partial}{\partial \mathbf{r}} \left(\sum_{i=1}^{N} E_{\text{ML}}(i) \right) \right| \tag{5.2}$$

The second MLIP type, $\text{SchNet}(H)$, was trained on Hirshfeld volume ratios (H) on a per-atom basis, which were extracted from $\text{PBE}^{+\text{MBD@rsSCS}}$ calculations. For each atom, H was calculated as the ratio between the effective volume of the atom within a molecule and the volume of the free atom, as expressed in Equation (2.24). The Hirshfeld partitioning of the electronic density [284] was then used to rescale the atomic polarisabilities and vdW radii, which was subsequently used to evaluate a vdW correction based on the MBD@rsSCS [287] scheme, as explained in Section 2.2.2. It should be noted that this model could also be used to evaluate any vdW correction based on the Hirshfeld partitioning of the electronic density [284], such as the TS [262] schemewhich used H to rescale atomic polarisabilities and the C_6 coefficients, as explained in Section 2.2.1. Due to the high accuracy in the H predictions, only one model was trained and the query-by-committee approach was not required. Equation (5.3) expresses the squared loss function, $\ell^2(H)$, that was used to train the $\text{SchNet}(H)$ model for a system with N atoms.

$$\ell^2(H) = \sum_{i=1}^{N} |H(i) - H_{\text{ML}}(i)|^2 \tag{5.3}$$

Adaptive sampling [319] was used to extend the training dataset by running global structure searches using a basin-hopping algorithm, as implemented within the Atomic Simulation Environment [339], with the initial model. NCs of various sizes, including those featured in the training dataset (15, 20, 30, 35, 40, 45 and 50 atoms) and some new sizes (6, 25, 28, 44 and 66 atoms) were used as the starting geometry for each basin-hopping run. The variances of the model predictions after every geometry optimisation step of the basin-hopping run were computed, and geometries with the largest variances were further optimised using $PBE^{+MBD@rsSCS}$. The DFT optimisation steps were subsequently added to the training dataset to improve the performance of the MLIP. In total, three adaptive sampling runs were conducted. The collection of MLIPs after the first adaptive sampling run, $SchNet_{adapt1}$, which were obtained after conducting basin-hopping using the $SchNet_{initial}$ MLIPs, were trained on 7700 data points and validated using 800 data points. In addition, 243 single-point calculations of structures with the largest model errors where conducted to inform the model where not to go during optimisations. Using these single-point calculations and after conducting a second adaptive sampling run with the $SchNet_{adapt1}$ MLIPs, the $SchNet_{adapt2}$ collection of MLIPs was obtained, which resulted in 9757 data points in total, of which 8500 were used for training and 800 were once again used for validation. After the third and final adaptive sampling run with the $SchNet_{adapt2}$ MLIPs to obtain the collection of $SchNet_{adapt3}$ MLIPs, 15293 data points were procured, of which 12500 were used for training and 1500 were used for validation.

The installation instructions and the codes used to develop the SchNet-based MLIPs are freely available from the `maurergroup/SchNet-vdW` GitHub repository under https://github.com/maurergroup/SchNet-vdW.

5.2.3 Structure Generation and Analysis

Supported gold NC structures were constructed using the Atomic Simulation Environment [339] by placing NCs of various sizes onto the centre of an unterminated diamond (110) surface, with all carbon atoms being fully frozen during optimisations. An unterminated surface was used as the oxygenation state of the diamond (110) surface had not yet been characterised at the time of development. The structures for gold NCs themselves were manually created by adding atoms at random sites above the diamond (110) surface substrate using the Atomic Simulation Envi-

ronment [339], with the exception of Au$_{147}$ NCs. The icosahedral and cuboctahedral Au$_{147}$ NCs were constructed using the `ase.cluster` package provided within the Atomic Simulation Environment [339]. The number of shells to create the icosahedron was set to 4, while the number of layers cut at each vertex and the number of atoms on the square edges of the complete octahedron were set to 3 and 7 respectively to create the cuboctahedron. Cartesian coordinates for the amorphous Au$_{147}$ NC were taken from Tarrat *et al.* [465]. All three Au$_{147}$ NC isomers are visualised in Figure 5.1.

| (a) | (b) | (c) |
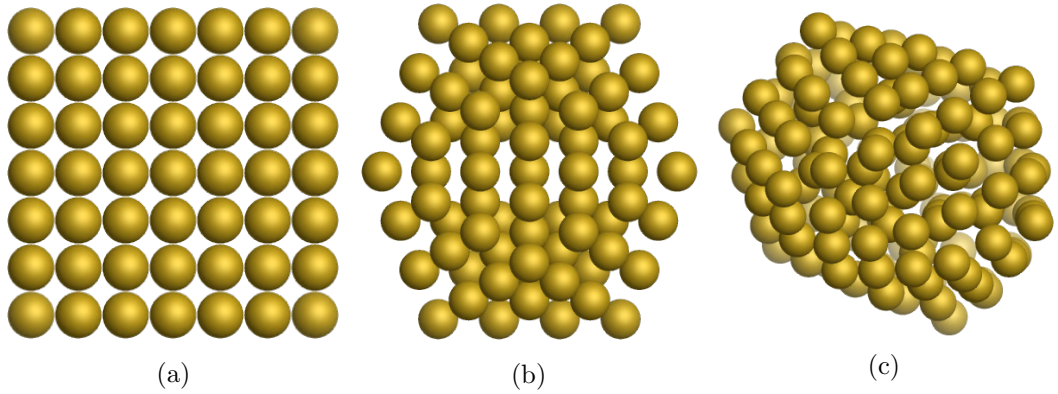
Figure 5.1. Orthographic visualisations of the (a) cuboctahedral, (b) icosahedral and (c) amorphous 147-atom gold NCs investigated. Cartesian coordinates for structure (c) were taken from Tarrat *et al.* [465].

To analyse the compactness of the Au$_{147}$ NCs, Equation (5.4) was used to calculate the sphericity, $\mathcal{S}$, of a NC, which is defined as ratio of the surface area of a sphere with the same volume as the given NC to the surface area of the NC [492]. $\mathcal{S} = 1$ for a perfect sphere and as a result of the isoperimetric inequality [493], $\mathcal{S} < 1$ for any other shape.

$$\mathcal{S} = \frac{\pi^{1/3}(6V)^{2/3}}{A} \tag{5.4}$$

where V and A are the volume and surface area of the NC respectively [492]. For an edge length $l \in \mathbb{R}$, V and A for regular icosahedra ($V = \frac{5}{12}(3 + \sqrt{5})l^3$, $A = 5\sqrt{3}l^2$) and cuboctahedra ($V = (6 + 2\sqrt{3})l^3$, $A = \frac{5}{3}\sqrt{2}l^3$) are trivially defined. However, for every irregular shape, such as the amorphous NCs and the NCs after geometry optimisations, V and A were calculated from the convex hull of the shape, which is defined as the smallest convex set that contains the shape.
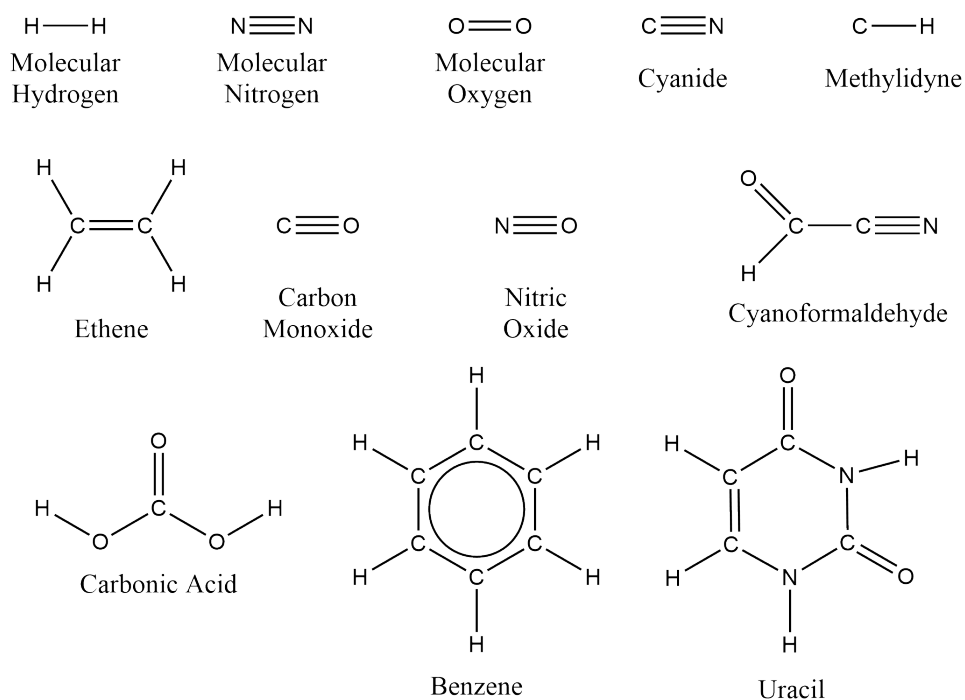
Figure 5.2. Atomic structures of the twelve organic molecular adsorbates used to validate the ML-generated au_full DFTB2 parameter set provided by Francis He. Structures were selected with Dr. Adam McSloy.

Figure 5.2 shows the twelve organic molecular adsorbates that were used as the validation set, which was selected with Dr. Adam McSloy, for the ML-generated au_full DFTB2 parameterisation provided by Francis He. Of these twelve molecules, all but three (molecular hydrogen, benzene, and uracil) were present within the original training set for au_full. These organic molecules were placed onto 10-, 18- and 34-atom gold clusters, as well as a periodic gold (111) surface. For heteronuclear diatomic molecules, A–B (where the letters denote their chemical symbols), both coaxial binding modes were investigated and were termed 'AB top' and 'AB bottom' respectively, where the noun refers to the position of B within the molecule with respect to the surface.

5.3 Results and Discussion

5.3.1 Viability of Dispersion-Corrected DFTB2

Having explained the need to account for beyond-pairwise interactions in large structures, two MBD@rsSCS-corrected DFTB2 parameter sets were benchmarked against MBD@rsSCS-corrected DFT, the first of them being the publicly-available auorg [86] parameterisation, which is the only (currently) publicly available DFTB2 parameterisation that accounts for interactions with gold [86]. However, this parameterisation was designed to describe optical excitations of thiolates on gold NCs [86], so another parameterisation was sought after that would, in theory, be more applicable to describe gold NCs on diamond. ML-based DFTB parameterisations focusing on the repulsive potential, as detailed in Equation (2.39), have been reported to outperform existing DFTB2 parameterisations [468; 494; 495; 496], while constrained splines have successfully been used to generate repulsive pair potentials with a high level of accuracy and transferability for a range of systems [497; 498]. For this reason, the performance of an ML-generated, spline-constrained reparameterisation of the auorg parameter set, au_full, provided by Francis He was investigated. The au_full parameter set was designed to describe interactions between organic molecules and gold surfaces, and should in principle outperform auorg. The accuracy of both DFTB2 parameterisations was determined by benchmarking them against DFT. This was done first for the adsorption of the organic molecules, shown in Figure 5.2, on various gold NC substrates. While these are different systems to gold NCs on diamond, they can still provide some understanding into how accurately the parameter sets describe elemental interactions with gold. Both DFTB2 parameterisations were subsequently used to optimise 4-atom gold NCs on diamond (110) surfaces and benchmarked against DFT.

Organic Molecules on Gold Nanoclusters

First, the structural differences after optimisations with auorg, au_full and PBE (all MBD@rsSCS-corrected) investigated. This was done by comparing the RMSDs of the two DFTB2 parameter sets with respect to PBE$^{+\text{MBD@rsSCS}}$. Figures 5.3(a) and (b) show the mean average RMSDs with respect to PBE$^{+\text{MBD@rsSCS}}$ of all the organic molecules, substrates and combined systems after auorg$^{+\text{MBD@rsSCS}}$ and au_full$^{+\text{MBD@rsSCS}}$ optimisations, respectively. Across all structures, the au_full parameterisation was found to result in a larger mean average RMSD value (0.489 Å) than auorg (0.295 Å), indicating a greater disparity with PBE$^{+\text{MBD@rsSCS}}$. This is reflected by the larger range of RMSDs observed with au_full over auorg. The

auorg$^{+\text{MBD@rsSCS}}$-optimised geometries for the individual gold substrates and organic molecules are in close agreement with the DFT results. The RMSDs of optimised organic molecules on Au$_{18}$, Au$_{34}$ and Au(111) surfaces are all below 1.0 Å, though some structural disparities were observed. The greatest structural disparities with respect to DFT were from organic molecules on Au$_{10}$ NCs. In contrast, the au_full$^{+\text{MBD@rsSCS}}$-optimised larger gold substrates had RMSDs of around 1.0 Å with respect to DFT. The RMSDs of many adsorbed organic molecules were also greater than 1.0 Å, with four complexes having RMSDs greater than 2.0 Å. However, the structures of au_full$^{+\text{MBD@rsSCS}}$-optimised individual organic molecules, as well as the adsorbed molecules on gold (111) surfaces, were in close agreement with PBE$^{+\text{MBD@rsSCS}}$.
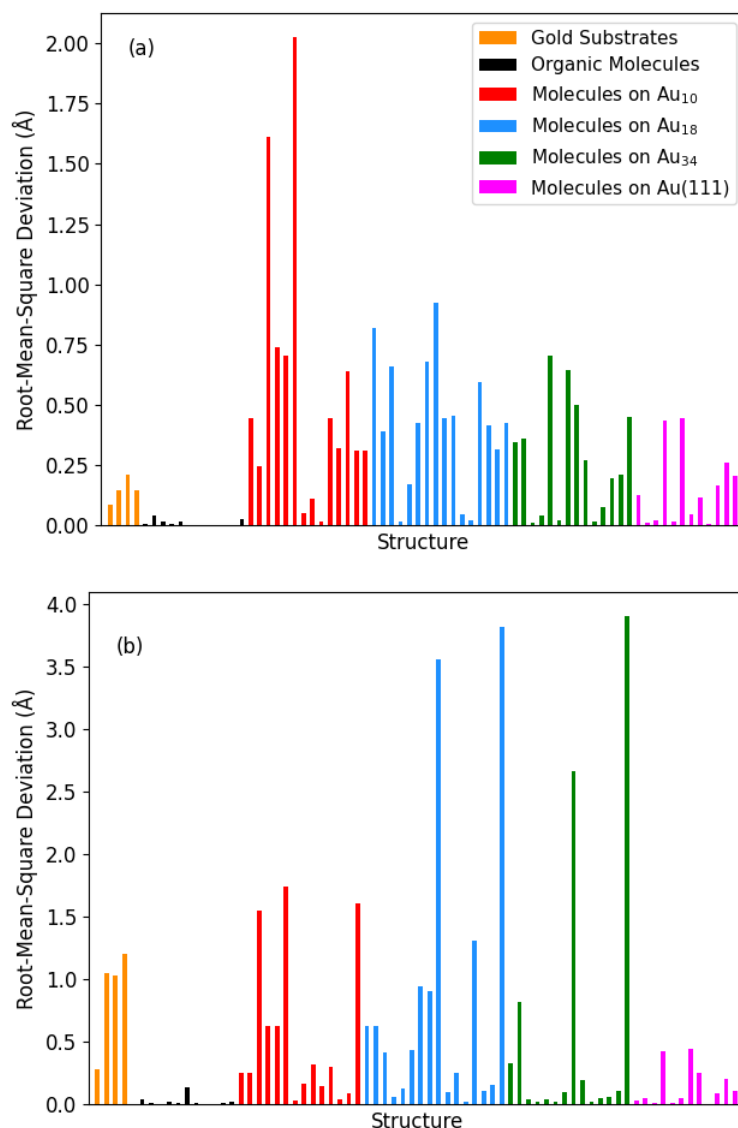
Figure 5.3. RMSDs for (a) auorg$^{+\text{MBD@rsSCS}}$- and (b) au_full$^{+\text{MBD@rsSCS}}$-optimised gold substrates, organic molecules and combined systems with respect to their corresponding PBE$^{+\text{MBD@rsSCS}}$-optimised structures.

Figures 5.4 and 5.5 summarise the adsorption energies of the molecules in Figure 5.2 on different gold substrates using au_full$^{+\text{MBD@rsSCS}}$ and auorg$^{+\text{MBD@rsSCS}}$ respectively as compared to PBE$^{+\text{MBD@rsSCS}}$. As can be seen from Figures 5.4 and 5.5, the mean absolute errors (MAEs) of adsorption energies with au_full$^{+\text{MBD@rsSCS}}$ are larger than auorg$^{+\text{MBD@rsSCS}}$ with respect to PBE$^{+\text{MBD@rsSCS}}$ for all substrates apart from the Au$_{34}$ cluster. This is also reflected in the Pearson correlation coefficients of adsorption energies, where au_full$^{+\text{MBD@rsSCS}}$ has values of 0.25 and 0.34 for the

Au$_{10}$ and Au$_{18}$ substrates respectively, which indicates very little correlation with PBE$^{+\text{MBD@rsSCS}}$. In contrast, auorg$^{+\text{MBD@rsSCS}}$ has Pearson correlation coefficients of 0.94 and 0.56 for the Au$_{10}$ and Au$_{18}$ substrates respectively, which indicates much higher correlation with PBE$^{+\text{MBD@rsSCS}}$. However, au_full$^{+\text{MBD@rsSCS}}$ has higher Pearson correlation coefficients for the Au$_{34}$ and Au(111) substrates.
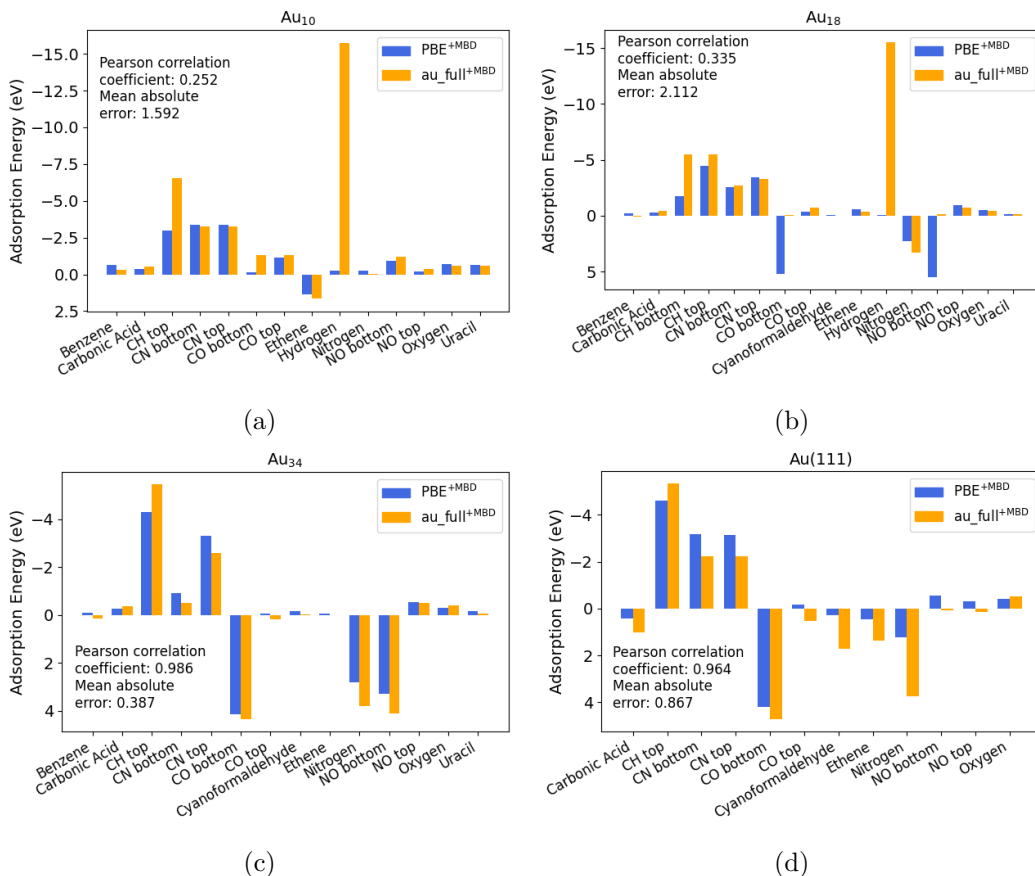
Figure 5.4. Comparison of adsorption energetics of different organic molecules on different gold substrates, calculated using PBE$^{+\text{MBD}}$ and au_full$^{+\text{MBD}}$. The Pearson correlation coefficients and MAEs for each substrate are also provided.

It can be argued that outliers are responsible for skewing the final MAEs and Pearson correlation coefficients. For example, the adsorption energies of molecular hydrogen on Au$_{10}$ and Au$_{18}$ are calculated to be around -15.0 eV with au_full$^{+\text{MBD@rsSCS}}$, whereas both MBD@rsSCS-corrected auorg and PBE predict very weak adsorption. The reason for these differences could be because short-range hydrogen-hydrogen interactions were not included within the training dataset for au_full. Molecular hydrogen was not included as an adsorbate on Au$_{34}$ and Au(111) surfaces due to its tendency to desorb on large gold clusters and bulk gold [160; 499]; for these

substrates, the au_full$^{+\text{MBD@rsSCS}}$ Pearson correlation coefficients are higher than auorg$^{+\text{MBD@rsSCS}}$, showing the potential effect of large outliers on correlation and errors.

(a)

(b)

(c)

(d)
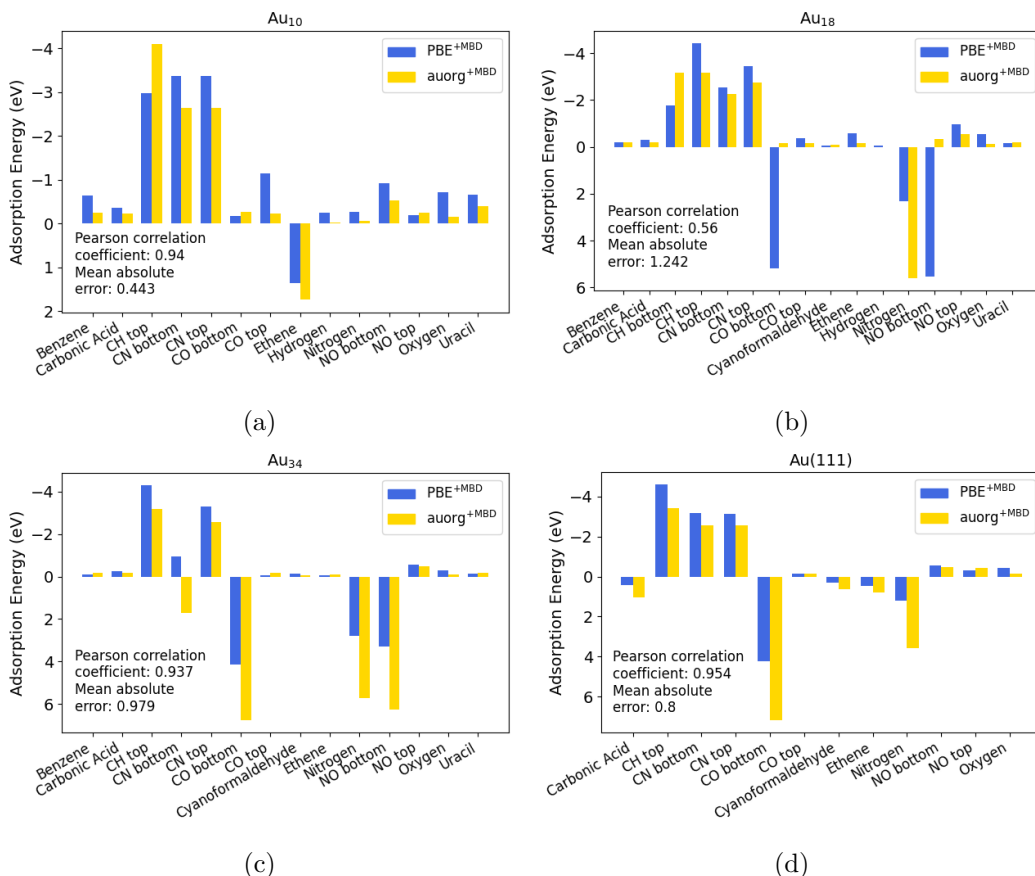
Figure 5.5. Comparison of adsorption energetics of different organic molecules on different gold substrates, calculated using PBE$^{+\text{MBD}}$ and auorg$^{+\text{MBD}}$. The Pearson correlation coefficients and MAEs for each substrate are also provided.

The disagreement between MBD@rsSCS-corrected auorg and PBE might be due to the properties of the auorg parameter set and the nature of its constituent parameters. While the mio parameterisation, which auorg is an extension of, was designed for the description of interactions within biological and organic molecules [78], the auorg parameterisation itself was designed to describe optical excitations of thiolates chemisorbed on gold NCs [86]. The applicability of the auorg parameterisation for the description of interactions between gold NCs and organic/carbonaceous materials is therefore questionable. It should also be noted that at the time of use, the au_full parameter set was still in development and not perfected, and its intrinsic properties could be responsible for the observed discrepancies with DFT. For

example, the smoothing constraint on the second derivative, which was set to be strictly positive, can result in piece-wise linear behaviour in some of the potentials. This constraint can interfere with the natural boundary conditions used, where the second derivative is equal to zero at end points. Furthermore, as observed with molecular hydrogen, au_full is currently unable to accurately model short-range interactions within molecules not in its training dataset. There is scope to improve the performance of au_full, but in its current version, it has been shown to outperform auorg, in terms of energies, for molecular adsorbates that were included within its training data. What remains now is to gauge the performance of both gold-based DFTB2 parameterisations for gold NCs adsorbed on diamond.

Gold Tetramers on Diamond (110) Surfaces

Both the auorg and au_full DFTB2 parameterisations were used to benchmark the adsorption of three gold tetramers adsorbed onto a diamond (110) surface. Gold atoms were arranged collinearly ('Geometry 1'), quadrilaterally ('Geometry 2') and tetrahedrally ('Geometry 3'). Geometry optimisations, with the carbon atoms fully constrained, were conducted on each system. Table 5.1 details the RMSDs of the structures as optimised with different methods with respect to the corresponding PBE$^{+\text{MBD@rsSCS}}$-optimised structure, as well as the adsorption energies of the gold NCs, while Table 5.2 visualises the three initial geometric structures as well as their optimised geometries with different methods.

First, the structural differences for the three supported gold tetramers after DFT and DFTB2 optimisations are analysed. As can be seen in Table 5.2, PBE$^{+\text{MBD@rsSCS}}$ conserves the collinear nature of the four gold atoms with respect to the initial Geometry 1. While three of the gold atoms are collinear with auorg$^{+\text{MBD@rsSCS}}$, the fourth atom moves above this plane to form an irregular tetrahedral structure, while auorg$^{+\text{DFT-D3}}$ results in a non-collinear arrangement of the gold atoms. In contrast, the au_full$^{+\text{MBD@rsSCS}}$-optimised Geometry 1 retains the initial collinear arrangement of gold atoms, similar to PBE$^{+\text{MBD@rsSCS}}$. These structural differences for Geometry 1 are reflected in the RMSDs of various methods, where both auorg$^{+\text{MBD@rsSCS}}$ and auorg$^{+\text{DFT-D3}}$ result in high RMSDs (2.21 Å and 1.93 Å respectively) with respect to PBE$^{+\text{MBD@rsSCS}}$. In contrast, the RMSD of the au_full$^{+\text{MBD@rsSCS}}$-optimised structure is 0.32 Å, which is much lower than both auorg methods and shows a much better agreement with PBE$^{+\text{MBD@rsSCS}}$. The geometrical differences for Geometry 2 are less profound, with all methods retaining the quadrilateral arrangement of the gold atoms. Both dispersion-corrected

138

auorg methods result in a more compact arrangement of the gold atoms as compared to PBE$^{+\text{MBD@rsSCS}}$, while au_full$^{+\text{MBD@rsSCS}}$ results in a structure that is more rectangular than the rhomboidal-like nature of the PBE$^{+\text{MBD@rsSCS}}$-optimised structure, as can be seen in Table 5.2. This is reflected in Table 5.1, where both auorg methods result in structures that possess lower RMSDs (less than 0.9 Å) than au_full$^{+\text{MBD@rsSCS}}$ (1.16 Å) with respect to PBE$^{+\text{MBD@rsSCS}}$. For Geometry 3, all the methods retain the tetrahedral shape of the initial structure, though both dispersion-corrected auorg methods result in a more compact and regular tetrahedral shape than PBE$^{+\text{MBD@rsSCS}}$. This structural difference is reflected by the RMSDs of both dispersion-corrected auorg methods being greater than 1 Å with respect to PBE$^{+\text{MBD@rsSCS}}$. In contrast, the structure obtained after au_full$^{+\text{MBD@rsSCS}}$ optimisation is more similar to PBE$^{+\text{MBD}}$, as is reflected in its RMSD of 0.52 Å.

Table 5.1 also details the adsorption energies of the gold NCs in the three geometries. For any given method, all three geometries have similar adsorption energies. However, DFT-D3-corrected auorg consistently overestimates adsorption energies; this phenomenon has previously been reported to occur with the DFT-D3 scheme [69]. Between the MBD@rsSCS-corrected methods, auorg predicts fairly similar adsorption energetics to PBE for all three geometries, despite the relatively large structural differences. In contrast, au_full$^{+\text{MBD@rsSCS}}$ consistently underestimates adsorption energies, which are at least $0.50\,\text{eV\,atom}^{-1}$ weaker than PBE$^{+\text{MBD@rsSCS}}$ despite the high agreement in optimised structures. This shows that the two MBD@rsSCS-corrected gold-based DFTB2 parameterisations investigated herein either approximate the adsorption energetics of gold NCs well, or result in a final optimised structure that is similar to the DFT-optimised structure, but never both (at least for the systems investigated). An alternative method must therefore be sought that can both optimise geometries correctly and calculate accurate energetics, and thus accurately explore the space of possible gold NC structures.

Method	RMSD (Å)	Adsorption Energy (eV atom^{-1})
Geometry 1		
PBE$^{+MBD@rsSCS}$	0	-2.64
au_full$^{+MBD@rsSCS}$	0.32	-2.00
auorg$^{+MBD@rsSCS}$	2.21	-2.59
auorg$^{+DFT-D3}$	1.93	-3.33
Geometry 2		
PBE$^{+MBD@rsSCS}$	0	-2.63
au_full$^{+MBD@rsSCS}$	1.16	-2.13
auorg$^{+MBD@rsSCS}$	0.89	-2.55
auorg$^{+DFT-D3}$	0.84	-3.62
Geometry 3		
PBE$^{+MBD@rsSCS}$	0	-2.77
au_full$^{+MBD@rsSCS}$	0.52	-2.09
auorg$^{+MBD@rsSCS}$	1.22	-2.57
auorg$^{+DFT-D3}$	1.05	-3.65

Table 5.1. Comparison of adsorption energetics and RMSDs of dispersion-corrected DFTB2 optimisations of the clusters visualised in Table 5.2 with respect to PBE$^{+MBD@rsSCS}$-optimised structures.

Method	Geometry 1	Geometry 2	Geometry 3
Initial			
PBE$^{+\text{MBD@rsSCS}}$			
au_full$^{+\text{MBD@rsSCS}}$			
auorg$^{+\text{MBD@rsSCS}}$			
auorg$^{+\text{DFT-D3}}$			

Table 5.2. Orthographic visualisations of the Geometry 1, Geometry 2 and Geometry 3 systems, as optimised using various methods. Each structure is a gold tetramer adsorbed onto a diamond (110) surface, with the gold atoms being arranged (a) collinearly (Geometry 1), (b) quadrilaterally (Geometry 2) and (c) tetrahedrally (Geometry 3). Structures are viewed 'top down' from the [110] direction, and gold and carbon atoms are shown in gold and grey respectively.

5.3.2 Accuracy and Efficiency of MLIPs

Development of MLIPs

As dispersion-corrected DFTB2 was determined to be an inaccurate alternative to PBE$^{+\text{MBD@rsSCS}}$, the development of an MLIP trained on PBE$^{+\text{MBD@rsSCS}}$ data was sought after instead. Two collections of MLIPs were trained: one set of models was trained on E and F, while the other was trained on Hirshfeld volume ratios, H. Four sets of MLIPs were trained (the initial model and three adaptively-sampled models), and every set comprised four models for (E, F) predictions, each differing in the choice of trade-off value, and one model for H predictions.

Table 5.3 details the MAEs of E, F and H predictions on a hold-out set that the models were not trained on. As can be seen, the largest MAE for H predictions was from the SchNet$_{\text{adapt3}}(H)$ model and was evaluated to be 1.1×10^{-4}, showcasing the high accuracy of all the models. For E and F, four SchNet(E, F) models were trained, each with a specific t valuebut with random weights initialised within the NN. The MAEs of the four SchNet(E, F) models range from 0.01–0.18 eV and 0.01–0.07 eV Å^{-1} for E and F respectively. The largest errors arise from the SchNet$_{\text{adapt2}}$ and SchNet$_{\text{adapt3}}$ models, but this was to be expected as these MLIPs included geometries that were energetically unfavourable to ensure the models could identify less favourable structures. Discounting just eight of these unfavourable geometries, the MAEs for the E predictions from the SchNet$_{\text{adapt2}}$ and SchNet$_{\text{adapt3}}$ MLIPs decrease by about a third, while the MAEs for F predictions are almost halved, making the errors more comparable to the SchNet$_{\text{initial}}$ and SchNet$_{\text{adapt1}}$ MLIPs.

Model	Energies, E (eV)	Forces, F (eV Å^{-1})	Hirshfeld Volume Ratios, H
SchNet$_{\text{initial}}$			
$(E,F)_{t=0.03}$	0.01	0.02	–
$(E,F)_{t=0.04}$	0.01	0.02	–
$(E,F)_{t=0.04}$	0.01	0.02	–
$(E,F)_{t=0.05}$	0.01	0.02	–
(H)	–	–	8.1×10^{-5}
SchNet$_{\text{adapt1}}$			
$(E,F)_{t=0.03}$	0.02	0.01	–
$(E,F)_{t=0.04}$	0.02	0.01	–
$(E,F)_{t=0.04}$	0.04	0.02	–
$(E,F)_{t=0.05}$	0.03	0.02	–
(H)	–	–	3.9×10^{-5}
SchNet$_{\text{adapt2}}$			
$(E,F)_{t=0.03}$	0.09	0.03	–
$(E,F)_{t=0.04}$	0.14	0.05	–
$(E,F)_{t=0.04}$	0.18	0.06	–
$(E,F)_{t=0.05}$	0.14	0.06	–
(H)	–	–	6.2×10^{-5}
SchNet$_{\text{adapt3}}$			
$(E,F)_{t=0.03}$	0.09	0.04	–
$(E,F)_{t=0.04}$	0.09	0.05	–
$(E,F)_{t=0.04}$	0.12	0.04	–
$(E,F)_{t=0.05}$	0.12	0.07	–
(H)	–	–	1.1×10^{-4}

Table 5.3. MAEs of E, F and H values compared to PBE$^{+\text{MBD@rsSCS}}$ as predicted by the various SchNet-based MLIPs on a hold-out set that the models were not trained on. The t parameter denotes the trade-off value that was used within the loss function to weight energies and forces.

As NCs can exhibit many metastable geometries, the performance of the final $SchNet_{adapt3}$+MBD@rsSCS model was assessed by comparing interatomic distances and evaluating the ability of the model to . Figure 5.6 shows the radial atom distributions of NCs comprising 6, 15, 20, 25, 28, 30, 35, 40, 44, 45, 60 and 66 atoms, and how the interatomic distances change after optimisations with $SchNet_{adapt3}$+MBD@rsSCS and PBE+MBD@rsSCS. As can be seen from Figure 5.6(a), initial gold-gold distances smaller than 2.6 Å disappear after optimisation with both methods. Furthermore, both methods result in a peak at around 2.8 Å, showcasing the accuracy of $SchNet_{adapt3}$+MBD@rsSCS with respect to PBE+MBD@rsSCS. Some deviations can be seen in Figure 5.6(b) at around 2.5 Å for the initial gold-carbon distances, but the peaks and troughs shift towards the PBE+MBD@rsSCS values after optimisation with $SchNet_{adapt3}$+MBD@rsSCS.
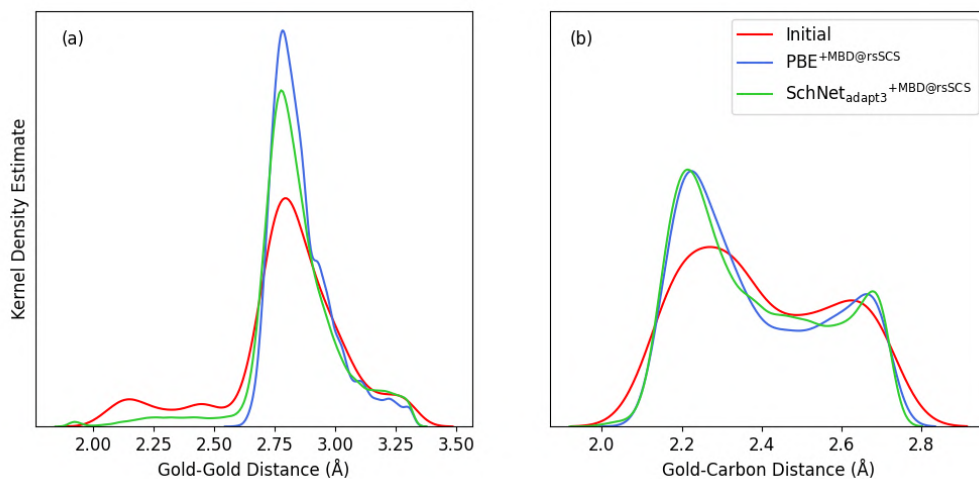
Figure 5.6. Kernel density estimates for the radial atom distributions of (a) gold-gold and (b) gold-carbon bonds.

Exploration of Possible Supported Gold NC Structures

To further validate the accuracy of the trained MLIPs, the global minimum and two random local minima were selected from basin-hopping runs conducted using $SchNet_{initial}$+MBD@rsSCS for the first adaptive sampling run. For 6-, 15- and 40-atom gold NCs, all three chosen minima were subsequently optimised using MBD@rsSCS-corrected DFT, DFTB2 and MLIPs. All optimised structures are visualised in Table B.2, while their RMSDs with respect to the PBE+MBD@rsSCS-optimised structures are detailed in Table 5.4.

Method	Global Minimum	Local Minimum 1	Local Minimum 2
Au_6			
SchNet$_{initial}$	1.37	0.66	1.40
SchNet$_{adapt3}$	1.35	0.11	0.34
auorg	1.31	2.31	2.86
au_full	1.87	1.62	2.48
Au_{15}			
SchNet$_{initial}$	0.14	0.38	4.82
SchNet$_{adapt3}$	0.13	0.16	4.95
auorg	0.74	1.04	4.95
au_full	1.84	1.19	5.64
Au_{40}			
SchNet$_{initial}$	0.41	0.32	0.36
SchNet$_{adapt3}$	0.20	0.18	0.31
auorg	0.91	0.87	0.93
au_full	1.81	1.94	1.80

Table 5.4. RMSDs (in angstroms) of 6-, 15- and 40-atom gold NCs, adsorbed onto a diamond (110) surface, optimised with different MBD@rsSCS-corrected methods with respect to PBE$^{+MBD@rsSCS}$. For each NC, the global minimum and the two randomly-chosen local minima are from a basin-hopping run conducted with SchNet$_{initial}$$^{+MBD@rsSCS}$ for the first adaptive sampling run, and structures were subsequently optimised with the listed methods.

Figure 5.7. Relative energies with respect to the energetically lowest NC calculated using various MBD@rsSCS-corrected methods. Structures are the global minimum as well as two randomly-chosen local minima from basin-hopping runs conducted using the $SchNet_{initial}+^{MBD@rsSCS}$ MLIP on supported (a) Au_6, (b) Au_{15} and (c) Au_{40} NCs visualised in Table B.2 for the first adaptive sampling run.

As can be seen in Table 5.4, the RMSDs of MLIP-optimised structures with respect to DFT generally improves after adaptive sampling. With the exception of the second Au_{15} local minimum, the RMSDs of MLIP-optimised geometries are generally quite low, indicating close structural agreement with $PBE^{+MBD@rsSCS}$-optimised NCs. The large disparities observed with the second Au_{15} local minimum, where RMSDs were calculated to be greater than 4.8 Å, could indicate that this structure lies within a region of the potential energy surface that was under-represented within the training dataset. This is supported by the energy variance from the query-by-committee approach for this structure (30 meV), which was around 4 times higher than for other clusters (around 7 meV). The RMSDs of MLIP-optimised geometries are also typically lower than those of DFTB2 parameterisations. The $au\_full^{+MBD@rsSCS}$ methods consistently result in structures with RMSDs greater than 1.0 Å with respect to DFT, and while the RMSDs of $auorg^{+MBD@rsSCS}$-optimised NCs are not as large, they are still larger than MLIP-optimised geometries. Figure 5.7 details the relative energies of the optimised minima as calculated using various methods. All adaptively-sampled MLIPs correctly order the energies as compared to $PBE^{+MBD@rsSCS}$, though the $SchNet_{initial}+^{MBD@rsSCS}$ MLIP incorrectly orders the Au_{15} structures. For all three NC sizes, MLIP-calculated relative ener-

146

gies become more similar to the PBE$^{+MBD@rsSCS}$-calculated values as the number of adaptive sampling steps increases, with the SchNet$_{adapt3}$$^{+MBD@rsSCS}$ MLIP resulting in relative energies that are very similar to DFT. The au_full parameterisation correctly orders 6- and 15-atom energies as compared to DFT, but the relative energies themselves are generally not as accurate compared to SchNet$_{initial}$$^{+MBD@rsSCS}$. However, au_full$^{+MBD@rsSCS}$ incorrectly orders the 40-atom minima, while auorg$^{+MBD@rsSCS}$ consistently orders energies incorrectly for all three NC sizes. This shows that once again, the DFTB2 parameterisations investigated here either perform well energetically or result in a final optimised structure that is similar to the DFT-optimised geometry, but never both (at least for the systems investigated). This is in contrast to the trained MLIPswhich generally perform well structurally and energetically with respect to DFT.

For the Au$_6$ NCs, the most stable geometry after PBE$^{+MBD@rsSCS}$ optimisation was found to be a rhomboidal frustum, with four gold atoms arranged on the diamond (110) surface and two gold atoms placed above these four atoms. This structure has been observed to be the most stable supported gold hexamer on substrates such as cerium(IV) oxide (111) surfaces [407]. Some structural disparity can be seen for the SchNet$_{adapt3}$$^{+MBD@rsSCS}$-optimised global minimum, which has an RMSD greater than 1.3 Å with respect to PBE$^{+MBD@rsSCS}$. However, both SchNet$_{adapt3}$$^{+MBD@rsSCS}$-optimised local minima are very similar to their respective PBE$^{+MBD@rsSCS}$-optimised structures, with RMSDs below 0.35 Å. While the auorg$^{+MBD@rsSCS}$-optimised global minimum has a marginally lower RMSD that is 0.04 Å smaller than SchNet$_{adapt3}$$^{+MBD@rsSCS}$, the RMSDs for the optimised local minima are greater than 2.3 Å, showing a large disparity with DFT. As can be seen in Table B.2, optimisation with auorg$^{+MBD@rsSCS}$ results in the two Au$_6$ local minima to form highly symmetrical triangles that lie parallel to the diamond (110) surface, and these structures are more energetically stable than the auorg$^{+MBD@rsSCS}$-optimised global minimum. Such triangular Au$_6$ structures have been shown to be the most stable Au$_6$ geometry on various substrates such as magnesium oxide [407; 450] and graphite (0001) surfaces [407], and have also been shown to be most stable gas-phase Au$_6$ structure [500]. However, of all investigated methods, auorg$^{+MBD@rsSCS}$ is the only one to result in such a triangular atomic arrangement, which indicates . As mentioned above, au_full$^{+MBD@rsSCS}$ has fairly large RMSDs with respect to DFT for all three minima and has the largest RMSD among all investigated methods for the optimised global minimum (1.87 Å)

For the Au_{15} NCs, the most stable structure was found to be pyramidal, with nine gold atoms arranged on the diamond (110) surface to form the polyhedron base, five gold atoms above these nine atoms and finally one atom at the apex of the geometry. A truncated version of this pyramidal geometry with nine gold atoms at the base and six atoms above them (i.e. no apex) has been observed to be the most stable supported geometry on cerium(IV) oxide (111) surfaces, though planar geometries parallel to the substrate surface were observed to be more stable on magnesium oxide (001) and graphite (0001) surfaces [407]. There is a high degree of agreement for the global minimum structure after optimisations with $PBE^{+MBD@rsSCS}$ and the MLIPs, with RMSDs less than $0.15\,\text{Å}$. The first local minimum after $SchNet_{adapt3}^{+MBD@rsSCS}$ optimisation is also very similar to the $PBE^{+MBD@rsSCS}$-optimised structure, with an RMSD of $0.16\,\text{Å}$, and the RMSDs of the structures optimised using MLIPs generally decreases as the number of adaptive sampling runs increases. Despite this discrepancy with DFT, the MLIPs outperform both DFTB2 parameterisations for all three Au_{15} minima. For the global minimum, $auorg^{+MBD@rsSCS}$ predicts a final structure with an RMSD of $0.74\,\text{Å}$, which is $0.60\,\text{Å}$ greater than the RMSDs of the MLIPs. For the first local minimum, the $auorg^{+MBD@rsSCS}$ RMSD is greater than $1.0\,\text{Å}$, while the RMSD for the second local minimum is $4.95\,\text{Å}$, which shows that it performs no better than the trained MLIPs. Once again, $au\_full^{+MBD@rsSCS}$ results in very different structures with RMSDs greater than $1.1\,\text{Å}$ for all three Au_{15} minima, and has the largest RMSD for any Au_{15} NC among all investigated methods.

For the Au_{40} NCs, the most stable structure after $PBE^{+MBD@rsSCS}$ optimisation was also found to be pyramidal, with 15 gold atoms forming the polyhedron base, 14 gold atoms above the base, 10 above them and one atom at the apex. Not much data exists to compare this supported structure against other substrates, though the pyramidal structure is quite different to the truncated octadehral structure observed for gas-phase Au_{40} NCs [501]. However, the pyramidal structure is similar to the twisted pyramidal gas-phase structure that was calculated to be the global minimum for Au_{40} NCs by Jiang and Walter [461] and is also qualitatively not too dissimilar to the STEM images of BDD-supported Au_{40} NCs reported by Hussein *et al.* [22].
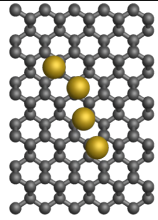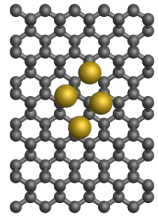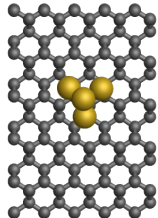
Structure	Optimised Geometry	RMSD (Å)
Geometry 1		1.04
Geometry 2		0.36
Geometry 3		0.77

Table 5.5. Comparison of the final structures after SchNet$_{adapt3}$$^{+MBD@rsSCS}$ optimisations of the Geometry 1, Geometry 2 and Geometry 3 gold tetramers visualised in Table 5.2. RMSD values are with respect to the corresponding PBE$^{+MBD@rsSCS}$ structure, and relative energies are given with respect to the energetically lowest tetramer (in this case, Geometry 3). Structures are shown from the [110] direction as a top view, and gold and carbon atoms are shown in gold and grey respectively.

The accuracy of SchNet$_{adapt3}$$^{+MBD@rsSCS}$ was further investigated by using the MLIP to optimise the three gold tetramers (Geometry 1, Geometry 2 and Geometry 3) shown in Table 5.2. Four-atom gold NCs were not included within the training dataset and can therefore be used to gauge the accuracy of the MLIP for structures outside its training regime. Table 5.5 visualises the final optimised structures and details their RMSDs with respect to the PBE$^{+MBD@rsSCS}$-optimised structures in Table 5.1. Some structural disparity can be seen for Geometry 1 with an RMSD of 1.04 Å, though this is still better than the RMSDs of 2.21 Å and 1.93 Å calculated with auorg$^{+MBD@rsSCS}$ and auorg$^{+DFT-D3}$ respectively. For Geometry 2, the quadrilateral arrangement of the atoms is retained after optimisation with SchNet$_{adapt3}$$^{+MBD@rsSCS}$ and this is reflecting by its low RMSD (0.36 Å) with respect to PBE$^{+MBD@rsSCS}$. Finally, for Geometry 3, the optimised geometry is tetrahedral,

but is more compact and regular than the PBE$^{+\text{MBD@rsSCS}}$-optimised structure. The RMSD was calculated to be 0.77 Å, which is not as low as the au_full$^{+\text{MBD@rsSCS}}$ value (0.52 Å) but is lower than values from dispersion-corrected auorg.

Figure 5.8 details the relative energies of the three tetramers as calculated with MBD@rsSCS-corrected DFT, DFTB2 and SchNet$_{\text{adapt3}}$. As can be seen, both DFTB2 methods incorrectly order the tetramer energies as compared to PBE. However, similar to PBE$^{+\text{MBD@rsSCS}}$, SchNet$_{\text{adapt3}}$$^{+\text{MBD@rsSCS}}$ correctly predicts Geometry 3 to be the most stable NC. However, the MLIP incorrectly orders the Geometry 1 and Geometry 2 energies as compared to DFT. Despite this discrepancy, it should be noted that the MLIP still outperforms both DFTB2 parameterisations, predicts final NCs that are in close agreement with the DFT-optimised geometries and was able to predict the most stable structure despite 4-atom NCs not being within its training dataset. Adaptive sampling could certainly be utilised to improve the accuracy of the MLIP predictions for structures either under-represented or neglected within the training dataset.
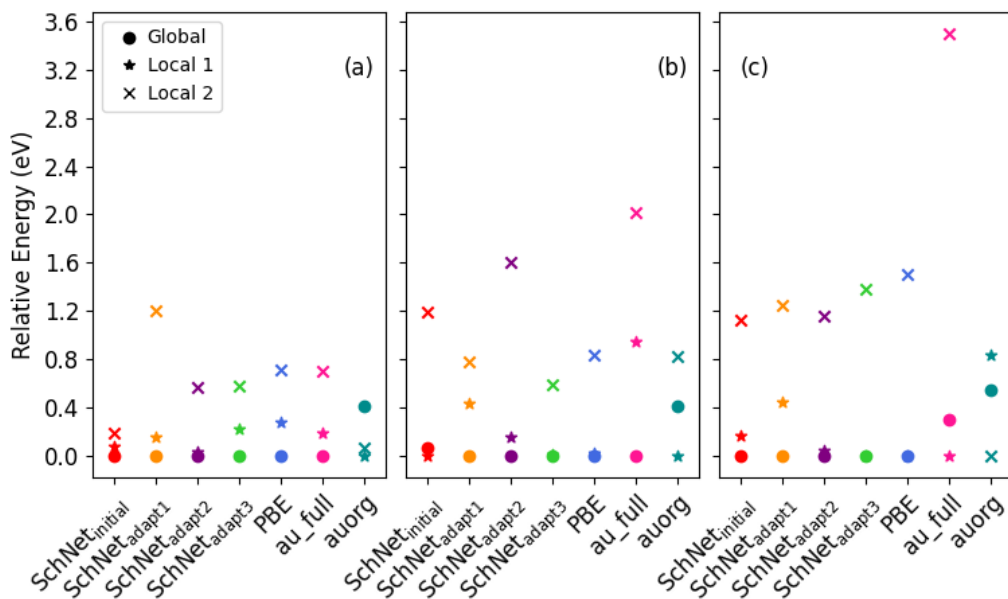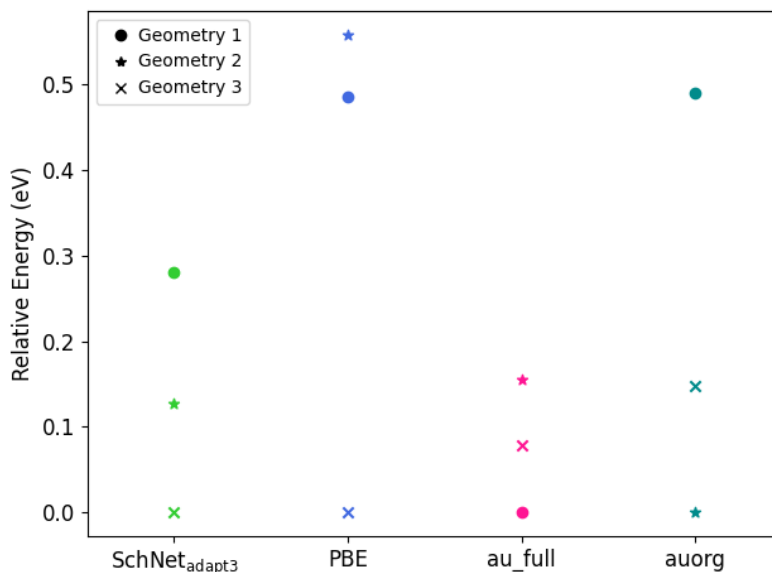

Figure 5.8. Relative energies with respect to the energetically lowest NC calculated using various MBD@rsSCS-corrected methods. Structures are the Geometry 1, Geometry 2 and Geometry 3 supported gold tetramers visualised in Table 5.2.

Having established the accuracy of the SchNet$_{\text{adapt3}}$$^{+\text{MBD@rsSCS}}$ MLIP with respect to PBE$^{+\text{MBD@rsSCS}}$, it is also important to compare its relative computational efficiency, and showcase its ability to facilitate fast structure optimisations.

Method	Calculation Time (s)
Geometry 1	
PBE	69543
auorg	8929
au_full	2784
SchNet$_{adapt3}$	407
Geometry 2	
PBE	74818
auorg	4501
au_full	1216
SchNet$_{adapt3}$	367
Geometry 3	
PBE	78534
auorg	2882
au_full	1652
SchNet$_{adapt3}$	212

Table 5.6. Computational costs (rounded to the nearest integer) of geometry optimisations conducted using various MBD@rsSCS-corrected approaches. Systems are the Geometry 1, Geometry 2 and Geometry 3 supported gold tetramers visualised in Table 5.2. The PBE$^{+\mathrm{MBD@rsSCS}}$ calculations were performed using 128 cores (8 nodes with 16 cores per node); all other methods used 1 core.

Table 5.6 shows the computational costs of conducting geometry optimisations on the Geometry 1, Geometry 2 and Geometry 3 supported gold tetramers using different methods. The DFTB2 and MLIP optimisations for these tetramers were performed using 1 core on 1 Fujitsu Primergy RX2530 M5 compute node provided by the Scientific Computing Research Technology Platform of the University of Warwick, while the DFT calculations were performed using 128 cores (8 nodes with 16 cores per node) on Lenovo NeXtScale nx360 M5 servers with dual Intel Xeon E5-2630 v3 2.4 GHz (Haswell) 8-core processors, also provided by the Scientific Computing Research Technology Platform within their Tinis high-performance computing cluster. Even with 128 cores used in parallel, the PBE$^{+\mathrm{MBD@rsSCS}}$ calculations took over 19 h to complete. In contrast, using 1 core, MBD@rsSCS-corrected

DFTB2 optimisations were at least an order of magnitude cheaper than DFT calculations, while $SchNet_{adapt3}+^{MBD@rsSCS}$ optimisations required less than 7 min. This highlights that the MLIPs can be used to facilitate both fast and accurate structure optimisations.

5.3.3 Supported Au_{147} Nanoclusters

Having shown the ability of $SchNet_{adapt3}+^{MBD@rsSCS}$ to facilitate fast and accurate optimisations, the MLIP was subsequently used to optimise large gold NCs on diamond. 147-atom gold NCs were chosen for this purpose due to their size, which would be computationally to optimise using DFT. Furthermore, NCs comprising certain special numbers of atoms, such as 147, are much more abundant than others when generated in typical cluster experiments [502; 503]. These numbers are termed 'magic' and have been studied extensively in literature for their unique properties [465; 502; 503; 504; 505]. To investigate the stability of Au_{147} NCs supported on diamond, three isomers were constructed, placed onto a diamond (110) surface and subsequently optimised using $SchNet_{adapt3}+^{MBD@rsSCS}$. Two of these isomers were highly-symmetric and of regular icosahedral and cuboctahedral shapes respectively, while a third amorphous isomer identified by Tarrat *et al.* [465] was also investigated; all three isomers are visualised in Figure 5.1. Table 5.7 details the structural and energetic differences between the supported isomers after optimisation with $SchNet_{adapt3}+^{MBD@rsSCS}$, which are visualised in Figure 5.1.

The cuboctahedral NC was found to be the energetically lowest structure of the three isomers investigated. The icosahedral isomer was calculated to be 0.97 eV higher in energy, while the energy of the amorphous isomer was 6.37 eV higher than the cuboctahedral isomer. This indicates that, at least for 147-atom NCs, ordered structures are more energetically favourable than amorphous structures on diamond (110) surfaces. Table 5.7 also suggests that as the sphericity decreases, the more energetically favourable the supported NC is. This is most likely because NCs with lower sphericities have more atoms closer to the substrate surface, which results in a greater interaction between the NC adsorbate and diamond surface. Indeed, as can be seen from the visualisations of the isomers in Figure 5.9, the cuboctahedral NC has the most atoms (16) directly above the surface, which is six more than the icosahedral NC (10). In contrast, the amorphous NC only has 4 atoms directly above the surface. It should be noted that the observed trend between sphericity and relative energies is most likely not universal, and there will most likely be bounds on both the NC size and sphericity for this trend to be valid. However, the sphericity

Figure 5.9. Orthographic visualisations of the optimised Au_{147} isomers, as optimised using $SchNet_{adapt3}{}^{+MBD@rsSCS}$. Each structure is adsorbed onto a diamond (110) surface, with the gold atoms being arranged (a) cuboctahedrally, (b) icosahedrally and (c) amorphously. Structures are viewed 'front on' from the [001] direction and 'top down' from the [110] direction, with surface axes presented. Also shown on the right are the convex hulls (blue) of each surface-adsorbed NC. Gold and carbon atoms are shown in gold and grey respectively.

153

measure can be used to explain the trends in relative energies observed between the three isomers.

Initial NC Structure	NC Sphericity	Relative Energy (eV)
Cuboctahedral	0.920	0
Icosahedral	0.951	0.97
Amorphous	0.955	6.37

Table 5.7. Sphericities and relative energies of three Au_{147} isomers supported on a diamond (110) surface after optimisation with $SchNet^{+MBD@rsSCS}$. Relative energies are given with respect to the energetically lowest supported isomer.

As shown in Table 5.7, the optimised cuboctahedral and icosahedral isomers have sphericities of 0.920 and 0.951, which are both greater than the sphericities for regular cuboctahedra (0.905) and icosahedra (0.939). The sphericity of the amorphous isomer was also observed to marginally increase from 0.953 to 0.955. This indicates that the adsorbed NCs are more compact and spherical than the initial atomic arrangement upon adsorption. It should be noted that the small differences in absolute sphericity values do not correlate with geometrical similarity. While the RMSD between the pre- and post-optimised cuboctahedral NCs was evaluated to be $0.22\,\text{Å}$, indicating very similar structures, the RMSD between the initial and final icosahedral isomers was $0.84\,\text{Å}$. The RMSD between the amorphous NCs was $0.86\,\text{Å}$, indicating some structural disparity despite the sphericity 'only' increasing by 0.002. Moreover, the sphericities reported herein are for irregular polyhedra, and are thus approximations to the true sphericities. This is because for each irregular NC, the volume and surface area that enter Equation (5.4) are of its convex hulls, and are therefore upper bounds on the true volume and surface area of the NC. Further investigation would therefore be required to ascertain if there exists a true correlation between sphericity and energetic stability.

While the accuracy and utility of the SchNet-based MLIPs have been showcased, it should be noted that there are factors that were not accounted for during their development. For example, while the MLIPs account for long-range vdW interactions, they do not explicitly treat electrostatic interactions. However, this could theoretically be achieved by extending the MLIPs to learn partial atomic charges, similar to the SpookyNet [317] and Behler's fourth-generation NNs [487] architectures, along with energies, forces and Hirshfeld volume ratios. Furthermore, while

the MLIPs were able to calculate relative energies and perform structure optimisations, they do not provide other information that electronic structure methods such as DFT offer e.g. band gaps.

The effect of the DFA that the models are trained on should also be investigated. In this chapter, the performance of the MLIPs were only benchmarked against DFTB2 parameterisations and the PBE [229] GGA. Using the procedure outlined in this chapter, MLIPs should be trained on different DFAs, including higher-rung MGGAs and HGGAs. It could also be worthwhile to investigate other ML and NN architectures for the training of MLIPs. For example, the descriptors within the SchNet NN are rotationally invariant and therefore do not account for any directional, equivariant information [333; 506]. The architecture could be used to train MLIPs using rotationally invariant descriptors that can propagate directional information [506]. Architectures such as could also be used to train MLIPs that are even faster than the SchNet-based MLIPs [507] developed herein.

Finally, as was mentioned above, an unterminated substrate was used *in lieu* of an oxygen-terminated surface as the diamond (110) surface oxygenation state had not yet been characterised at the time of MLIP training. The presence of oxygen groups on the surface would most likely influence the shapes and sizes of surface-adsorbed NCs. Defects and dopants were also not considered within the MLIP, which were shown to promote the adsorption of single gold atoms in Chapter 4.

5.4 Conclusions

There is a great need for efficient methods that can be used to conduct accurate structure searches and explore the large configuration space of possible NC structures. While *ab initio* methods such as DFT are accurate, they are very computationally expensive and typically infeasible to be used as calculators in structure exploration methods. This chapter investigated the accuracy of two possible alternative methods, DFTB2 parameterisations and MLIPs, that could be used to conduct efficient structure searches.

First, the auorg and au_full DFTB2 parameterisations were benchmarked against DFT for two sets of systems: organic molecules on gold NCs, and gold NCs on diamond (110) surfaces. Both DFTB2 parameterisations were deemed to be unviable alternatives to PBE$^{+\text{MBD@rsSCS}}$ for these systems, and typically either

performed well structurally or energetically, but both. For this reason, a SchNet-based ML approach that could be combined with long-range dispersion schemes was developed and used to conduct structure searches of both small and large gold NCs on diamond (110) surfaces. By making use of adaptive sampling, the SchNet$^{+\text{MBD@rsSCS}}$ MLIPs were able to calculate radial atomic distributions close to the PBE$^{+\text{MBD@rsSCS}}$ optimum. The MLIPs were also shown to outperform the aforementioned DFTB2 parameterisations for a variety of NC sizes and were mostly able to identify (un)favourable NC geometries. While some disparities were observed with respect to DFT, it should be noted that the accuracy of the MLIPs could certainly be improved with further training and adaptive sampling. Nevertheless, the MLIPs developed in this chapter were shown to have a high degree of accuracy and were also significantly computationally cheaper than both DFT and DFTB2 approaches.

Having benchmarked the accuracy of SchNet$_{\text{adapt3}}$$^{+\text{MBD@rsSCS}}$, the MLIPs were subsequently used to analyse the adsorption of a large 147-atom gold NC onto a diamond (110) surface. Such large systems would be computationally with DFT but was able to be investigated using the ML approach developed in this chapter. Ordered isomers were calculated to be significantly more energetically favourable than the amorphous NC, with the cuboctahedral isomer found to be the most favourable geometry. This could be attributed to its lower sphericity, which means there is a greater interaction between the NC and the surface.

While there are still aspects that are neglected within the MLIPs, the ML-based approach presented in this chapter is of general utility for the computational surface science community and has the potential to drastically reduce the computational effort of some of the most common tasks in this field, and study the nucleation of both small and large NCs on surface substrates.

Chapter 6

Conclusion and Outlook

This thesis sought to characterise the adsorption of gold NCs at the interface between gold NCs, which have been shown to have many applications due to their unique optoelectronic properties, and the surfaces of polycrystalline BDD, which is an attractive substrate for metal electrodeposition due to its electrical conductivity and high stability. This would help to elucidate the initial stages of NC formation, as well as study the structures and stabilities of surface-adsorbed NCs.

However, this first required knowledge of the BDD surface oxygenation state, an understanding of which has been lacking to-datedue to the challenges associated with growing single-crystal diamond samples of (110) orientation [150; 155], which has been shown to be the dominant surface crystallography of CVD-grown polycrystalline BDD [22; 156]. This is in contrast to (111)- and (100)-oriented samples which, due to their slower growth rates [139; 150; 151], are much easier to grow and have thus been the focus of many single-crystal diamond studies. A joint computational-experimental study was conducted in Chapter 3 to identify the most stable oxygenation state of diamond (110) surfaces. Using *ab initio* thermodynamics [335; 373] based on DFT [65; 66], a phase diagram of the most stable oxygenic surface phases was established. Within the 0–1000 K temperature range, this phase diagram was found to be dominated by a highly stable phase of coexistent carbonyl and ether functional groups, while peroxide groups become more stable at low temperatures and high oxygen pressures. These findings were found to agree with experimental spectroscopic data in literature [152; 369; 370; 371], as well as to be robust with respect to the choices of DFA and dispersion correction. The high stability of the adjacent carbonyl and ether groups was found to arise from cooperative effects that mitigate surface deformation and satisfy all valencies. However, a lower-

coverage form of this phase was found to produce simulated core-level shifts that were in closer agreement with experimental XPS data collected after the removal of surface-adsorbed contaminants. This could indicate that the full-coverage phase may not be able to be realised due to kinetic hindrance and coverage limitations on realistic surfaces. The combined computational and experimental analyses outlined in Chapter 3 provide a much improved understanding of the oxygen-terminated diamond (110) surface, which has been lacking to-date, and showcase how the synergy between theory and experiment can be used to enrich the field.

Chapter 4 focused on identifying stabilisation mechanisms for single gold atoms on oxygen-terminated diamond (110) surfaces. This was done by using embedded cluster models within a hybrid QM/MM framework to investigate the adsorption and kinetic stability of single gold atoms. Such an approach allowed for the effects of local surface defects and dopants to be analysed. The fully-oxygenated surface is unlikely to stabilise single gold atoms due to the very weak adsorption interaction as well as the fairly low kinetic barriers associated with it, which is in agreement with other DFT calculations from literature [22]. The low kinetic barriers for the idealised surface are unlikely to be large enough to resist the highly energetic electron beam that is used in STEM experiments, this surface is unlikely to stabilise single gold atoms. This is supported by the significantly stronger adsorption energies and larger kinetic barriers of single gold atomthat were observed after the introduction of defects or dopants into the substrate surface. In the former case, the introduction of a SCOV into the surface was found to result in very high stability of the gold atom, though it should be noted that realistic surfaces would possess a myriad of coexistent defects and dopants. The strong adsorption and large kinetic barriers associated with the SCOV-defective surface can be attributed to the formation of a bond between the gold atom and a surface ether oxygen atom. This means that the high stability of single gold atoms on BDD observed by Hussein *et al.* [22] is most likely due to surface defects and dopants that were not seen within the STEM images or accounted for within their DFT calculations. The observed adsorption and kinetic trends were found to be robust with respect to the choices of embedding forcefield and dispersion correction. Adsorption energies were found to have some dependency on the choice of DFA. The PBE [229] GGA was generally found to perform well with respect to other GGAs, as well as higher-rung MGGAs and HGGAs, after the inclusion of long-range dispersion effects, for the calculation of adsorption energies and kinetic barriers on the idealised and SCOV-defective surfaces, while other DFAs such as the RPBE [233] GGA and TPSS [234] MGGA

were generally found to perform poorly for these systems. The agreement between PBE and higher-rung DFAs is of particular importance due to the lack of reliable experimental data that exists about adsorption structures and energetics on semi-conducting materials such as BDD. While the PBE [229] GGA was able to capture the adsorption of the single gold atom on delocalised triel-doped surface with respect to the idealised surface, some discrepancies were observed between PBE [229] and higher-rung DFAs, which means higher-rung HGGAs are perhaps more appropriate for this surface.

The SCOV-defective oxygen-terminated diamond (110) surface was then used in Section 4.3.5 to study the formation of surface-adsorbed gold dimers, trimers and tetramers by iteratively building NCs. The NC structure was found to be closer to its gas-phase optimum as the NC size increased, while the interaction between the NC and the surface was found to weaken . The ratio between the cohesion and adsorption energies indicated that the growth of gold NCs is preferentially determined by their interaction with the SCOV-defective surface rather than by the gold-gold interactions within the NCs. Preferred structural trends of NCs of a given atom count were also established, such as the preferential binding of a second gold to the first gold atom irrespective of its initial adsorption site, which resulted in a dimer structure tilted towards the surface. Here, the second gold atom was observed to have a strong adsorption energy as well as large diffusion barriers. For trimers, two stable structures were found depending on the initial site of the third atom: a near-equilateral triangular arrangement perpendicular to the surface, and a near-isosceles triangular arrangement that bridged over the surface. The number of structures however significantly increased upon the addition of a fourth gold atom. Of the seven tetramers observed, Y-shaped structures were generally found to be more stable than rhomboidal ones. Chapter 4 shows how the QM/MM framework can be used to study the adsorption and nucleation of single metal atoms and small metal NCs on surfaces whilst accounting for surface defects and dopants that are likely to, and have been shown to, stabilise adsorbates. The methodology used in Chapter 4 shows how computational approaches can be used to study the early stages of metal nucleation on an atomistic level whilst explicitly accounting for some conditions that are encountered in experiment such as surface heterogeneity and dopants.

However, the manual method used to build NCs in Chapter 4 is not efficient, and while it can be used to study NCs comprising only a few atoms, the number of

possible sites for an additional gold atom increases at a superlinear rate as the NC size increases. For this reason, Chapter 5 looked into methods that could be used alongside existing structure search algorithms [98; 99] to explore the configurational space of NC shapes and sizes, whilst retaining the accuracy of *ab initio* methods such as DFT at a fraction of their computational cost. Dispersion-corrected DFTB2 was determined to be an inaccurate alternative to dispersion-corrected DFT, which necessitated the development of an MLIP based on the SchNet [315; 333] NN trained on DFT data. By making use of adaptive sampling and training one set of models on vdW-free energies and forces, and another on Hirshfeld volume ratios for the calculation of a dispersion correction, the combined ML approach was able to calculate radial atomic distributions close to the DFT optimum. Furthermore, the MLIPs were shown to outperform the aforementioned DFTB2 parameterisations for a variety of NC sizes and enable accurate structure optimisations of supported gold NC at a fraction of the computational cost of DFT. The MLIPs were mostly able to identify (un)favourable NC geometries, though some disparities were observed with respect to DFT. However, it should be noted that the number of structures within the training dataset was very low, and the accuracy of the MLIPs could certainly be improved with further training and adaptive sampling. The MLIPs were finally used to analyse the structures and stabilities of three 147-atom gold NC isomers on diamond, systems that would be computationally intractable with DFT. Ordered isomers were calculated to be significantly more energetically favourable than the amorphous NC, with the cuboctahedral isomer found to be the most favourable geometry. This could be attributed to its lower sphericity, which means there is a greater interaction between the NC and the surface.

Overall, the aim of this thesis was to use computational approaches to investigate the nucleation of gold NCs on diamond surfaces. To this end, the initial stages of gold NC formation has been studied, which will aid in the development of atomistic nucleation theories along with the scientific effort in understanding the interaction between metal-diamond interfaces. This thesis will contribute to a variety of fields and can act as the foundation for future works.

That being the case, future works should aim to extend and complement the work presented in this thesis as there are still many existing factors that can further current understanding. If possible, all work in this thesis should be benchmarked against higher-level methods such as coupled cluster theory [508], Møller–Plesset perturbation theory [509] or the random-phase approximation [510; 511; 512] which,

despite being very computationally demanding, can aid in improving the accuracy of computational studies. The QM/MM framework used in Chapter 4 should be extended to investigate other defects such as interstitials, lattice vacancies and combinations thereof, as well as other dopants, to elucidate the best stabilisation mechanism for metal NC growth. The manual iterative preparation of gold NCs used in Chapter 4 could also be transformed into a much more efficient ML method similar to the G-SchNet architecture, which generates molecules in an autoregressive manner by placing one atom after another in 3D Euclidean space [513].

The vast multitude of experimental techniques that exist should also be used to validate the predictions made using any computational simulations, such as those in Chapters 4 and 5, as well as for any future work. Experimental methods such as temperature-programmed desorption and adsorption calorimetry have been used to evaluate the adsorption energies of metals on well-defined oxide surfaces [514; 515; 516; 517] and could be used to validate some of the findings within this thesis. The work in this thesis should also be used to complement experimental data: as was explained in Section 1.1.1, it is often difficult to assess the 3D structure of surface-adsorbed structures from experimental STEM images, which means there is an opportunity to use the MLIP outlined in Chapter 5 to augment existing STEM experiments of gold NCs on BDD surfaces [22] and improve the synergy between experiment and simulation.

Future atomistic simulations based on this work should also seek to replicate more realistic conditions and aim to account for experimental variables such as the electrochemical potential, the solvent and ions. The surfaces studied in this thesis were also modelled as vacuum-exposed rather than electrochemical solid-liquid interfaces. It would certainly be of great interest to see how the trends and results presented within this thesis would alter, if at all, after the inclusion of such electrochemical variables. To achieve this goal though, it should be noted that significant software developments would be required. This was briefly discussed in Chapter 4, where no infrastructure currently exists in the Py-ChemShell [301] software to conduct global structure searches with embedded cluster models. Another example is the lack of infrastructure to conduct grand-canonical QM/MM calculations, which would allow for the explicit inclusion of an electrochemical potential as well as local surface defects and dopants for covalent systems. While the GPAW [72; 73] software package has been used alongside the Atomic Simulation Environment [339] to run hybrid QM/MM calculations [518; 519; 520] and can be used to model more

realistic electrochemical environments [23; 24; 67; 521; 522; 523], it currently can not be used to run QM/MM calculations on covalently-bound systems such as BDD.

There are many open avenues and opportunities that remain, but this thesis has contributed to the development of modern atomistic theories of atom-by-atom particle growth. All being well, this thesis will play a useful part in solving any outstanding questions that remain, as well as help guide the controlled design of nanostructured catalysts in the future.

Appendix A

Codes and Scripts

A.1 Example Software Input Files

Listing A.4. Example `dftb_in.hsd` input file for DFTB+ calculations.

```
1   Geometry = GenFormat {
2       <<< "geometry.gen"
3   }
4   Driver = LBFGS{
5       AppendGeometries = YES
6       LatticeOpt = NO
7       MaxSteps = -1
8   }
9   Hamiltonian = DFTB{
10      Filling = Fermi{
11          Temperature [eV] = 0.1
12      }
13      MaxAngularMomentum = {
14          C = "p"
15          Au = "d"
16      }
17      SpinPolarisation = {
18      }
19      Mixer = Broyden{
20          MixingParameter = 0.01
21          InverseJacobiWeight = 0.01
22          MinimalWeight = 1.0
23          MaximalWeight = 1E5
24          WeightFactor = 1E-2
25      }
26      ShellResolvedSCC = YES
27      SCC = YES
```

```
28      SCCTolerance = 1E-6
29      MaxSCCIterations = 3000
30      SlaterKosterFiles = Type2FileNames{
31          Prefix = "[path to/auorg-1-1]"
32          Separator = "-"
33          Suffix = ".skf"
34      }
35  }
36  Analysis{
37      CalculateForces = YES
38  }
```

Listing A.5. Example `control.in` input file for FHI-aims calculations.

```
1   xc                     pbe
2   occupation_type        gaussian 0.1
3   relativistic           atomic_zora scalar
4   spin                   none    # Can also be 'collinear'
5   default_initial_moment 0
6
7   # Self-consistency settings
8   sc_accuracy_etot       1E-06
9   sc_accuracy_eev        1E-02
10  sc_accuracy_rho        1E-05
11
12  # For geometry optimisations
13  sc_accuracy_forces     1E-04
14  relax_geometry trm     1E-02
15
16  # For periodic surface calculations
17  k_grid                 16 16 1
18  use_dipole_correction
19
20  # 'tight' basis set for carbon
21
22  ################################################################################
23  #
24  #  FHI-aims code project
25  # Volker Blum, Fritz Haber Institute Berlin, 2009
26  #
27  #  Suggested "tight" defaults for C atom (to be pasted into control.in file)
28  #
29  ################################################################################
30    species        C
31  #     global species definitions
32      nucleus            6
33      mass               12.0107
34  #
35      l_hartree          6
36  #
37      cut_pot            4.0  2.0  1.0
38      basis_dep_cutoff   1e-4
39  #
40      radial_base        34 7.0
41      radial_multiplier  2
42      angular_grids specified
43        division  0.2187    50
```

```
44        division    0.4416   110
45        division    0.6335   194
46        division    0.7727   302
47        division    0.8772   434
48  #     division    0.9334   590
49  #     division    0.9924   770
50  #     division    1.0230   974
51  #     division    1.5020  1202
52  #    outer_grid   974
53       outer_grid   434
54  ################################################################################
55  #
56  #  Definition of "minimal" basis
57  #
58  ################################################################################
59  #     valence basis states
60       valence      2  s   2.
61       valence      2  p   2.
62  #     ion occupancy
63       ion_occ      2  s   1.
64       ion_occ      2  p   1.
65  ################################################################################
66  #
67  #  Suggested additional basis functions. For production calculations,
68  #  uncomment them one after another (the most important basis functions are
69  #  listed first).
70  #
71  #  Constructed for dimers: 1.0 A, 1.25 A, 1.5 A, 2.0 A, 3.0 A
72  #
73  ################################################################################
74  #  "First tier" - improvements: -1214.57 meV to -155.61 meV
75       hydro 2 p 1.7
76       hydro 3 d 6
77       hydro 2 s 4.9
78  #  "Second tier" - improvements: -67.75 meV to -5.23 meV
79       hydro 4 f 9.8
80       hydro 3 p 5.2
81       hydro 3 s 4.3
82       hydro 5 g 14.4
83       hydro 3 d 6.2
84  #  "Third tier" - improvements: -2.43 meV to -0.60 meV
85  #     hydro 2 p 5.6
86  #     hydro 2 s 1.4
87  #     hydro 3 d 4.9
88  #     hydro 4 f 11.2
89  #  "Fourth tier" - improvements: -0.39 meV to -0.18 meV
```

```
90  #      hydro 2 p 2.1
91  #      hydro 5 g 16.4
92  #      hydro 4 d 13.2
93  #      hydro 3 s 13.6
94  #      hydro 4 f 17.6
95  #  Further basis functions - improvements: -0.08 meV and below
96  #      hydro 3 s 2
97  #      hydro 3 p 6
98  #      hydro 4 d 20
99  ###############################################################################
100 #
101 # For methods that use the localized form of the "resolution of identity" for
102 # the two-electron Coulomb operator (RI_method LVL), particularly Hartree-Fock and
103 # hybrid density functional calculations, the highest accuracy can be obtained by
104 # uncommenting the line beginning with "for_aux"  below, thus adding an extra g
    ↪  radial
105 # function to the construction of the product basis set for the expansion.
106 # See Ref. New J. Phys. 17, 093020 (2015) for more information, particularly Figs. 1
    ↪   and 6.
107 #
108 ###############################################################################
109 #
110 # for_aux hydro 5 g 6.0
```

A.2 Py-ChemShell Source Code Changes

Listing A.6. The amended `getGeomBuff()` Python function used to parse in initial spin moments and initial charges within the `chemsh/interfaces/FHIaims/__init__.py` file in the 21.0 release of Py-ChemShell. Amendments to the file by the author are on lines 131-142.

```python
121     def getGeomBuff(self):
122         '''String buff containing the geometry information'''
123
124         from numpy.core import defchararray, full
125
126         strbuff = '#FHI-aims ChemShell calculation'+ '\n'
127         # Print atoms
128         strbuff_atoms = self.frag.coords2str(coords=9, atoms='names', order='ca',
        ↪  indent=' ', delimiter=' ',
129                                             inserts=[(0,'atom')], unit='angstrom')
130
131         # Enumerate as index i directly corresponds to atom number
132         for i, line in enumerate(strbuff_atoms.splitlines()):
133             line += '\n'
134             if line.split()[4] in self.initial_moment:
135                 line += 'initial_moment ' +
            ↪  str(self.initial_moment[line.split()[4]]) + '\n'
136             if i in self.initial_moment:
137                 line += 'initial_moment ' + str(self.initial_moment[i]) + '\n'
138             if line.split()[4] in self.initial_charge:
139                 line += 'initial_charge ' +
            ↪  str(self.initial_charge[line.split()[4]]) + '\n'
140             if i in self.initial_charge:
141                 line += 'initial_charge ' + str(self.initial_charge[i]) + '\n'
142             strbuff += line
143
144         # Print ECPs. Only occurs if doing QMMM, else self.ecp is undefined.
145         if self.ecps:
146             self._ecprange = self.frag.bqs.getRegion(2)
147             # prefix with 'bq_' and suffix with '_e'
148             ecpnames = defchararray.add(full((self.ecps.natoms,), 'bq_',
            ↪  dtype='S3'), self.ecps.names)
149             ecpnames = defchararray.add(ecpnames, full((self.ecps.natoms,), '_e',
            ↪  dtype='S2'))
150             strbuff += self.ecps.coords2str(coords=9, atoms=None, order='ca',
            ↪  indent=' ', delimiter=' ',
151                                             inserts=[(0,'pseudocore'),(4,ecpnames)],
                                            ↪  unit='angstrom')
152
```

```python
153          # Print BQs
154      if self.frag.nbqs:
155          multipole_str = 'multipole'
156          multipole_order = '0'
157          atomstr = 'empty'
158          self._ghost_list = []
159
160          # nbqs is initialised to zero so no harm in just setting this for loop
        ↪   to run over this variable
161          for ibq in range(self.frag.nbqs):
162
163              # Check BQ is not an ECP centre. If no ECPs, this list is empty
164              if ibq not in self._ecprange:
165                  # AJL, March 2019: Get BQ coords, convert to Angstrom
166                  # Made this an if statement so I can match settings for the TCL
                ↪   version of ChemShell (giving Bohr2Angstrom) and validate
                ↪   setup
167                  coords = [ (float(coord)*unitconvert.Bohr2Angstrom) for coord in
                ↪   list(self.frag.bqs.coords[ibq]) ]
168                  strbuff = ''.join([strbuff, '%10s    %16.9f    %16.9f    %16.9f
                ↪   %2s   %12.8f\n'%(multipole_str, coords[0], coords[1],
                ↪   coords[2],
169                                  multipole_order,
                                ↪   self.frag.bqs.charges[ibq])])
170
171                  # AJL, April 2019: Check to see if we need an empty site on this
                ↪   BQ.
172                  # In short, this is necessary as it removes the integration
                ↪   grids from being right on top of the point charge during
                ↪   the
173                  # numerical integrations. Without it, you get spurious,
                ↪   uncontrollable results with singularities if the BQs are
                ↪   close to QM atoms
174                  # The default ghost_cutoff is set to 5 angstrom, so by default
                ↪   this check is made. First we look at atom-bq interactions
175                  if self.frag.selectByRadius(radius=self.ghost_max_cutoff *
                ↪   unitconvert.Angstrom2Bohr,
                ↪   centre=list(self.frag.bqs.coords[ibq])).size > 0:
176
177                      if self.ghost_species:
178                          # Now we need to check for bq-bq interactions. We cannot
                        ↪   put empty sites on top of each other so if we have
179                          # BQs very close together (0.001 au) then we only add an
                        ↪   empty site over the first instance
```

```python
180                             ghost_close_neighbours =
      ↪    self.frag.bqs.selectByRadius(radius=self.ghost_min_cutoff
      ↪    * unitconvert.Angstrom2Bohr,
      ↪    centre=list(self.frag.bqs.coords[ibq]))
181
182                             if len(ghost_close_neighbours) == 1 or any(jbq > ibq for
      ↪    jbq in ghost_close_neighbours):
183                                 #print('**** Adding an integration grid to protect
      ↪    from QM-BQ energy singularities')
184                                 strbuff = ''.join([strbuff, '%10s    %16.9f
      ↪    %16.9f    %16.9f    %7s\n'%(atomstr, coords[0],
      ↪    coords[1], coords[2],
185                                             self.ghost_species)])
186                                 #Need to keep count of the empty sites we add so
      ↪    that we can include gradients information
      ↪    post-calculation
187                                 self._ghost_list.append(ibq)
188
189                         else:
190                             # I've made this verbose so the user understands the
      ↪    error
191                             print('****')
192                             print('**** A QM and BQ site have been identified in
      ↪    close proximity. Due to the numerical')
193                             print('**** integrations used in FHI-aims, an "empty"
      ↪    site **must** be placed on the BQ to protect')
194                             print('**** against Coulomb singularities. Please define
      ↪    the "empty" site species (ghost_species)')
195                             print('**** Note: this error can be silenced by setting
      ↪    ghost_max_cutoff = 0.')
196                             print('****')
197                             # I'd like a hard stop here but not sure this works as
      ↪    desired?
198                             # Someone should check the coding for me as I can't find
      ↪    a good example in the codebase
199                             #errors.ChemShModuleError('ghost_species must have a
      ↪    species name for the "empty" sites')
200                             assert false, 'Ghost_species must have a species name
      ↪    for the empty sites. Check control.in and update
      ↪    input'
201
202         return strbuff
```

Listing A.7. The amended `chemsh/dl_find/opt.py` Python file within Py-ChemShell used to parse in a `frozen_partial` mask. Corresponding line numbers are also included. Amendments to the file by the author are on lines 58 and 308–338.

```python
1    #  Copyright (C) 2017 The authors of Py-ChemShell
2    #
3    #  This file is part of Py-ChemShell.
4    #
5    #  Py-ChemShell is free software: you can redistribute it and/or modify
6    #  it under the terms of the GNU Lesser General Public License as
7    #  published by the Free Software Foundation, either version 3 of the
8    #  License, or (at your option) any later version.
9    #
10   #  Py-ChemShell is distributed in the hope that it will be useful,
11   #  but WITHOUT ANY WARRANTY; without even the implied warranty of
12   #  MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
13   #  GNU Lesser General Public License for more details.
14   #
15   #  You should have received a copy of the GNU Lesser General Public
16   #  License along with Py-ChemShell.  If not, see
17   #  <http://www.gnu.org/licenses/>.
18
19   from numpy    import asarray, full, zeros, float64, int64
20   from .dl_find import DL_FIND # this is the DL_FIND-like task class
21   from ..       import dl_find # the DL_FIND module
22   from ..utils  import objutils
23
24   # now we only got DL_FIND but will have more
25   # we provide users some synonyms
26   _synons = { 'dl_find': dl_find,
27               'dl-find': dl_find,
28               'dlfind' : dl_find,
29               'dlf'    : dl_find }
30
31   class Opt(DL_FIND):
32       '''Geometry Optimisation'''
33
34       _attrs = {
35                   'active'            :[],
36                   'algorithm'         :'lbfgs',
37   #                'buff'              : asarray([[1.0,2.0,3.0]]),
38                   'callback'          : lambda x:999,
39                   'carthessian'       : False,
40                   'cell_only'         : False,
41                   'coords2'           : zeros(shape=(0,3), dtype=float64),
```

171

```
42              'conint'         :'no',
43              'coordinates'    :'cartesian',
44              'coupled_states' : True,
45              'cycle'          : 0,
46              'delta'          : 0.01,
47              'derivatives'    : zeros((1,), dtype=float),
48              'dump'           : 5,
49              'dim'            :'2D',
50              'dimer'          : False,
51              'dimer_interpolate': True,
52              'distort'        : 0.0,
53              'frag2'          : None,
54              'fric0'          : 0.3,
55              'fricfac'        : 0.95,
56              'fricp'          : 0.3,
57              'frozen'         : full((1), -1, dtype=int64),
58              'frozen_partial' : [],
59              'genetic'        : False,
60              'gp_c3'          : 1.0,
61              'gp_c4'          : 0.9,
62              'icoord'         :-1,
63              'include_res'    : False,
64              'inithessian'    :-1,
65              'initial_hessian' :'external',
66              'inner_atoms'    : None,
67              'inner_residues' : None,
68              'iopt'           :-1,
69              'lbfgs_mem'      :-1,
70              'list_option'    :'medium',
71              'ln_t1'          : 0.0001,
72              'ln_t2'          : 1.0,
73              'mass'           : zeros(shape=(0,), dtype=float64),
74              'maxcycle'       : 100,
75              'maxene'         : 100000,
76              'maxmicrocycle'  : 100,
77              'maxrot'         : 10,
78              'maxstep'        : 0.5,
79              'maxupdate'      : 50,
80              'microiterative' : False,
81              'micro_esp_fit'  : False,
82              'minstep'        : 0.00001,
83              'neb'            :'no',
84              'nebk'           : 0.01,
85              'neb_cart'       : False,
86              'neb_climb_test' : 3.0,
87              'neb_freeze_test' : 1.0,
```

```
 88              'neb_path_sampling': -1.0,
 89              'nframes'          : 0,
 90              'nimages'          :-1,
 91              'nzero'            :-1,
 92              'optimiser'        :'dl_find',
 93              'parameterise'     : False,
 94              'parameters'       :[],
 95              'pf_c1'            : 5.0,
 96              'pf_c2'            : 5.0,
 97              'po_contraction'   : 0.9,
 98              'po_death_rate'    : 0.5,
 99              'po_distribution'  :'force_bias',
100              'po_init_pop_size' : 50,
101              'po_maxcycle'      : 10000,
102              'po_mutation_rate' : 0.15,
103              'po_nsave'         : 10,
104              'po_pop_size'      : 25,
105              'po_radius'        : 1.0,
106              'po_reset'         : 500,
107              'po_scalefac'      : 10.0,
108              'po_tolerance_g'   : 0.001,
109              'po_tolerance_r'   : 1.0E-8,
110              'qts'              : False,
111              'qtsrate'          : False,
112              'rate'             : False,
113              'restart'          : False,
114 # attribute name taken by `self.result`
115 #               'result'           :'dl_find.result',
116              'result_file'      :'dlf_result.pun',
117              'save_path'        : False,
118              'scale_step'       : 1.0,
119              'setParameters'    : lambda x:999,
120              'soft'             : 1.0E20,
121              'state_i'          : 1,
122              'state_j'          : 2,
123              'stochastic'       : False,
124              'thermal'          : False,
125              'timestep'         : 1.0,
126              'tolerance'        : -1.0,
127              'tolerance_e'      : -1.0,
128              'tolrot'           : 5.0,
129              'tsplit'           : False,
130              'tsrelative'       : False,
131              'trust_radius'     :'constant',
132              'update'           :-1,
133              'update_method'    :'bofill',
```

```python
134                    'weights'         : None
135                }
136
137        _internals = {
138                    'coords2'       : zeros(shape=(0,3), dtype=float64),
139                    '_iter'         : 0,
140                    '_task'         :'standard',
141                    '_convergences': zeros(shape=(0,5), dtype=float64), # convergence
                    ↪   information: energy, max grad, max step, RMS grad, RMS step
142                }
143
144        _synons = { 'coordstype': 'coordinates',
145                    'coordtype' : 'coordinates',
146                }
147
148        _priorities = 'theory', 'cell_only', 'dim', 'parameters'
149
150
151        @property
152        def frag2(self):
153            '''Auxiliary second fragment (needed by, e.g., NEB)'''
154
155            return self._frag2
156
157        @frag2.setter
158        def frag2(self, val):
159            '''Setter of frag2'''
160
161            try:
162                self.coords2 = val.coords
163                self._frag2 = val
164                self.nframes = 1
165            except:
166                from ..objects import fragment
167                merged = fragment.Fragment()
168                merged.merge(*val, noself=True)
169                self._frag2 = merged
170                self.coords2 = merged.coords
171                self.nframes = len(val)
172
173
174        @property
175        def neb(self):
176            '''NEB method'''
177
178            # set default when frag2 given
```

```python
179            if not self.frag2 and self._neb != 'no':
180                self._neb = 'no'
181                print(" ChemShell >>> NEB turned off since frag2 not given")
182
183            if self.frag2 and self._neb != 'no':
184                self._neb = 'frozen'
185
186            return self._neb
187
188        @neb.setter
189        def neb(self, val):
190            '''Setter of NEB method'''
191
192            # force to switch off NEB
193            if val.lower() == 'no':
194                self.frag2 = None
195
196            self._neb = val.lower()
197
198        @property
199        def coords2(self):
200            '''Second coords'''
201
202            return self._coords2
203
204        @coords2.setter
205        def coords2(self, val):
206            '''Setter of second coords'''
207            from ..base import nparray
208
209  # TODO
210  #          nparray.setField(self._coords2, val)
211            self._coords2 = asarray(val).flatten()
212
213
214        @property
215        def neb_path_sampling(self):
216            '''Return the sampling density of images along the reaction path'''
217
218  # AJL Aug 2020
219  # Unclear if this should be the private or public equivalent of this variable?
220  # There isn't a clear guide as to what is used where - e.g. coords2 and _coords2 are
221  #   used above, as are neb/_neb
222            return self._neb_path_sampling
223
224        @neb_path_sampling.setter
```

```python
224         def neb_path_sampling(self, val):
225             '''Setter for NEB path sampling - will also then work out a value for
                ↪   nimages'''
226
227             self._neb_path_sampling = val
228

229     # It would make more sense to me if this property was a redirection of
        ↪   self.theory.frag,
230     # rather than using different calls for each sets of coords, but I presume there is
        ↪   a reason this is absent?
231     #       print(self._coords)
232     #       print(self.theory.frag.coords)
233     #       print(self._coords2.reshape(self.theory.frag.coords.shape))
234     #
        ↪   print(self.theory.frag.coords-self._coords2.reshape(self.theory.frag.coords.shape))
235
236             from numpy.linalg import norm
237             # None of this needs storing - it can be broken up if you want the variables
                ↪   to be more readable
238             linear_interpolation_path_distance =
                ↪   norm(self.theory.frag.coords-self._coords2.reshape(self.theory.frag.coords.shape))
239             self.nimages =
                ↪   int(linear_interpolation_path_distance/self.neb_path_sampling)
240
241         @property
242         def nimages(self):
243             '''Return the number of images along the reaction trajectory'''
244
245             if not self.frag2 and self._neb != 'no':
246                 if self._nimages == -1:
247                     self.neb_path_sampling = 0.5 # A very conservative default value
248
249             return self._nimages
250
251         @nimages.setter
252         def nimages(self, val):
253             '''Set the number of images, as this value is hidden by default'''
254
255             self._nimages = val
256
257         @property
258         def active(self):
259             '''List of active centres (complementary to frozen)'''
260
261             return self._active
262
```

```python
263
264        @active.setter
265        def active(self, val):
266            '''Setter of active'''
267
268            from numpy import arange, ones
269
270            self._active = asarray(self.theory.frag.getRange(val))
271
272            # only if val is non-empty
273            if self._active.size:
274                self.theory.frag.frozen[:] = 0
275                mask = ones(self.theory.frag.frozen.shape, dtype=bool)
276                try:
277                    mask[self._active] = False
278                except IndexError as e:
279                    print(e)
280                    print("\n >>> ERROR: it's possible theory.frag hasn't been
                        ↪  assigned.\n")
281                self.theory.frag.frozen[mask] = -1
282                self._frozen = arange(self.theory.frag.natoms)[mask]
283
284
285        @property
286        def frozen(self):
287            '''List of frozen centres (complementary to active)'''
288
289            return self._frozen
290
291
292        @frozen.setter
293        def frozen(self, val):
294            '''Setter of frozen'''
295
296            from numpy import arange, in1d, ones
297
298            self._frozen = asarray(self.theory.frag.getRange(val))
299
300            # only if val is non-emtpy
301            if self._frozen.size:
302                self.theory.frag.frozen[:] = 0
303                self.theory.frag.frozen[self._frozen] = -1
304                mask = ones(self.theory.frag.frozen.shape, dtype=bool)
305                mask[self._frozen] = False
306                self._active = arange(self.theory.frag.natoms)[mask]
307
```

```python
308         @property
309         def frozen_partial(self):
310
311             return self._frozen_partial
312
313         @frozen_partial.setter
314         def frozen_partial(self, val):
315             '''val should be entered as a list of tuples e.g. frozen_partial = [(A,
                 ax),...], where A is either
316             an integer corresponding to the atom number (e.g. 1) or a string of the atom
     label (e.g. 'C'), and
317             ax is a string containing the axes that should be constrained (e.g. 'xy')
     for atom A'''
318
319             from numpy import ones
320
321             for entry in val:
322                 if isinstance(entry[0], str):
323                     indices = [i for i, name in enumerate(self.theory.frag.names) if
                         name==bytes(entry[0].encode())]
324                     for index in indices:
325                         val.append((index, entry[1]))
326                     val.remove(entry)
327             val2 = [(A, ''.join(sorted(ax)).lower()) for A, ax in val]
328
329             self._frozen_partial = ones((self.theory.frag.natoms, 3), dtype=float)
330             for i in range(self.theory.frag.natoms):
331                 for entry in val2:
332                     if i == entry[0]:
333                         if 'x' in entry[1]:
334                             self._frozen_partial[i, 0] = 0
335                         if 'y' in entry[1]:
336                             self._frozen_partial[i, 1] = 0
337                         if 'z' in entry[1]:
338                             self._frozen_partial[i, 2] = 0
339
340
341         @property
342         def lbfgs_mem(self):
343             ''''''
344
345             return self._lbfgs_mem
346
347         @lbfgs_mem.setter
348         def lbfgs_mem(self, val):
349             ''''''
```

```python
350
351          self._lbfgs_mem = val
352
353
354      @property
355      def tolerance_e(self):
356          '''Convergence criterion for the energy change (a.u.)'''
357
358          if self._tolerance_e < 0:
359              return self.tolerance/450.0
360          else:
361              return self._tolerance_e
362
363      @tolerance_e.setter
364      def tolerance_e(self, val):
365          '''Setter of tolerance_e'''
366
367          self._tolerance_e = val
368
369
370      @property
371      def nframes(self):
372          '''Number of frames'''
373
374          if self._nframes > 100:
375              print(" >>> ERROR: nframes cannot be greater than 100")
376              quit(1)
377          return self._nframes
378
379      @nframes.setter
380      def nframes(self, val):
381          '''Setter of nframes'''
382
383          self._nframes = val
384
385
386      @property
387      def parameterise(self):
388          '''Parameterisation mode'''
389
390          return self.cell_only or len(self.parameters) > 0
391
392
393      @property
394      def parameters(self):
395          '''Variables to parameterise'''
```

```python
396
397          from numpy import array, copy
398
399          if self.cell_only:
400              self._task = 'cell_only'
401              if self.dim.upper().strip() in [ '2D' ]:
402                  self.theory.dim = '2D'
403                  if len(self._parameters) != 3:
404                      self._parameters = array([0.0,0.0,0.0])
405                      self.derivatives.resize(3, refcheck=False)
406              elif self.dim.upper().strip() in [ '3D' ]:
407                  self.theory.dim = '3D'
408                  if len(self._parameters) != 6:
409                      self._parameters = array([0.0,0.0,0.0,0.0,0.0,0.0])
410                      self.derivatives.resize(6, refcheck=False)
411              else:
412                  print("\n ChemShell ERROR: we currently only support 2D and 3D
                     ↪   cell-only optimisation\n")
413                  exit(112)
414
415          return self._parameters
416
417
418      @parameters.setter
419      def parameters(self, val):
420          '''Setter of parameters'''
421
422          from ..utils import iterutils
423
424          # flatten
425          self._parameters = iterutils.getFlattenedIter(val, sort=False)
426
427
428      @property
429      def thermal(self):
430          '''Hessian and thermal corrections only'''
431
432          return self._thermal
433
434
435      @thermal.setter
436      def thermal(self, val):
437          '''Setter of thermal'''
438
439          self._thermal = bool(val)
440          if self._thermal:
```

```python
441                self.algorithm = 'thermal'
442
443
444        def run(self, dryrun=False, info_filename='_profile_opt.npz', **kwargs):
445            '''Run geometry optimization'''
446
447            from sys      import stdout
448            from numpy    import copy, savez_compressed, vstack
449            from ..base   import run
450            from ..utils  import strutils
451
452            # YL: cell-only optimisation
453            if self.cell_only:
454                # compute the strains
455                self.theory._strains = True
456                # save the original vectors: the strains being optimised are targeting
                 ↪   the original vectors!
457                self._vectors0 = copy(self.theory.frag.cell.vectors)
458
459            # make sure gradients are always computed for optimisation
460            self.theory._gradients = True
461
462            # if path can be imported as module call run.runLib(), otherwise
             ↪   run.runExec() (see <dict> run.run().runTypes)
463            optimiser = strutils.importAsModule(_synons[self.optimiser].__name__)
464
465            # YL 26/01/2021: it's safe to cite the reference because callback funtions
             ↪   are not invoked unless during an optimisation
466            optimiser.callback._taskObj      = self
467            optimiser.callback._dryrun       = dryrun
468            optimiser.callback._info_filename  = info_filename
469            optimiser.setParameters._taskObj   = self
470
471            # YL 26/01/2021: handles of functions
472            self.callback      = optimiser.callback
473            self.setParameters = optimiser.setParameters
474
475            # run
476            optimiser.run(self.theory, self.options, **kwargs)
477
478            # report final results
479            if self.cell_only:
480                _printFinalCell(self)
481
482            # convergence information: max_gradient, max_step, RMS_gradient, RMS_step
```

```python
483          self._convergences = vstack((self._convergences,
    ↪   self.theory._result.convergences.T))
484          # dump the information of convergence path to a compressed NumPy .npz file
485          # we do it every step in case the optimisation job fails
486          savez_compressed(info_filename,
487                           energies      = self._convergences[:,0],
488                           max_gradient  = self._convergences[:,1],
489                           max_step      = self._convergences[:,2],
490                           RMS_gradient  = self._convergences[:,3],
491                           RMS_step      = self._convergences[:,4]
492                          )

494          print()
495          stdout.flush()

497          # switch off strains by default
498          self.theory._strains = False

500          return self


503  # aliases
504  Optimisation = Opt
505  Optimization = Opt
506  Optimise     = Opt
507  Optimize     = Opt


510  def _printFinalCell(_self):
511      '''Print the final results of cell-only optimisation'''

513      if _self.dim.upper().strip() in [ '2D' ]:
514          print("\n Optimised 2D cell vectors (a.u.):\n")
515          print(" X {:>15.9f} {:>15.9f}".format(*_self.theory.frag.cell.vectors[0]))
516          print(" Y {:>15.9f} {:>15.9f}".format(*_self.theory.frag.cell.vectors[1]))
517          print("\n Original 2D cell vectors (a.u.):\n")
518          print(" X {:>15.9f} {:>15.9f}".format(*_self._vectors0[0]))
519          print(" Y {:>15.9f} {:>15.9f}".format(*_self._vectors0[1]))
520          print("\n Strain tensor matrix applied to the original cell:\n")
521          print(" {:>15.9f} {:>15.9f}".format(*_self.theory.frag.cell.strains[0]))
522          print(" {:>15.9f} {:>15.9f}".format(*_self.theory.frag.cell.strains[1]))
523      elif _self.dim.upper().strip() in [ '3D' ]:
524          print("\n Optimised 3D cell vectors (a.u.):\n")
525          print(" X {:>15.9f} {:>15.9f}
    ↪   {:>15.9f}".format(*_self.theory.frag.cell.vectors[0]))
```

```
526        print(" Y {:>15.9f} {:>15.9f}
    ↪    {:>15.9f}".format(*_self.theory.frag.cell.vectors[1]))
527        print(" Z {:>15.9f} {:>15.9f}
    ↪    {:>15.9f}".format(*_self.theory.frag.cell.vectors[2]))
528        print("\n Original 3D cell vectors (a.u.):\n")
529        print(" X {:>15.9f} {:>15.9f} {:>15.9f}".format(*_self._vectors0[0]))
530        print(" Y {:>15.9f} {:>15.9f} {:>15.9f}".format(*_self._vectors0[1]))
531        print(" Z {:>15.9f} {:>15.9f} {:>15.9f}".format(*_self._vectors0[2]))
532        print("\n Strain tensor matrix applied to the original cell:\n")
533        print(" {:>15.9f} {:>15.9f}
    ↪    {:>15.9f}".format(*_self.theory.frag.cell.strains[0]))
534        print(" {:>15.9f} {:>15.9f}
    ↪    {:>15.9f}".format(*_self.theory.frag.cell.strains[1]))
535        print(" {:>15.9f} {:>15.9f}
    ↪    {:>15.9f}".format(*_self.theory.frag.cell.strains[2]))
536        print("\n Final stress tensor matrix (GPa):\n")
537        print("              X              Y              Z")
538        print(" X {:>15.9f} {:>15.9f}
    ↪    {:>15.9f}".format(*_self.theory.frag.cell.stress_tensor[0]))
539        print(" Y {:>15.9f} {:>15.9f}
    ↪    {:>15.9f}".format(*_self.theory.frag.cell.stress_tensor[1]))
540        print(" Z {:>15.9f} {:>15.9f}
    ↪    {:>15.9f}".format(*_self.theory.frag.cell.stress_tensor[2]))
```

Listing A.8. The amended `dlf_get_gradient()` Fortran subroutine used to run constrained optimisations, which was used within the `chemsh/dl_find/dlf_routines.f90` file in Py-ChemShell. Corresponding line numbers are also included. New additions to the subroutine are on lines 575, 643, 647, 668–669 and 682.

```fortran
552    subroutine dlf_get_gradient(nvar,      &
553                               coords,    &
554                               energy,    &
555                               gradients, &
556                               iimage,    &
557                               kiter,     &
558                               status)
559
560        use iso_c_binding
561        use dlf_parameter_module, only: rk
562        use ChemShellDLFModule   , only: PyFrag, PyResult, PyCallback, PyOptions,
         ↪   PySetParams, ierror
563        use dlf_convergence      , only: valg, valrmsg, vals, valrmss
564
565        implicit none
566        integer          , intent(in)  :: nvar
567        real(rk), target , intent(in)  :: coords(nvar)
568        real(rk)         , intent(out) :: energy
569        real(rk)         , intent(out) :: gradients(nvar)
570        integer          , intent(in)  :: iimage
571        integer          , intent(in)  :: kiter
572        integer          , intent(out) :: status
573
574        integer                        :: i, debug, natoms, iparameterise
575        real(rk), pointer              :: onedimdbl(:), twodimdbl(:,:),
         ↪   gradients_frozen(:,:)
576        real(rk), pointer              :: temp(:)
577        integer(kind=8), pointer       :: onedimint(:)
578
579        real :: start_time, end_time
580
581        logical                        :: parameterise
582
583        status = 0
584
585        call PyOptions%get("debug", debug)
586
587        call PyOptions%get("parameterise", iparameterise)
588        parameterise = iparameterise.eq.1
```

184

```fortran
589
590        ! atomic mode
591        if(.not.parameterise) then
592
593            call PyFrag%get('npts' , natoms)
594
595            allocate(onedimint(natoms))
596            call PyFrag%get("znums", onedimint)
597
598            if(debug > 3) then
599                allocate(twodimdbl(3,natoms))
600                call PyFrag%get('xyz', twodimdbl)
601                write(*,'(/" DL_FIND original coords (a.u.):"/)')
602                write(*,'(3X,"index",1X,"Z-number",12X,"X",15X,"Y",15X,"Z"/)')
603                do i = 1, natoms
604                    write(*,'(1X,I6,3X,I6,3X,3F16.9)') i-1, onedimint(i), twodimdbl(:,i)
605                enddo
606                call flush(6)
607                deallocate(twodimdbl)
608            endif
609
610            ! update coords
611            call PyFrag%set('xyz', reshape(coords, (/3,natoms/)))
612
613            ! CALLBACK
614            if(debug > 0) then
615                write(*,'(/" >>> DL_FIND module is calling callback function for
    ↪   gradients evaluation..."/)')
616            endif
617
618            call cpu_time(start_time)
619
620            ! energy of the previous step
621            allocate(onedimdbl(1))
622            call PyResult%get('energy', onedimdbl)
623            energy = onedimdbl(1)
624            ! save the energy, max grad, max step, RMS grad, RMS step
625            call PyResult%set('convergences', (/energy,valg,vals,valrmsg,valrmss/))
626
627            ! do calculations on the Python side
628            ierror = PyCallback()
629            call flush(5)
630            call flush(6)
631            if(ierror.ne.0) then
632    !           print *, ">>> ERROR: DL_FIND failed to run callback function.
        ↪   Exiting..."
```

```fortran
633                 call exit(ierror)
634             endif
635             call cpu_time(end_time)
636             if (debug > 4) then
637                 print *, ">>> DL_FIND: PyCallback() costs ", end_time - start_time, "
    ↪  seconds"
638             endif
639             ! END OF CALLBACK
640
641             ! note the shape!
642             allocate(twodimdbl(3,natoms))
643             allocate(gradients_frozen(3,natoms))
644
645             call PyResult%get('energy'   , onedimdbl)
646             call PyResult%get('gradients', twodimdbl)
647             call PyOptions%get('frozen_partial', gradients_frozen)
648
649             ! update energy and gradients
650             ! note: energy in Python is an array!
651             energy = onedimdbl(1)
652
653             ! can use c_f_pointer to reshape without copying, too
654             ! call c_f_pointer(c_loc(twodimdbl), onedimdbl, [nvar])
655             ! gradients = onedimdbl(:)
656
657             if(debug > 3) then
658                 write(*,'(/" DL_FIND energy (a.u.) of updated coordinates:"/)')
659                 write(*,'(4XF16.9)') energy
660             endif
661
662             if(debug > 3) then
663                 write(*,'(/" DL_FIND gradients (a.u.) of updated coordinates:"/)')
664                 write(*,'(3X,"index",1X,"Z-number",12X,"X",15X,"Y",15X,"Z"/)')
665             endif
666
667             do i = 1, natoms
668                 gradients(i*3-2:i*3) = twodimdbl(:,i)*gradients_frozen(:,i)
669                 write(*,*) i, gradients(i*3-2:i*3)
670                 if(debug > 3) then
671                     ! atomic numbers
672                     write(*,'(1X,I6,3X,I6,3X,3F16.9)') i-1, onedimint(i),
                        ↪  gradients(i*3-2:i*3)
673                 endif
674             enddo
675
676             if(debug > 3) then
```

186

```fortran
677              write(*,'(//)')
678          endif
679          call flush(6)
680
681          deallocate(onedimdbl, twodimdbl)
682          deallocate(onedimint, gradients_frozen)
683
684      ! parameterisation mode
685      else
686
687          allocate(onedimdbl(nvar))
688          call PyOptions%get('parameters', onedimdbl)
689          write(*,'(/" DL_FIND original parameters:"/)')
690          do i = 1, nvar
691              write(*,'(3X,I6,F20.12)') i-1, onedimdbl(i)
692          enddo
693
694          ! update parameters
695          call PyOptions%set('parameters', coords)
696          call flush(6)
697          if(PySetParams() /= 0) then
698              write(*, '(A)') ">>> ChemShell ERROR: DL_FIND failed to run
     ↪  PySetParams(). Exiting..."
699              status = 1
700          endif
701
702          ! CALLBACK
703          write(*,'(/" >>> DL_FIND is running the callback function for derivatives
             ↪  evaluation..."/)')
704          ! do calculations on the Python side
705
706          call cpu_time(start_time)
707          call flush(6)
708
709          ! run callback function to calculate the gradients
710          if(PyCallback() /= 0) then
711              print *, ">>> ERROR: DL_FIND failed to run callback function.
     ↪  Exiting..."
712              status = 1
713              stop
714          endif
715
716          call cpu_time(end_time)
717          if(debug > 3) then
718              write (*, '(1X,A,F12.3,1X,A)') ">>> DL_FIND: CPU time for the callback
     ↪  function is", end_time-start_time, "s"
```

```fortran
719        endif
720        call PyOptions%get('derivatives', onedimdbl)
721        gradients(:) = onedimdbl(:)
722
723        if(debug > 3) then
724            write(*,'(/" DL_FIND derivatives of updated parameters:"/)')
725            write(*,'(3X,"index",7X,"derivative"/)')
726            do i = 1, nvar
727                write(*,'(1X,I6,3X,F16.9)') i-1, onedimdbl(i)
728            enddo
729            write(*,'(//)')
730            call flush(6)
731        endif
732        deallocate(onedimdbl)
733
734        ! note: energy in Python is an array!
735        allocate(onedimdbl(1))
736        call PyResult%get('energy', onedimdbl)
737        energy = onedimdbl(1)
738
739        deallocate(onedimdbl)
740        ! END OF CALLBACK
741
742    endif
743
744 endsubroutine dlf_get_gradient
```
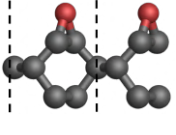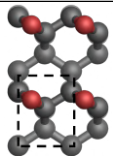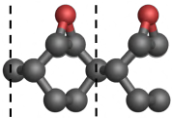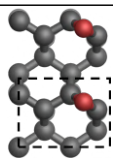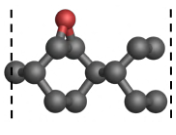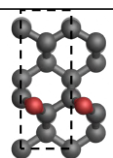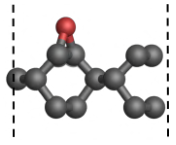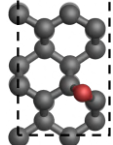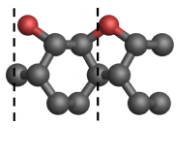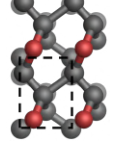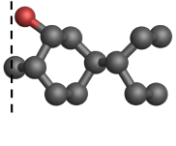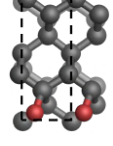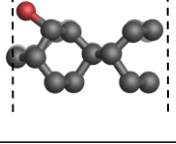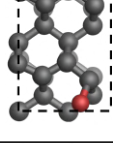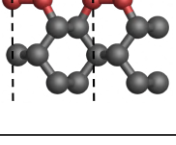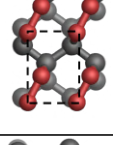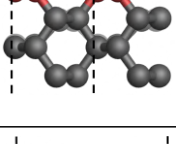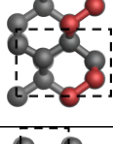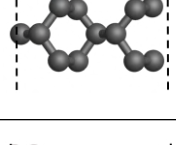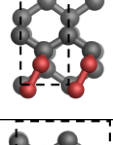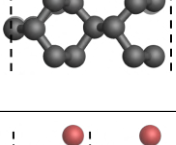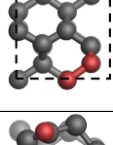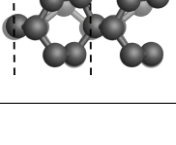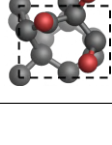
Appendix B
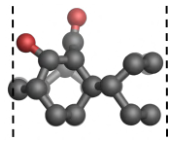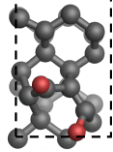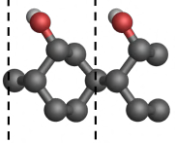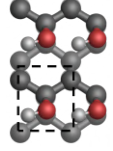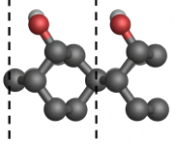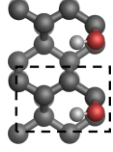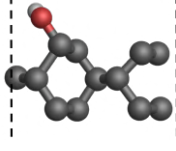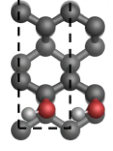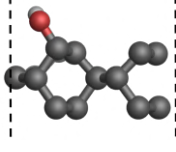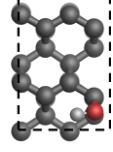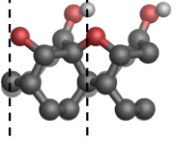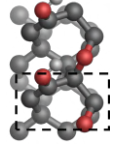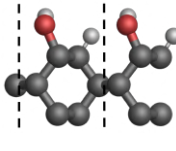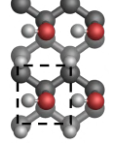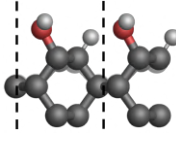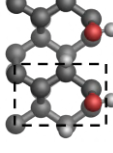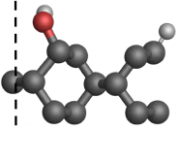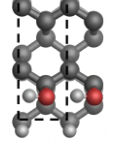
Additional Results
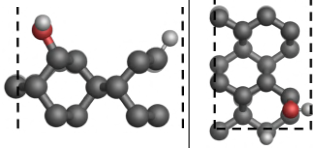
B.1 Oxygen-Terminated Diamond (110) Surface Phases

Table B.1. Table showing all the distinct identified optimised oxygenated diamond (110) surfaces, with their orthographic ball-and-stick visualisations of the first three carbon layers shown from the [001] and [110] directions; their adsorption energies per oxygen adsorbate, E_{ads}, and per surface area, E_{ads}/A; and their oxygen coverages. Carbon, oxygen and hydrogen atoms are shown in grey, red and white respectively, and the unit cell outline is shown as dashed black lines.

Surface Phase	Visualisation Direction		E_{ads}	E_{ads}/A	Coverage
	[001]	[110]	(eV atom^{-1})	(eV Å^{-2})	(atoms nm^{-2})
$p(1 \times 1)$ Ether$_3$			1.77	0.20	11.1
$p(2 \times 1)$ Ether$_3$			1.77	0.10	5.5
$p(1 \times 2)$ Ether$_3$			1.79	0.10	5.5

$p(2 \times 2)$ Ether$_3$			1.78	0.05	2.8
$p(1 \times 1)$ Ether$_5$			1.52	0.17	11.1
$p(1 \times 2)$ Ether$_5$			1.56	0.09	5.5
$p(2 \times 2)$ Ether$_5$			1.68	0.05	2.8
$p(1 \times 1)$ Peroxide			1.42	0.31	22.2
$p(2 \times 1)$ Peroxide			2.04	0.23	11.1
$p(1 \times 2)$ Peroxide			1.14	0.13	11.1
$p(2 \times 2)$ Peroxide			1.78	0.10	5.5
$p(2 \times 1)$ Keto– Ether$_5$			3.03	0.34	5.5

$p(2 \times 2)$ Keto– Ether$_5$			2.25	0.25	2.8
$p(1 \times 1)$ OH$_{single}$			0.52	0.06	11.1
$p(2 \times 1)$ OH$_{single}$			0.46	0.03	5.5
$p(1 \times 2)$ OH$_{single}$			0.62	0.03	5.5
$p(2 \times 2)$ OH$_{single}$			0.47	0.01	2.8
$p(2 \times 1)$ Ether$_5$–OH			1.26	0.14	11.1
$p(1 \times 1)$ OH–H			0.52	0.06	11.1
$p(2 \times 1)$ OH–H			0.46	0.03	5.5
$p(1 \times 2)$ OH–H			0.62	0.03	5.5

$p(2 \times 2)$ OH–H			0.47	0.01	2.8
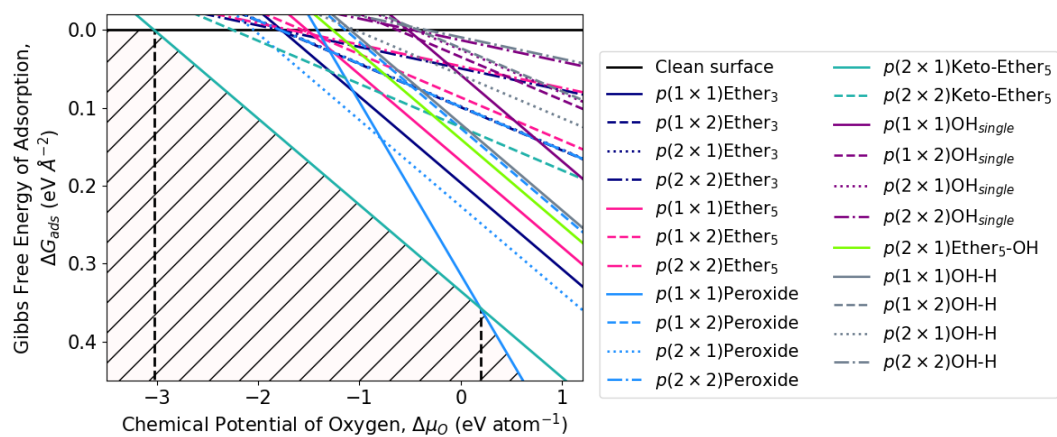
Figure B.1. Free energy plot showing the relationship between ΔG_{ads} and $\Delta \mu_{\mathrm{O}}$ for all of the identified oxygenated (110) surface phases. All $(\Delta \mu_{\mathrm{O}}, \Delta G_{\mathrm{ads}})$ coordinates within the convex hull of ΔG_{ads} are shaded, with the dashed black lines showing the boundaries between the thermodynamically most stable phases.

B.2 Cluster Models for Core-Level Binding Energy Simulations
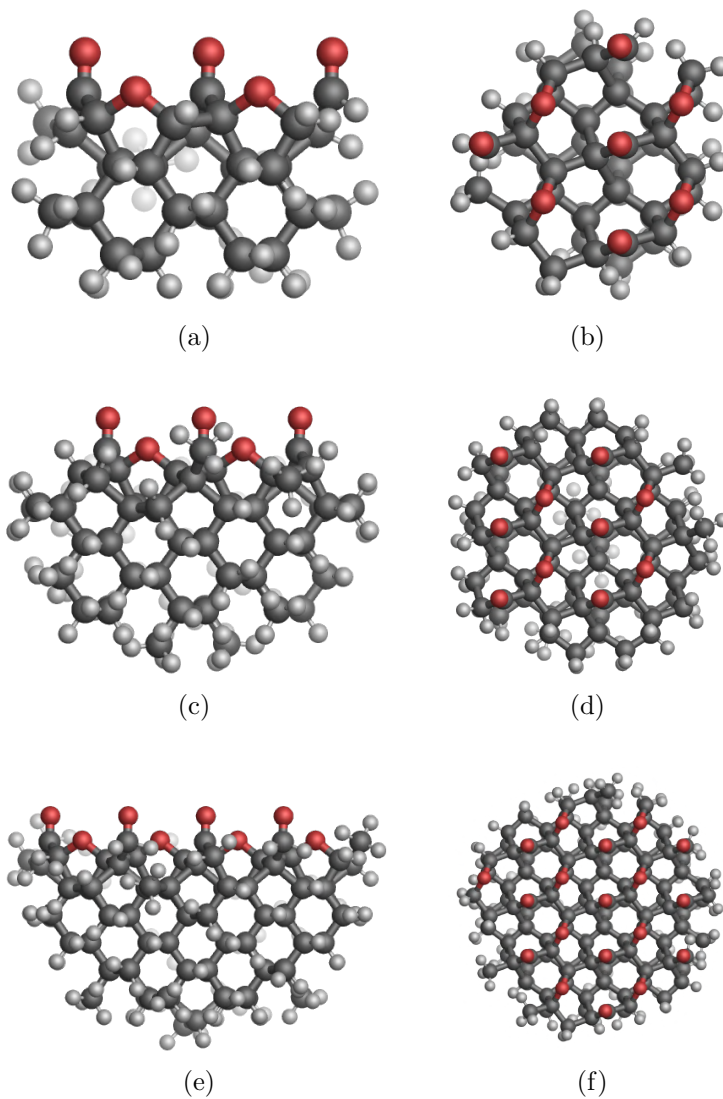
Figure B.2. Orthographic ball-and-stick visualisations of the cluster models of radii (a–b) $9\,a_0$, (c–d) $12\,a_0$, and (e–f) $15\,a_0$ of the $p(2\times1)$Keto-Ether$_5$ phase, shown from two different perspectives (the first column is in the [001] direction as a 'front-on' view, and the third column is in the [110] direction as a 'top-down' view). Carbon, oxygen and hydrogen atoms are shown in grey, red and white respectively.
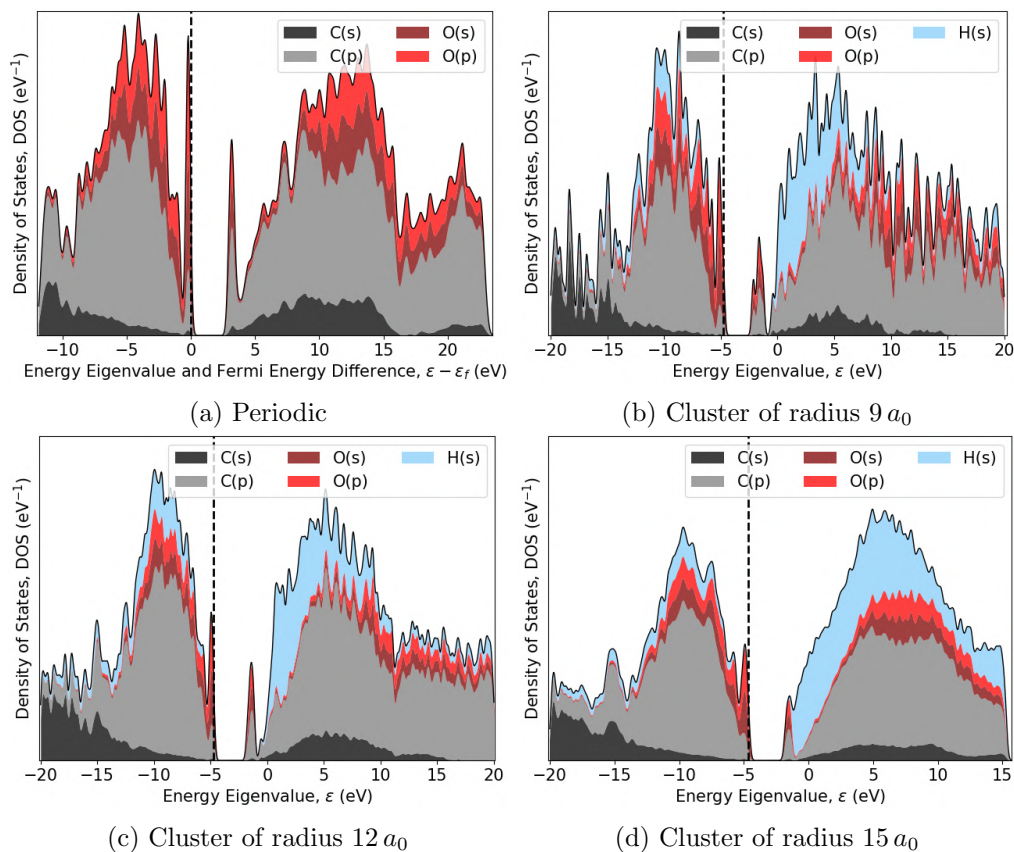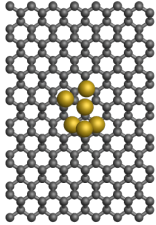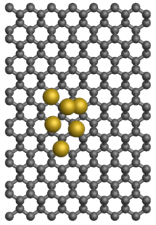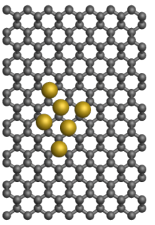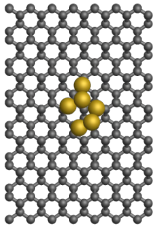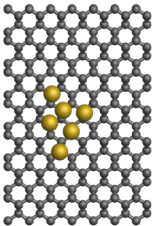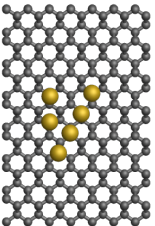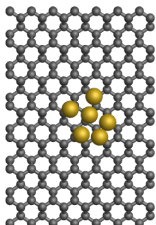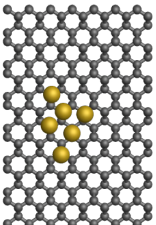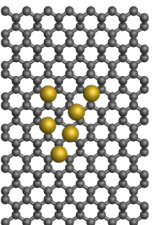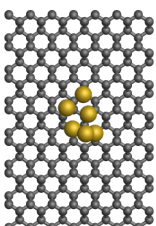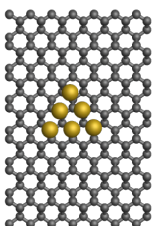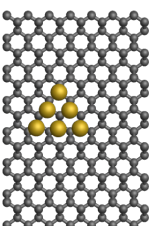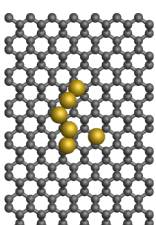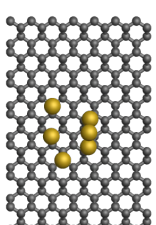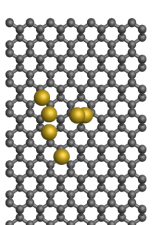
Figure B.3. A comparison of the projected density of states for the periodic and cluster models of various radii of the $p(2 \times 1)$Keto-Ether$_5$ phase. The black dashed vertical line indicates the position of the highest occupied molecular orbital.

B.3   Optimised Structures of Supported Gold Nanoclusters

Method	Global Minimum	Local Minimum 1	Local Minimum 2
Au$_6$			
SchNet$_{\text{initial}}$+MBD@rsSCS			
PBE+MBD@rsSCS			
SchNet$_{\text{adapt3}}$+MBD@rsSCS			
auorg+MBD@rsSCS			
au_full+MBD@rsSCS			
Au$_{15}$			

SchNet$_{\text{initial}}$$^{+\text{MBD@rsSCS}}$			
PBE$^{+\text{MBD@rsSCS}}$			
SchNet$_{\text{adapt3}}$$^{+\text{MBD@rsSCS}}$			
auorg$^{+\text{MBD@rsSCS}}$			
au_full$^{+\text{MBD@rsSCS}}$			

Au$_{40}$

SchNet$_{\text{initial}}$$^{+\text{MBD@rsSCS}}$			

Table B.2. Orthographic visualisations of Au_6, Au_{15} and Au_{40} NCs optimised using PBE$^{+MBD@rsSCS}$ and SchNet$_{adapt3}$$^{+MBD@rsSCS}$. Initial structures are the global minima and two randomly-chosen local minima from basin-hopping runs conducted using SchNet$_{initial}$$^{+MBD@rsSCS}$ for the first adaptive sampling run. Structures are shown from the [110] direction as a top view, and gold and carbon atoms are shown in gold and grey respectively.

Bibliography

[1] Ko, S. H. *et al.* Direct Nanoimprinting of Metal Nanoparticles for Nanoscale Electronics Fabrication. *Nano Lett.* **7**, 1869–1877 (2007). URL https://doi.org/10.1021/nl070333v. (page: 1)

[2] Chen, H. L., Chuang, S. Y., Cheng, H. C., Lin, C. H. & Chu, T. C. Directly patterning metal films by nanoimprint lithography with low-temperature and low-pressure. *Microelectron. Eng.* **83**, 893–896 (2006). URL https://doi.org/10.1016/j.mee.2006.01.095. (page: 1)

[3] Sharma, J., Chhabra, R., Yan, H. & Liu, Y. pH-driven conformational switch of "i-motif" DNA for the reversible assembly of gold nanoparticles. *Chem. Commun.* 477–479 (2007). URL https://doi.org/10.1039/B612707J. (page: 1)

[4] Ahn, J. & Jung, J. H. pH-driven Assembly and Disassembly Behaviors of DNA-modified Au and Fe_3O_4@SiO_2 Nanoparticles. *Bull. Korean Chem. Soc.* **36**, 1922–1925 (2015). URL https://doi.org/10.1002/bkcs.10353. (page: 1)

[5] Lalander, C. H., Zheng, Y., Dhuey, S., Cabrini, S. & Bach, U. DNA-Directed Self-Assembly of Gold Nanoparticles onto Nanopatterned Surfaces: Controlled Placement of Individual Nanoparticles into Regular Arrays. *ACS Nano* **4**, 6153–6161 (2010). URL https://doi.org/10.1021/nn101431k. (page: 1)

[6] Sharma, J., Chhabra, R., Liu, Y., Ke, Y. & Yan, H. DNA-Templated Self-Assembly of Two-Dimensional and Periodical Gold Nanoparticle Arrays. *Angew. Chem.* **118**, 744–749 (2006). URL https://doi.org/10.1002/ange.200503208. (page: 1)

[7] Zhang, J., Liu, Y., Ke, Y. & Yan, H. Periodic Square-Like Gold Nanoparticle Arrays Templated by Self-Assembled 2D DNA Nanogrids on a Surface.

Nano Lett. **6**, 248–251 (2006). URL https://doi.org/10.1021/nl052210l. (page: 1)

[8] Liu, S., Maoz, R. & Sagiv, J. Planned Nanostructures of Colloidal Gold via Self-Assembly on Hierarchically Assembled Organic Bilayer Template Patterns with In-situ Generated Terminal Amino Functionality. *Nano Lett.* **4**, 845–851 (2004). URL https://doi.org/10.1021/nl049755k. (page: 1)

[9] Khatri, O. P., Han, J., Ichii, T., Murase, K. & Sugimara, H. Self-Assembly Guided One-Dimensional Arrangement of Gold Nanoparticles: A Facile Approach. *J. Phys. Chem. C* **112**, 16182–16185 (2008). URL https://doi.org/10.1021/jp806932b. (page: 1)

[10] Li, B. *et al.* Controlled Assembly of Gold Nanoparticles and Graphene Oxide Sheets on Dip Pen Nanolithography-Generated Templates. *Langmuir* **25**, 10455–10458 (2009). URL https://doi.org/10.1021/la902601v. (page: 1)

[11] Liu, S., Maoz, R., Schmid, G. & Sagiv, J. Template Guided Self-Assembly of [Au$_{55}$] Clusters on Nanolithographically Defined Monolayer Patterns. *Nano Lett.* **2**, 1055–1060 (2002). URL https://doi.org/10.1021/nl025659c. (page: 1)

[12] Ofir, Y. *et al.* Polyelectrolyte Negative Resist Patterns as Templates for the Electrostatic Assembly of Nanoparticles and Electroless Deposition of Metallic Films. *Adv. Mater.* **20**, 2561–2566 (2008). URL https://doi.org/10.1002/adma.200703095. (page: 1)

[13] Mariano, R. G., McKelvey, K., White, H. S. & Kanan, M. W. Selective increase in CO_2 electroreduction activity at grain–boundary surface terminations. *Science* **358**, 1187–1192 (2017). URL https://doi.org/10.1126/science.aao3691. (page: 1)

[14] Kim, T. W. & Choi, K. Nanoporous $BiVO_4$ Photoanodes with Dual–Layer Oxygen Evolution Catalysts for Solar Water Splitting. *Science* **343**, 990–996 (2014). URL https://doi.org/10.1126/science.1246913. (page: 1)

[15] Wang, C. *et al.* Proton Exchange Membrane Fuel Cells with Carbon Nanotube Based Electrodes. *Nano Lett.* **4**, 345–348 (2004). URL https://doi.org/10.1021/nl034952p. (page: 1)

[16] Ferrando, R., Jellinek, J. & Johnston, R. L. Nanoalloys: From Theory to Applications of Alloy Clusters and Nanoparticles. *Chem. Rev.* **108**, 845–910 (2008). URL https://doi.org/10.1021/cr040090g. (page: 1)

[17] Trindell, J. A., Duan, Z., Henkelman, G. & Crooks, R. M. Well-Defined Nanoparticle Electrocatalysts for the Refinement of Theory. *Chem. Rev.* **120**, 814–850 (2020). URL https://doi.org/10.1021/acs.chemrev.9b00246. (pages: 1, 3)

[18] Li, H. *et al.* Oxygen Reduction Reaction on Classically Immiscible Bimetallics: A Case Study of RhAu. *J. Phys. Chem. C* **122**, 2712–2716 (2018). URL https://doi.org/10.1021/acs.jpcc.7b10974. (page: 1)

[19] Gawande, M. B. *et al.* Core–shell nanoparticles: synthesis and applications in catalysis and electrocatalysis. *Chem. Soc. Rev.* **44**, 7540–7590 (2015). URL https://doi.org/10.1039/C5CS00343A. (page: 1)

[20] Bottari, F. & De Wael, K. Electrodeposition of gold nanoparticles on boron doped diamond electrodes for the enhanced reduction of small organic molecules. *J. Electroanal. Chem.* **801**, 521–526 (2017). URL https://doi.org/10.1016/j.jelechem.2017.07.053. (page: 1)

[21] Cobb, S. J., Ayres, Z. J. & Macpherson, J. V. Boron Doped Diamond: A Designer Electrode Material for the Twenty-First Century. *Annu. Rev. Anal. Chem.* **11**, 463–484 (2018). URL https://doi.org/10.1146/annurev-anchem-061417-010107. (pages: 1, 9, 14, and 74)

[22] Hussein, H. E. M. *et al.* Tracking Metal Electrodeposition Dynamics from Nucleation and Growth of a Single Atom to a Crystalline Nanoparticle. *ACS Nano* **12**, 7388–7396 (2018). URL https://doi.org/10.1021/acsnano.8b04089. (pages: 2, 3, 4, 6, 7, 9, 14, 15, 47, 59, 70, 71, 74, 82, 106, 108, 110, 114, 116, 123, 148, 157, 158, and 161)

[23] Melander, M. M., Kuisma, M. J., Christensen, T. E. K. & Honkala, K. Grand-canonical approach to density functional theory of electrocatalytic systems: Thermodynamics of solid-liquid interfaces at constant ion and electrode potentials. *J. Chem. Phys.* **150**, 041706 (2019). URL https://doi.org/10.1063/1.5047829. (pages: 2, 5, and 162)

[24] Sundararaman, R., Goddard III, W. A. & Arias, T. A. Grand canonical electronic density-functional theory: Algorithms and applications to electro-

chemistry. *J. Chem. Phys.* **146**, 114104 (2017). URL https://doi.org/10.1063/1.4978411. (pages: 2, 162)

[25] Inkson, B. J. 2 - Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) for materials characterization, 17–43 (Woodhead Publishing, 2016). URL https://doi.org/10.1016/B978-0-08-100040-3.00002-X. (page: 3)

[26] Ustarroz, J., Gupta, U., Hubin, A., Bals, S. & Terryn, H. Electrodeposition of Ag nanoparticles onto carbon coated TEM grids: A direct approach to study early stages of nucleation. *Electrochem. Commun.* **12**, 1706–1709 (2010). URL https://doi.org/10.1016/j.elecom.2010.10.002. (page: 3)

[27] Ustarroz, J., Ke, X., Hubin, A., Bals, S. & Terryn, H. New Insights into the Early Stages of Nanoparticle Electrodeposition. *J. Phys. Chem. C* **116**, 2322–2329 (2012). URL https://doi.org/10.1021/jp210276z. (page: 3)

[28] Cherigui, E. A. M. *et al.* Comprehensive Study of the Electrodeposition of Nickel Nanostructures from Deep Eutectic Solvents: Self-Limiting Growth by Electrolysis of Residual Water. *J. Phys. Chem. C* **121**, 9337–9347 (2017). URL https://doi.org/10.1021/acs.jpcc.7b01104. (page: 3)

[29] Ustarroz, J. *et al.* Electrodeposition of Highly Porous Pt Nanoparticles Studied by Quantitative 3D Electron Tomography: Influence of Growth Mechanisms and Potential Cycling on the Active Surface Area. *ACS Appl. Mater. Interfaces* **9**, 16168–16177 (2017). URL https://doi.org/10.1021/acsami.7b01619. (page: 3)

[30] Leenheer, A. J., Jungjohann, K. L., Zavadil, K. R., Sullivan, J. P. & Harris, C. T. Lithium Electrodeposition Dynamics in Aprotic Electrolyte Observed *in Situ* via Transmission Electron Microscopy. *ACS Nano* **9**, 4379–4389 (2015). URL https://doi.org/10.1021/acsnano.5b00876. (page: 3)

[31] Wang, X. *et al.* New Insights on the Structure of Electrochemically Deposited Lithium Metal and Its Solid Electrolyte Interphases via Cryogenic TEM. *Nano Lett.* **17**, 7606–7612 (2017). URL https://doi.org/10.1021/acs.nanolett.7b03606. (page: 3)

[32] Pei, A., Zheng, G., Shi, F., Li, Y. & Cui, Y. Nanoscale Nucleation and Growth of Electrodeposited Lithium Metal. *Nano Lett.* **17**, 1132–1139 (2017). URL https://doi.org/10.1021/acs.nanolett.6b04755. (page: 3)

[33] Lukaczynska, M. *et al.* Influence of water content and applied potential on the electrodeposition of Ni coatings from deep eutectic solvents. *Electrochim. Acta* **319**, 690–704 (2019). URL https://doi.org/10.1016/j.electacta.2019.06.161. (page: 3)

[34] Ross, F. M. Opportunities and challenges in liquid cell electron microscopy. *Science* **350**, aaa98861–aaa98869 (2015). URL https://doi.org/10.1126/science.aaa9886. (page: 3)

[35] Hodnik, N., Dehm, G. & Mayrhofer, K. J. J. Importance and Challenges of Electrochemical *in Situ* Liquid Cell Electron Microscopy for Energy Conversion Research. *Acc. Chem. Res.* **49**, 2015–2022 (2016). URL https://doi.org/10.1021/acs.accounts.6b00330. (page: 3)

[36] Li, Z. Y. *et al.* Three-dimensional atomic-scale structure of size-selected gold nanoclusters. *Nature* **451**, 46–48 (2008). URL https://doi.org/10.1038/nature06470. (page: 4)

[37] Wang, X. *et al.* AutoDetect-mNP: An Unsupervised Machine Learning Algorithm for Automated Analysis of Transmission Electron Microscope Images of Metal Nanoparticles. *JACS Au* **1**, 316–327 (2021). URL https://doi.org/10.1021/jacsau.0c00030. (page: 4)

[38] Schiøtz, J. *et al.* Using Neural Networks to Identify Atoms in HRTEM Images. *Microsc. Microanal.* **25 (S2)**, 216–217 (2019). URL https://doi.org/10.1017/S1431927619001818. (page: 4)

[39] Ede, J. M. & Beanland, R. Partial Scanning Transmission Electron Microscopy with Deep Learning. *Sci. Rep.* **10**, 8332 (2020). URL https://doi.org/10.1038/s41598-020-65261-0. (page: 4)

[40] Ede, J. M. Deep learning in electron microscopy. *Mach. Learn. Sci. Technol.* **2**, 011004 (2021). URL https://iopscience.iop.org/article/10.1088/2632-2153/abd614. (page: 4)

[41] Timoshenko, J., Lu, D., Lin, Y. & Frenkel, A. I. Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles. *J. Phys. Chem, Lett.* **8**, 5091–5098 (2017). URL https://doi.org/10.1021/acs.jpclett.7b02364. (page: 4)

[42] Khan, M. K., Wang, Q. Y. & Fitzpatrick, M. E. 1 - Atomic force microscopy (AFM) for materials characterization, 1–16 (Woodhead Publishing, 2016). URL https://doi.org/10.1016/B978-0-08-100040-3.00001-8. (page: 4)

[43] Wang, S. *et al.* Nanoscale observation of the solid electrolyte interface and lithium dendrite nucleation–growth process during the initial lithium electrodeposition. *J. Mater. Chem. A* **8**, 18348–18357 (2020). URL https://doi.org/10.1039/D0TA06141G. (page: 4)

[44] Matsushima, H., Lin, S., Morin, S. & Magnussen, O. M. *In situ* video–STM studies of the mechanisms and dynamics of electrochemical bismuth nanostructure formation on Au. *Faraday Discuss.* **193**, 171–185 (2016). URL https://doi.org/10.1039/C6FD00086J. (page: 4)

[45] Lachenwitzer, A., Morin, S., Magnussen, O. M. & Behm, R. J. *In situ* STM study of electrodeposition and anodic dissolution of Ni on Ag(111). *Phys. Chem. Chem. Phys.* **3**, 3351–3363 (2001). URL https://doi.org/10.1039/B103354A. (page: 4)

[46] Morin, S., Lachenwitzer, A., Möller, F. A., Magnussen, O. M. & Behm, R. J. Comparative In Situ STM Studies on the Electrodeposition of Ultrathin Nickel Films on Ag(111) and Au(111) Electrodes. *J. Electrochem. Soc.* **146**, 1013–1018 (1999). URL https://doi.org/10.1149/1.1391714. (page: 4)

[47] Möller, F. A., Magnussen, O. M. & Behm, R. J. Electrodeposition and Anodic Dissolution of Ni on Au(100): an *in situ* STM Study. *Z. Phys. Chem.* **208**, S57–S75 (1999). URL https://doi.org/10.1524/zpch.1999.208.Part_1_2.057. (page: 4)

[48] Lachenwitzer, A., Vogt, M. R., Magnussen, O. M. & Behm, R. J. Electrodeposition of Ni on Cu(100): an in-situ STM study. *Surf. Sci.* **382**, 107–115 (1997). URL https://doi.org/10.1016/S0039-6028(97)00110-6. (page: 4)

[49] Möller, F. A., Kintrup, J., Lachenwitzer, A., Magnussen, O. M. & Behm, R. J. *In situ* study of the electrodeposition and anodic dissolution of ultrathin epitaxial Ni films on Au(111). *Phys. Rev. B* **56**, 12506–12518 (1997). URL https://doi.org/10.1103/PhysRevB.56.12506. (page: 4)

[50] Strbac, S., Magnussen, O. M. & Behm, R. J. Nanoscale Pattern Formation during Electrodeposition: Ru on Reconstructed Au(111). *Phys. Rev. Lett.* **83**, 3246–3249 (1999). URL https://doi.org/10.1103/PhysRevLett.83.3246. (page: 4)

[51] Ustarroz, J. *et al.* Mobility and Poisoning of Mass–Selected Platinum Nanoclusters during the Oxygen Reduction Reaction. *ACS Catal.* **8**, 6775–6790 (2018). URL https://doi.org/10.1021/acscatal.8b00553. (page: 4)

[52] Daviddi, E. *et al.* Nanoscale Electrochemistry in a Copper/Aqueous/Oil Three-phase System: Surface Structure-Activity-Corrosion Potential Relationships. *Chem. Sci.* **12**, 3055–3069 (2021). URL https://doi.org/10.1039/D0SC06516A. (page: 4)

[53] Aaronson, B. D. B. *et al.* Pseudo-Single-Crystal Electrochemistry on Polycrystalline Electrodes: Visualizing Activity at Grains and Grain Boundaries on Platinum for the Fe^{2+}/Fe^{3+} Redox Reaction. *J. Am. Chem. Soc.* **135**, 3873–3880 (2013). (page: 4)

[54] Lai, S. C. S., Lazenby, R. A., Kirkman, P. M. & Unwin, P. R. Nucleation, aggregative growth and detachment of metal nanoparticles during electrodeposition at electrode surfaces. *Chem. Sci.* **6**, 1126–1138 (2015). URL https://doi.org/10.1039/c4sc02792b. (pages: 4, 7)

[55] Ebejer, N. *et al.* Scanning Electrochemical Cell Microscopy: A Versatile Technique for Nanoscale Electrochemistry and Functional Imaging. *Annu. Rev. Anal. Chem.* **6**, 329–351 (2013). URL https://doi.org/10.1146/annurev-anchem-062012-092650. (page: 4)

[56] Bentley, C. L., Kang, M. & Unwin, P. R. Scanning electrochemical cell microscopy: New perspectives on electrode processes in action. *Curr. Opin. Electrochem.* **6**, 23–30 (2017). URL https://doi.org/10.1016/j.coelec.2017.06.011. (page: 4)

[57] Bentley, C. L., Kang, M. & Unwin, P. R. Nanoscale Surface Structure–Activity in Electrochemistry and Electrocatalysis. *J. Am. Chem. Soc.* **141**, 2179–2193 (2019). URL https://doi.org/10.1021/jacs.8b09828. (page: 4)

[58] Bentley, C. L. *et al.* Nanoscale Electrochemical Mapping. *Anal. Chem.* **91**, 84–108 (2019). URL https://doi.org/10.1021/acs.analchem.8b05235. (page: 4)

[59] Kang, M., Momotenko, D., Page, A., Perry, D. & Unwin, P. R. Frontiers in Nanoscale Electrochemical Imaging: Faster, Multifunctional, and Ultrasensitive. *Langmuir* **32**, 7993–8008 (2016). URL https://doi.org/10.1021/acs.langmuir.6b01932. (page: 4)

[60] Laurinavichyute, V. K., Nizamov, S. & Mirsky, V. M. Real time tracking of the early stage of electrochemical nucleation. *Electrochim. Acta* **382**, 138278 (2021). URL https://doi.org/10.1016/j.electacta.2021.138278. (pages: 4, 5)

[61] Hill, C. M. & Pan, S. A Dark-Field Scattering Spectroelectrochemical Technique for Tracking the Electrodeposition of Single Silver Nanoparticles. *J. Am. Chem. Soc.* **135**, 17250–17253 (2013). URL https://doi.org/10.1021/ja4075387. (pages: 4, 5)

[62] Hill, C. M., Clayton, D. A. & Pan, S. Combined optical and electrochemical methods for studying electrochemistry at the single molecule and single particle level: recent progress and perspectives. *Phys. Chem. Chem. Phys.* **15**, 20797–20807 (2013). URL https://doi.org/10.1039/C3CP52756E. (pages: 4, 5)

[63] Sugino, O. *et al.* First–principles molecular dynamics simulation of biased electrode/solution interface. *Surf. Sci.* **601**, 5237–5240 (2007). URL https://doi.org/10.1016/j.susc.2007.04.208. (page: 5)

[64] Yang, K., Yiacoumi, S. & Tsouris, C. Canonical Monte Carlo simulations of the fluctuating–charge molecular water between charged surfaces. *J. Chem. Phys.* **117**, 337–345 (2002). URL https://doi.org/10.1063/1.1480858. (page: 5)

[65] Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136**, B864–B871 (1964). URL https://doi.org/10.1103/PhysRev.136.B864. (pages: 5, 18, 19, 20, 37, 38, 48, 71, 76, 127, 128, and 157)

[66] Kohn, W. & Sham, L. Self–Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **140**, A1133–A1138 (1965). URL https://doi.org/10.1103/PhysRev.140.A1133. (pages: 5, 20, 37, 38, 48, 71, 76, 127, 128, and 157)

[67] Melander, M. M. Grand canonical ensemble approach to electrochemical thermodynamics, kinetics, and model Hamiltonians. *Curr. Opin. Electrochem.* **29**, 100749 (2021). URL https://doi.org/10.1016/j.coelec.2021.100749. (pages: 5, 162)

[68] Hofmann, O. T., Zojer, E., Hörmann, L., Jeindl, A. & Maurer, R. J. First-principles calculations of hybrid inorganic–organic interfaces: from state-of-the-art to best practice. *Phys. Chem. Chem. Phys.* **23**, 8132–8180 (2021).

205

URL https://doi.org/10.1039/D0CP06605B. (pages: 5, 21, 22, 23, 30, 31, 57, 70, 71, 76, 90, 91, and 124)

[69] Maurer, R. J. *et al.* Advances in Density-Functional Calculations for Materials Modeling. *Annu. Rev. Mater. Sci.* **49**, 1–30 (2019). URL https://doi.org/10.1146/annurev-matsci-070218-010143. (pages: 5, 21, 22, 30, 57, 71, 90, 91, 93, 96, and 139)

[70] Blum, V. *et al.* *Ab initio* molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **180**, 2175–2196 (2009). URL https://doi.org/10.1016/j.cpc.2009.06.022. (pages: 5, 24, 40, 48, 51, 73, 76, 77, 127, and 128)

[71] Clark, S. *et al.* First principles methods using CASTEP. *Z. Kristallogr.* **220**, 567–570 (2005). URL https://doi.org/10.1524/zkri.220.5.567.65075. (pages: 5, 50)

[72] Mortensen, J. J., Hansen, L. B. & Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **71**, 035109 (2005). URL https://doi.org/10.1103/PhysRevB.71.035109. (pages: 5, 23, and 161)

[73] Enkovaara, J. *et al.* Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys. Condens. Matter* **22**, 253202 (2010). URL https://doi.org/10.1088/0953-8984/22/25/253202. (pages: 5, 23, and 161)

[74] Ganesh Balasubramani, S. *et al.* TURBOMOLE: Modular program suite for *ab initio* quantum-chemical and condensed-matter simulations. *J. Chem. Phys.* **152**, 184107 (2020). URL https://doi.org/10.1063/5.0004635. (page: 5)

[75] Kühne, T. *et al.* CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103 (2020). URL https://doi.org/10.1063/5.0007045. (page: 5)

[76] Guest, M. F. *et al.* The GAMESS-UK electronic structure package: algorithms, developments and applications. *Mol. Phys.* **103**, 719–747 (2005). URL https://doi.org/10.1080/00268970512331340592. (page: 5)

[77] Valiev, M. *et al.* NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **181**,

1477–1489 (2010). URL https://doi.org/10.1016/j.cpc.2010.04.018. (page: 5)

[78] Elstner, M. *et al.* Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58**, 7260–7268 (1998). URL https://doi.org/10.1103/PhysRevB.58.7260. (pages: 5, 28, 30, 40, 125, 127, and 137)

[79] Hourahine, B. *et al.* DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **152**, 124101 (2020). URL https://doi.org/10.1063/1.5143190. (pages: 5, 28, 29, 30, 40, 125, and 127)

[80] Grimme, S., Bannwarth, C. & Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ($Z = 1$–86). *J. Chem. Theory Comput.* **13**, 1989–2009 (2017). URL https://doi.org/10.1021/acs.jctc.7b00118. (page: 5)

[81] Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019). URL https://doi.org/10.1021/acs.jctc.8b01176. (page: 5)

[82] Li, Y. & Qi, Y. Transferable Self-Consistent Charge Density Functional Tight-Binding Parameters for Li–Metal and Li-Ions in Inorganic Compounds and Organic Solvents. *J. Phys. Chem. C* **20**, 10755–10764 (2018). URL https://doi.org/10.1021/acs.jpcc.8b01839. (page: 5)

[83] Kurban, H., Kurban, M. & Dalkılıç. Density-functional tight-binding approach for the structural analysis and electronic structure of copper hydride metallic nanoparticles. *Mater. Today Commun.* **21**, 100648 (2019). URL https://doi.org/10.1016/j.mtcomm.2019.100648. (page: 5)

[84] Moreira, N. H., Dolgonos, G., Aradi, B., da Rosa, A. L. & Frauenheim, T. Toward an Accurate Density-Functional Tight-Binding Description of Zinc-Containing Compounds. *J. Chem. Theory Comput.* **5**, 605–614 (2009). URL https://doi.org/10.1021/ct800455a. (page: 5)

[85] Dolgonos, G., Aradi, B., Moreira, N. H. & Frauenheim, T. An Improved Self-Consistent-Charge Density-Functional Tight-Binding (SCC-DFTB) Set of

Parameters for Simulation of Bulk and Molecular Systems Involving Titanium. *J. Chem. Theory Comput.* **6**, 266–278 (2010). URL https://doi.org/10.1021/ct900422c. (page: 5)

[86] Fihey, A. *et al.* SCC-DFTB parameters for simulating hybrid gold-thiolates compounds. *J. Comput. Chem.* **36**, 2075–2087 (2015). URL https://doi.org/10.1002/jcc.24046. (pages: 5, 40, 125, 127, 133, and 137)

[87] Santos, E., Nazmutdinov, R. & Schmickler, W. Electron transfer at different electrode materials: Metals, semiconductors, and graphene. *Curr. Opin. Electrochem.* **19**, 106–112 (2020). URL https://doi.org/10.1016/j.coelec.2019.11.003. (page: 5)

[88] Perdew, J. P., Ruzsinszky, A., Constantin, L. A., Sun, J. & Csonka, G. I. Some Fundamental Issues in Ground-State Density Functional Theory: A Guide for the Perplexed. *J. Chem. Theory Comput.* **5**, 902–908 (2009). URL https://doi.org/10.1021/ct800531s. (pages: 5, 57)

[89] Marom, N. *et al.* Dispersion Interactions with Density-Functional Theory: Benchmarking Semiempirical and Interatomic Pairwise Corrected Density Functionals. *J. Chem. Theory Comput.* **7**, 3944–3951 (2011). URL https://doi.org/10.1021/ct2005616. (pages: 5, 21)

[90] Meldgaard, S. A., Kolsbjerg, E. L. & Hammer, B. Machine learning enhanced global optimization by clustering local environments to enable bundled atomic energies. *J. Chem. Phys.* **149**, 134104 (2018). URL https://doi.org/10.1063/1.5048290. (page: 6)

[91] Jørgensen, M. S., Groves, M. N. & Hammer, B. Combining Evolutionary Algorithms with Clustering toward Rational Global Structure Optimization at the Atomic Scale. *J. Chem. Theory Comput.* **13**, 1486–1493 (2017). URL https://doi.org/10.1021/acs.jctc.6b01119. (page: 6)

[92] Pihlajamäki, A. *et al.* Monte Carlo Simulations of $Au_{38}(SCH_3)_{24}$ Nanocluster Using Distance–Based Machine Learning Methods. *J. Phys. Chem. A* **124**, 4827–4836 (2020). URL https://doi.org/10.1021/acs.jpca.0c01512. (page: 6)

[93] Vilhelmsen, L. B. & Hammer, B. A genetic algorithm for first principles global structure optimization of supported nano structures. *J. Chem. Phys.* **141**, 044711 (2014). URL https://doi.org/10.1063/1.4886337. (page: 6)

[94] Wolf, M. D. & Landman, U. Genetic Algorithms for Structural Cluster Optimization. *J. Phys. Chem. A* **102**, 6129–6137 (1988). URL https://doi.org/10.1021/jp9814597. (page: 6)

[95] Shao, G. *et al.* An improved genetic algorithm for structural optimization of Au–Ag bimetallic nanoparticles. *Appl. Soft Comput.* **73**, 39–49 (2018). URL https://doi.org/10.1016/j.asoc.2018.08.019. (page: 6)

[96] Logsdail, A., Paz-Borbón, L. O. & Johnston, R. L. Structures and Stabilities of Platinum–Gold Nanoclusters. *J. Comput. Theor. Nanosci.* **6**, 857–866 (2009). URL https://doi.org/10.1166/jctn.2009.1118. (page: 6)

[97] Sørensen, K. H., Jørgensen, M. S., Bruix, A. & Hammer, B. Accelerating atomic structure search with cluster regularization. *J. Chem. Phys.* **28**, 241734 (2018). URL https://doi.org/10.1063/1.5023671. (page: 6)

[98] Wales, D. J. & Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997). URL https://doi.org/10.1021/jp970984n. (pages: 6, 41, 75, 112, and 160)

[99] Doye, J. P. K. & Wales, D. J. Thermodynamics of Global Optimization. *Phys. Rev. Lett.* **80**, 1357–1360 (1998). URL https://doi.org/10.1103/PhysRevLett.80.1357. (pages: 6, 41, 75, 112, and 160)

[100] Goedecker, S. Minima Hopping: An Efficient Search Method for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems. *J. Chem. Phys.* **120**, 9911–9917 (2004). URL https://doi.org/10.1063/1.1724816. (page: 6)

[101] Ahmad, Z., Xie, T., Maheshwari, C., Grossman, J. C. & Viswanathan, V. Machine Learning Enabled Computational Screening of Inorganic Solid Electrolytes for Suppression of Dendrite Formation in Lithium Metal Anodes. *ACS Cent. Sci.* **4**, 996–1006 (2018). URL https://doi.org/10.1021/acscentsci.8b00229. (page: 6)

[102] Westermayr, J., Gastegger, M., Schütt, K. T. & Maurer, R. J. Perspective on integrating machine learning into computational chemistry and materials science. *J. Chem. Phys.* **154**, 230903 (2021). URL https://doi.org/10.1063/5.0047760. (pages: 6, 33, and 34)

[103] Sear, R. P. Nucleation: theory and applications to protein solutions and colloidal suspensions. *J. Phys. Condens. Matter* **19**, 033101 (2007). URL https://iopscience.iop.org/article/10.1088/0953-8984/19/3/033101. (page: 6)

[104] Milchev, A. Nucleation phenomena in electrochemical systems: kinetic models. *ChemTexts* **2**, 1–9 (2016). URL https://link.springer.com/article/10.1007/s40828-015-0021-1. (pages: 6, 7)

[105] Milchev, A. Electrochemical phase formation: classical and atomistic theoretical models. *Nanoscale* **8**, 13867–13872 (2016). URL https://doi.org/10.1039/C6NR02354A. (pages: ix, 6, 7, and 8)

[106] Milchev, A. & Stoyanov, S. Classical and atomistic models of electrolytic nucleation: Comparison with experimental data. *J. Electroanal. Chem.* **72**, 33–43 (1976). URL http://doi.org/10.1016/S0022-0728(76)80073-3. (page: 6)

[107] Budevski, E., Staikov, G. & Lorenz, W. J. Electrocrystallization: Nucleation and growth phenomena. *Electrochim. Acta* **45**, 2559–2574 (2000). URL http://doi.org/10.1016/S0013-4686(00)00353-4. (page: 6)

[108] Ustarroz, J. *et al.* A Generalized Electrochemical Aggregative Growth Mechanism. *J. Am. Chem. Soc.* **135**, 11550–11561 (2013). URL https://doi.org/10.1021/ja402598k. (page: 7)

[109] Williamson, M. J., Tromp, R. M., Vereecken, P. M., Hull, R. & Ross, F. M. Dynamic microscopy of nanoscale cluster growth at the solid–liquid interface. *Nature Mater.* **2**, 532–536 (2003). URL https://doi.org/10.1038/nmat944. (page: 7)

[110] Radisic, A., Vereecken, P. M., Hannon, J. B., Searson, P. C. & Ross, F. M. Quantifying Electrochemical Nucleation and Growth of Nanoscale Clusters Using Real–Time Kinetic Data. *Nano Lett.* **6**, 238–242 (2006). URL https://doi.org/10.1021/nl052175i. (page: 7)

[111] Velmurugan, J., Noël, J., Nogala, W. & Mirkin, M. V. Nucleation and growth of metal on nanoelectrodes. *Chem. Sci.* **3**, 3307–3314 (2012). URL https://doi.org/10.1039/C2SC21005C. (page: 7)

[112] Kim, Y. *et al.* Nucleation and Aggregative Growth of Palladium Nanoparticles on Carbon Electrodes: Experiment and Kinetic Model. *J. Phys. Chem. C* **119**,

17389–17397 (2015). URL https://doi.org/10.1021/acs.jpcc.5b03513. (page: 7)

[113] Sosso, G. C. *et al.* Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations. *Chem. Rev.* **116**, 7078–7116 (2016). URL https://doi.org/10.1021/acs.chemrev.5b00744. (page: 7)

[114] Kiang, C. S. *et al.* A Reexamination of Homogeneous Nucleation Theory. *J. Atmos. Sci.* **28**, 1222–1232 (1971). URL https://doi.org/10.1175/1520-0469(1971)028<1222:AROHNT>2.0.CO;2. (page: 7)

[115] Laaksonen, A. & Napari, I. Breakdown of the Capillarity Approximation in Binary Nucleation: A Density Functional Study. *J. Phys. Chem. B* **105**, 11678–11682 (2001). URL https://doi.org/10.1021/jp0116454. (page: 7)

[116] Lee, J. K., Abraham, F. F. & Pound, G. M. On the validity of the capillarity approximation in the rate theory of homogeneous nucleation. *Surf. Sci.* **34**, 745–758 (1973). URL https://doi.org/10.1016/0039-6028(73)90041-1. (page: 7)

[117] Schenter, G. K., Kathmann, S. M. & Garrett, B. C. Dynamical Nucleation Theory: A New Molecular Approach to Vapor–Liquid Nucleation. *Phys. Rev. Lett.* **82**, 3484–3487 (1999). URL https://doi.org/10.1103/PhysRevLett.82.3484. (page: 7)

[118] Kalikmanov, V. I. Mean–field kinetic nucleation theory. *J. Chem. Phys.* **124**, 124505 (2006). URL https://doi.org/10.1063/1.2178812. (page: 7)

[119] Russell, K. C. Linked flux analysis of nucleation in condensed phases. *Acta Metall.* **16**, 761–769 (1968). URL https://doi.org/10.1016/0001-6160(68)90148-X. (page: 7)

[120] Peters, B. On the coupling between slow diffusion transport and barrier crossing in nucleation. *J. Chem. Phys.* **135**, 044107 (2011). URL https://doi.org/10.1063/1.3613674. (page: 7)

[121] Wei, P. F. & Kelton, R., K. F. Falster. Coupled–flux nucleation modeling of oxygen precipitation in silicon. *J. Appl. Phys.* **88**, 5062–5070 (2000). URL https://doi.org/10.1063/1.1311309. (page: 7)

[122] Kelton, K. F. Time–dependent nucleation in partitioning transformations. *Acta Mater.* **48**, 1967–1980 (2000). URL https://doi.org/10.1016/S1359-6454(99)00455-3. (page: 7)

[123] Gránásy, L. Diffuse interface theory of nucleation. *J. Non-Cryst. Solids* **162**, 301–303 (1993). URL https://doi.org/10.1016/0022-3093(93)91250-7. (page: 7)

[124] Gránásy, L. & Herlach, D. M. Diffuse interface approach to crystal nucleation in glasses. *J. Non-Cryst. Solids* **192–193**, 470–473 (1995). URL https://doi.org/10.1016/0022-3093(95)00430-0. (page: 7)

[125] Blow, K. E., Quigley, D. & Sosso, G. C. The Seven Deadly Sins: when computing crystal nucleation rates, the devil is in the details (2021). 2104.13104. (page: 7)

[126] Isaev, V. A., Grishenkova, O. V. & Zaykov, Y. P. Theoretical modeling of electrochemical nucleation and growth of a single metal nanocluster on a nanoelectrode. *RSC Adv.* **10**, 6979–6984 (2020). URL https://doi.org/10.1039/D0RA00608D. (page: 9)

[127] Henkelman, G., Uberuaga, B. P. & Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **113**, 9901–9904 (2000). URL https://doi.org/10.1063/1.1329672. (page: 9)

[128] Wooddell, C. E. Method of Comparing the Hardness of Electric Furnace Products and Natural Abrasives. *Trans. Electrochem. Soc.* **68**, 111–130 (1935). URL https://iopscience.iop.org/article/10.1149/1.3493860. (page: 9)

[129] Wei, L., Kuo, P. K., Thomas, R. L., Anthony, T. R. & Banholzer, W. F. Thermal conductivity of isotopically modified single crystal diamond. *Phys. Rev. Lett.* **70**, 3764–3767 (1993). URL https://doi.org/10.1103/PhysRevLett.70.3764. (page: 9)

[130] Tohei, T., Kuwabara, A., Oba, F. & Tanaka, I. Debye temperature and stiffness of carbon and boron nitride polymorphs from first principles calculations. *Phys. Rev. B* **73**, 064304 (2006). URL https://doi.org/10.1103/PhysRevB.73.064304. (page: 9)

[131] Macpherson, J. V. A practical guide to using boron doped diamond in electrochemical research. *Phys. Chem. Chem. Phys.* **17**, 2935–2949 (2015). URL https://doi.org/10.1039/c4cp04022h. (pages: 9, 14, 47, 51, 71, and 75)

[132] Williams, R. J. *et al.* High Power Diamond Raman Lasers. *IEEE J. Sel. Top. Quantum* **24**, 1–14 (2018). URL https://doi.org/10.1109/JSTQE.2018.2827658. (page: 9)

[133] Kianinia, M. & Aharonovich, I. Diamond photonics is scaling up. *Nat. Photonics* **14**, 599–600 (2020). URL https://doi.org/10.1038/s41566-020-0695-9. (page: 9)

[134] Nizovtsev, A. P. *et al.* A quantum computer based on NV centers in diamond: Optically detected nutations of single electron and nuclear spins. *Opt. Spectrosc.* **99**, 233–244 (2005). URL https://doi.org/10.1134/1.2034610. (page: 9)

[135] Dutt, M. V. G. *et al.* Quantum Register Based on Individual Electronic and Nuclear Spin Qubits in Diamond. *Science* **316**, 1312–1316 (2007). URL https://doi.org/10.1126/science.1139831. (page: 9)

[136] Neumann, P. *et al.* Multipartite Entanglement Among Single Spins in Diamond. *Science* **320**, 1326–1329 (2008). URL https://doi.org/10.1126/science.1157233. (page: 9)

[137] Bundy, F. P. *et al.* The pressure-temperature phase and transformation diagram for carbon; updated through 1994. *Carbon* **34**, 141–153 (1996). URL https://doi.org/10.1016/0008-6223(96)00170-4. (pages: 9, 10, and 59)

[138] Bulanova, G. P. The formation of diamond. *J. Geochem. Explor.* **53**, 1–23 (1995). URL https://doi.org/10.1016/0375-6742(94)00016-5. (page: 10)

[139] Achard, J. *et al.* High quality MPACVD diamond single crystal growth: high microwave power density regime. *J. Phys. D Appl. Phys.* **40**, 6175–6188 (2007). URL https://doi.org/10.1088/0022-3727/40/20/S04. (pages: 10, 12, 13, 47, and 157)

[140] Badzian, A. & Badzian, T. Diamond homoepitaxy by chemical vapor deposition. *Diam. Relat. Mater.* **2**, 147–157 (1993). URL https://doi.org/10.1016/0925-9635(93)90046-5. (page: 10)

[141] Auciello, O. & Aslam, D. M. Review on advances in microcrystalline, nanocrystalline and ultrananocrystalline diamond films-based micro/nanoelectromechanical systems technologies. *J. Mater. Sci.* **56**, 7171–7230 (2021). URL https://doi.org/10.1007/s10853-020-05699-9. (page: 10)

[142] Jin, S. & Mavoori, H. Processing and properties of CVD diamond for thermal management. *J. Electron. Mater.* **27**, 1148–1153 (1998). URL https://doi.org/10.1007/s11664-998-0063-x. (page: 10)

[143] Ivandinia, T. A. & Einaga, Y. Polycrystalline boron-doped diamond electrodes for electrocatalytic and electrosynthetic applications. *Chem. Commun.* **53**, 1338–1347 (2017). URL https://doi.org/10.1039/C6CC08681K. (pages: 10, 14)

[144] Werner, M. & Locher, R. Growth and application of undoped and doped diamond films. *Rep. Prog. Phys.* **61**, 1665–1710 (1998). URL https://doi.org/10.1088/0034-4885/61/12/002. (pages: 10, 46)

[145] Balmer, R. S. *et al.* Chemical vapour deposition synthetic diamond: materials, technology and applications. *J. Phys. Condens. Matter* **21**, 364221 (2009). URL https://doi.org/10.1088/0953-8984/21/36/364221. (pages: 10, 46, 51, 74, and 97)

[146] Wood, G. F. *et al.* High pressure high temperature synthesis of highly boron doped diamond microparticles and porous electrodes for electrochemical applications. *Carbon* **171**, 845–856 (2021). URL https://doi.org/10.1016/j.carbon.2020.09.038. (page: 10)

[147] Kiflawi, I., Kanda, H. & Lawson, S. C. The effect of the growth rate on the concentration of nitrogen and transition metal impurities in HPHT synthetic diamonds. *Diam. Relat. Mater.* **11**, 204–211 (2002). URL https://doi.org/10.1016/S0925-9635(01)00569-6. (page: 11)

[148] Martineau, P. M. *et al.* High crystalline quality single crystal chemical vapour deposition diamond. *J. Phys. Condens. Matter* **21**, 364205 (2009). URL https://doi.org/10.1088/0953-8984/21/36/364205. (page: 11)

[149] Nagy, B. & Strand, R. Neighborhood Sequences in the Diamond Grid – Algorithms with Four Neighbors. In Wiederhold, P. & Barneva, R. P. (eds.) Combinatorial Image Analysis, 109–121 (Springer, 2005). URL https://doi.org/10.1007/978-3-642-10210-3_9. (page: 11)

[150] Brinza, O. *et al.* Dependence of CVD diamond growth rate on substrate orientation as a function of process parameters in the high microwave power density regime. *Phys. Status Solidi A* **205**, 2114–2120 (2008). URL https://doi.org/10.1002/pssa.200879716. (pages: 12, 13, and 157)

[151] Hartman, P. Structure and Morphology. In Hartman, P. (ed.) Crystal growth: An introduction, 367–402 (North-Holland Publishing Company, 1973). (pages: 12, 13, and 157)

[152] Mackey, B. L. *et al.* Oxygen Adsorption on the (110)-Oriented Diamond Surface. *J. Phys. Chem. B* **105**, 3803–3812 (2001). URL https://doi.org/10.1021/jp003586k. (pages: 12, 13, 47, 54, 57, 68, and 157)

[153] Giling, L. J. & Van Enckevort, W. J. P. On the influence of surface reconstruction on crystal growth processes. *Surf. Sci.* **161**, 567–583 (1985). URL https://doi.org/10.1016/0039-6028(85)90828-3. (page: 12)

[154] Friel, I. *et al.* Control of surface and bulk crystalline quality in single crystal diamond grown by chemical vapour deposition. *Diam. Relat. Mater.* **18**, 808–815 (2009). URL https://doi.org/10.1016/j.diamond.2009.01.013. (page: 13)

[155] Silva, F. *et al.* Geometric modeling of homoepitaxial CVD diamond growth: I. The {100}{111}{110}{113} system. *J. Cryst. Growth* **310**, 187–203 (2008). URL https://doi.org/10.1016/j.jcrysgro.2007.09.044. (pages: 13, 157)

[156] Liu, D.-Q. *et al.* Facet-Resolved Electrochemistry of Polycrystalline Boron-Doped Diamond Electrodes: Microscopic Factors Determining the Solvent Window in Aqueous Potassium Chloride Solutions. *ChemElectroChem* **5**, 3028–3035 (2018). URL https://doi.org/10.1002/celc.201800770. (pages: 14, 47, and 157)

[157] Muzyka, K. *et al.* Boron-doped diamond: current progress and challenges in view of electroanalytical applications. *Anal. Methods* **11**, 397–414 (2019). URL https://doi.org/10.1039/c8ay02197j. (page: 14)

[158] Yang, N. *et al.* Conductive diamond: synthesis, properties, and electrochemical applications. *Chem. Soc. Rev.* **48**, 157–204 (2019). URL https://doi.org/10.1039/C7CS00757D. (page: 14)

[159] Yannopoulos, J. C. Physical and Chemical Properties of Gold. In The Extractive Metallurgy of Gold, 11–23 (Springer, 1991). URL https://doi.org/10.1007/978-1-4684-8425-0_2. (page: 14)

[160] Hammer, B. & Norskov, J. K. Why gold is the noblest of all the metals. *Nature* **376**, 238–240 (1995). URL https://doi.org/10.1038/376238a0. (pages: 14, 136)

[161] Li, C. *et al.* Electrocatalysis of gold-based nanoparticles and nanoclusters. *Mater. Horiz.* **8**, 1657–1682 (2021). URL https://doi.org/10.1039/d0mh01947j. (page: 14)

[162] Rodriguez, P. & Koper, M. T. M. Electrocatalysis on gold. *Phys. Chem. Chem. Phys.* **16**, 13583–13594 (2014). URL https://doi.org/10.1039/C4CP00394B. (page: 14)

[163] Balasubramanian, S., Sheelam, A., Ramanujam, K. & Dhamodharan, R. Green, Seed-Mediated Synthesis of Au Nanowires and Their Efficient Electrocatalytic Activity in Oxygen Reduction Reaction. *ACS Appl. Mater. Interfaces* **9**, 28876–28886 (2017). URL https://doi.org/10.1021/acsami.7b075537. (page: 14)

[164] Li, M., Zhao, G., Geng, R. & Hu, H. Facile electrocatalytic redox of hemoglobin by flower-like gold nanoparticles on boron-doped diamond surface. *Bioelectrochemistry* **74**, 217–221 (2008). URL https://doi.org/10.1016/j.bioelechem.2008.08.004. (pages: 14, 15)

[165] Yu, Y. *et al.* Electrochemistry and Electrocatalysis at Single Gold Nanoparticles Attached to Carbon Nanoelectrodes. *ChemElectroChem* **2**, 58–63 (2015). URL https://doi.org/10.1002/celc.201402312. (page: 14)

[166] Kauffman, D. R., Alfonso, D., Matranga, C., Qian, H. & Jin, R. Experimental and Computational Investigation of Au_{25} Clusters and CO_2: A Unique Interaction and Enhanced Electrocatalytic Activity. *J. Am. Chem. Soc.* **134**, 10237–10243 (2012). URL https://doi.org/10.1021/ja303259q. (page: 14)

[167] Sumner, L. *et al.* Electrocatalytic Oxygen Reduction Activities of Thiol-Protected Nanomolecules Ranging in Size from $Au_28(SR)_{20}$ to $Au_{279}(SR)_{84}$. *J. Phys. Chem. C* **122**, 24809–24817 (2018). URL https://doi.org/10.1021/acs.jpcc.8b07962. (page: 14)

[168] Zhu, W. *et al.* Monodisperse Au Nanoparticles for Selective Electrocatalytic Reduction of CO_2 to CO. *J. Am. Chem. Soc.* **135**, 16833–16836 (2013). URL https://doi.org/10.1021/ja409445p. (page: 14)

[169] Yu, L. *et al.* Hollow AuAg Alloy Nanourchins: Twin Plane and Surface Treatment for Enhanced Methanol Electrooxidation Performance. *ACS Appl. Energy Mater.* **3**, 723–732 (2020). URL https://doi.org/10.1021/acsaem.9b01912. (page: 14)

[170] Celorrio, V. *et al.* Strain Effects on the Oxidation of CO and HCOOH on Au–Pd Core–Shell Nanoparticles. *ACS Catal.* **7**, 1673–1680 (2017). URL https://doi.org/10.1021/acscatal.6b03237. (page: 14)

[171] Chen, L. Y. *et al.* Geometrically Controlled Nanoporous PdAu Bimetallic Catalysts with Tunable Pd/Au Ratio for Direct Ethanol Fuel Cells. *ACS Catal.* **3**, 1220–1230 (2013). URL https://doi.org/10.1021/cs400135k. (page: 14)

[172] Wang, J., Chen, F., Jin, Y. & Johnston, R. L. Gold–Copper Aerogels with Intriguing Surface Electronic Modulation as Highly Active and Stable Electrocatalysts for Oxygen Reduction and Borohydride Oxidation. *ChemSusChem* **11**, 1354–1364 (2018). URL https://doi.org/10.1002/cssc.201800052. (page: 14)

[173] Yuan, L. *et al.* Gold-iridium bifunctional electrocatalyst for oxygen reduction and oxygen evolution reactions. *J. Energy Chem.* **25**, 805–810 (2016). URL https://doi.org/10.1016/j.jechem.2016.04.013. (page: 14)

[174] Li, Y. *et al.* Synthesis of core–shell Au–Pt nanodendrites with high catalytic performance via overgrowth of platinum on *in situ* gold nanoparticles. *J. Mater. Chem. A* **3**, 368–376 (2015). URL https://doi.org/10.1039/C4TA04940C. (page: 14)

[175] Kwak, K. *et al.* A molecule-like $PtAu_{24}(SC_6H_{13})_{18}$ nanocluster as an electrocatalyst for hydrogen production. *Nat. Commun.* **8**, 14723 (2017). URL https://doi.org/10.1038/ncomms14723. (page: 14)

[176] Kusada, K. *et al.* Emergence of high ORR activity through controlling local density-of-states by alloying immiscible Au and Ir. *Chem. Sci.* **10**, 652–656 (2019). URL https://doi.org/10.1039/C8SC04135K. (page: 14)

[177] Chang, F. *et al.* Composition Tunability and (111)-Dominant Facets of Ultrathin Platinum–Gold Alloy Nanowires toward Enhanced Electrocatalysis. *J. Am. Chem. Soc.* **138**, 12166–12175 (2016). URL https://doi.org/10.1021/jacs.6b05187. (page: 14)

[178] Pizzutilo, E. *et al.* Electrocatalytic synthesis of hydrogen peroxide on Au-Pd nanoparticles: From fundamentals to continuous production. *Chem. Phys. Lett.* **683**, 436–442 (2017). URL https://doi.org/10.1016/j.cplett.2017.01.071. (page: 14)

[179] Pizzutilo, E. *et al.* Addressing stability challenges of using bimetallic electrocatalysts: the case of gold–palladium nanoalloys. *Catal. Sci. Technol.* **7**, 1848–1856 (2017). URL https://doi.org/10.1039/C7CY00291B. (page: 14)

[180] Liang, Z. *et al.* Direct 12-Electron Oxidation of Ethanol on a Ternary Au(core)-PtIr(Shell) Electrocatalyst. *J. Am. Chem. Soc.* **141**, 9629–9636 (2019). URL https://doi.org/10.1021/jacs.9b03474. (page: 14)

[181] Nahar, L., Farghaly, A. A., Alan Esteves, R. J. & Arachchige, I. U. Shape Controlled Synthesis of Au/Ag/Pd Nanoalloys and Their Oxidation-Induced Self-Assembly into Electrocatalytically Active Aerogel Monoliths. *Chem. Mater.* **29**, 7704–7715 (2017). URL https://doi.org/10.1021/acs.chemmater.7b017317. (page: 14)

[182] Zhang, T. *et al.* Laser-irradiation induced synthesis of spongy AuAgPt alloy nanospheres with high-index facets, rich grain boundaries and subtle lattice distortion for enhanced electrocatalytic activity. *J. Mater. Chem. A* **6**, 13735–13742 (2018). URL https://doi.org/10.1039/C8TA04087G. (page: 14)

[183] Li, L., Chang, X., Lin, X., Zhao, Z.-J. & Gong, J. Theoretical insights into single-atom catalysts. *Chem. Soc. Rev.* **49**, 8156–8178 (2020). URL https://doi.org/10.1039/D0CS00795A. (pages: 14, 70)

[184] Mitchell, S. & Pérez-Ramírez, J. Single atom catalysis: a decade of stunning progress and the promise for a bright future. *Nat. Commun.* **11**, 4302 (2020). URL https://doi.org/10.1038/s41467-020-18182-5. (pages: 14, 70)

[185] Speck, F. D. *et al.* Single-Atom Catalysts: A Perspective toward Application in Electrochemical Energy Conversion. *JACS Au* **1**, 1086–1100 (2021). URL https://doi.org/10.1021/jacsau.1c00121. (pages: 14, 70)

[186] Zhou, X. *et al.* Unraveling Charge State of Supported Au Single-Atoms during CO Oxidation. *J. Am. Chem. Soc.* **140**, 554–557 (2018). URL https://doi.org/10.1021/jacs.7b10394. (page: 15)

[187] Ferrando, R. & Fortunelli, A. Diffusion of adatoms and small clusters on magnesium oxide surfaces. *J. Phys. Condens. Matter* **21**, 264001 (2009). URL https://doi.org/10.1088/0953-8984/21/26/264001. (pages: 15, 121)

[188] Yulikov, M., Sterrer, M., Risse, T. & Freund, H.-J. Gold atoms and clusters on MgO(100) films; an EPR and IRAS study. *Surf. Sci.* **603**, 1622–1628 (2009). URL https://doi.org/10.1016/j.susc.2008.10.056. (page: 15)

[189] Zhang, C., Michaelides, A. & Jenkins, S. J. Theory of gold on ceria. *Phys. Chem. Chem. Phys.* **13**, 22–33 (2011). URL https://doi.org/10.1039/C0CP01123A. (page: 15)

[190] Tosoni, S. & Pacchioni, G. Trends in Adhesion Energies of Gold on MgO(100), Rutile TiO$_2$(110), and CeO$_2$(111) Surfaces: A Comparative DFT Study. *J. Phys. Chem. C* **121**, 28328–28338 (2017). URL https://doi.org/10.1021/acs.jpcc.7b09429. (page: 15)

[191] Chen, Y., Hu, P., Lee, M.-H. & Wang, H. Au on (1 1 1) and (1 1 0) surfaces of CeO$_2$: A density-functional theory study. *Surf. Sci.* **602**, 1736–1741 (2008). URL https://doi.org/10.1016/j.susc.2008.02.036. (page: 15)

[192] Teng, B.-T. *et al.* A DFT Study of the Structures of Au$_x$ Clusters on a CeO$_2$(111) Surface. *ChemPhysChem* **13**, 1261–1271 (2012). URL https://doi.org/10.1002/cphc.201101007. (pages: 15, 113)

[193] Tang, Y. *et al.* Electronic states of metal (Cu, Ag, Au) atom on CeO$_2$(1 1 1) surface: The role of local structural distortion. *J. Power Sources* **197**, 28–37 (1997). URL https://doi.org/10.1016/j.jpowsour.2011.09.026. (page: 15)

[194] Zhu, K.-J. *et al.* Substrate-dependent Au$_x$ cluster: A new insight into Au$_x$/CeO$_2$. *Appl. Surf. Sci.* **387**, 557–568 (2016). URL https://doi.org/10.1016/j.apsusc.2016.06.163. (page: 15)

[195] Liu, J.-C., Wang, Y.-G. & Li, J. Toward Rational Design of Oxide-Supported Single-Atom Catalysts: Atomic Dispersion of Gold on Ceria. *J. Am. Chem. Soc.* **139**, 6190–6199 (2017). URL https://doi.org/10.1021/jacs.7b01602. (page: 15)

[196] Ming Wang, G., BelBruno, J. J., Kenny, S. D. & Smith, R. Gold adatoms and dimers on relaxed graphite surfaces. *Phys. Rev. B* **69**, 195412 (2004). URL https://doi.org/10.1103/PhysRevB.69.195412. (pages: 15, 113)

[197] Jensen, P., Blase, X. & Ordejón, P. First principles study of gold adsorption and diffusion on graphite. *Surf. Sci.* **564**, 173–178 (2004). URL https://doi.org/10.1016/j.susc.2004.06.188. (page: 15)

[198] Hardcastle, T. P. *et al.* Mobile metal adatoms on single layer, bilayer, and trilayer graphene: An *ab initio* DFT study with van der Waals corrections correlated with electron microscopy data. *Phys. Rev. B* **87**, 195430 (2013). URL https://doi.org/10.1103/PhysRevB.87.195430. (page: 15)

[199] Appy, D. *et al.* Transition metals on the (0 0 0 1) surface of graphite: Fundamental aspects of adsorption, diffusion, and morphology. *Prog. Surf. Sci.* **89**, 219–238 (2014). URL https://doi.org/10.1016/j.progsurf.2014.08.001. (page: 15)

[200] Amft, M., Lebègue, S., Eriksson, O. & Skorodumova, N. V. Adsorption of Cu, Ag, and Au atoms on graphene including van der Waals interactions. *J. Phys. Condens. Matter.* **23**, 395001 (2011). URL https://doi.org/10.1088/0953-8984/23/39/395001. (page: 15)

[201] Varns, R. & Strange, P. Stability of gold atoms and dimers adsorbed on graphene. *J. Phys. Condens. Matter.* **20**, 225005 (2008). URL https://doi.org/10.1088/0953-8984/20/22/225005. (page: 15)

[202] Chan, K. T., Neaton, J. B. & Cohen, M. L. First-principles study of metal adatom adsorption on graphene. *Phys. Rev. B* **77**, 235430 (2008). URL https://doi.org/10.1103/PhysRevB.77.235430. (page: 15)

[203] Wang, X. *et al.* Atomically dispersed Au_1 catalyst towards efficient electrochemical synthesis of ammonia. *Sci. Bull.* **63**, 1246–1253 (2018). URL https://doi.org/10.1016/j.scib.2018.07.005. (pages: 15, 70)

[204] Zhang, Q. & Guan, J. Single-Atom Catalysts for Electrocatalytic Applications. *Adv. Funct. Mater.* **30**, 2000768 (2020). URL https://doi.org/10.1002/adfm.202000768. (pages: 15, 70)

[205] Qin, Q., Heil, T., Antonietti, M. & Oschatz, M. Single-Site Gold Catalysts on Hierarchical N-Doped Porous Noble Carbon for Enhanced Electrochemical Reduction of Nitrogen. *Small Methods* **2**, 1800202 (2018). URL https://doi.org/10.1002/smtd.201800202. (pages: 15, 70)

[206] Liu, L., Su, H., Tang, F., Zhao, X. & Liu, Q. Confined organometallic Au_1N_x single-site as an efficient bifunctional oxygen electrocatalyst. *Nano Energy* **46**, 110–116 (2018). URL https://doi.org/10.1016/j.nanoen.2018.01.044. (pages: 15, 70)

[207] Bottari, F. & De Wael, K. Electrodeposition of gold nanoparticles on boron doped diamond electrodes for the enhanced reduction of small organic molecules. *J. Electroanal. Chem.* **801**, 521–526 (2017). URL https://doi.org/10.1016/j.jelechem.2017.07.053. (page: 15)

[208] Janegitz, B. C., Medeiros, R. A., Rocha-Filho, R. C. & Fatibello-Filho, O. Direct electrochemistry of tyrosinase and biosensing for phenol based on gold nanoparticles electrodeposited on a boron-doped diamond electrode. *Diam. Relat. Mater.* **25**, 128–133 (2012). URL https://doi.org/10.1016/j.diamond.2012.02.023. (page: 15)

[209] Wahyuni, W. T., Ivandini, T. A., Saepudin, E. & Einaga, Y. Development of neuraminidase detection using gold nanoparticles boron-doped diamond electrodes. *Anal. Biochem.* **497**, 68–75 (2016). URL https://doi.org/10.1016/j.ab.2015.12.003. (page: 15)

[210] Fauzillah, N. A., Abdullah, I. & Ivandini, T. A. Modification of boron-doped diamond electrode with gold nanoparticles synthesized by allyl mercaptan as the capping agent for arsenic sensors. *AIP Conf. Proc.* **2242**, 040031 (2020). URL https://doi.org/10.1063/5.0008495. (page: 15)

[211] Wei, M., Zeng, G. & Lu, Q. Determination of organophosphate pesticides using an acetylcholinesterase-based biosensor based on a boron-doped diamond electrode modified with gold nanoparticles and carbon spheres. *Microchim. Acta* **181**, 121–127 (2014). URL https://doi.org/10.1007/s00604-013-1078-4. (page: 15)

[212] Izquierdo, J., Mizaikoff, B. & Kranz, B. Surface-enhanced infrared spectroscopy on boron doped diamond modified with gold nanoparticles for spectroelectrochemical analysis. *Phys. Status Solidi A* **213**, 2056–2062 (2016). URL https://doi.org/10.1002/pssa.201600222. (page: 15)

[213] Song, M. J., Lee, S. K., Lee, J. Y., Kim, J. H. & Lim, D. S. Electrochemical sensor based on Au nanoparticles decorated boron-doped diamond electrode using ferrocene-tagged aptamer for proton detection. *J. Electroanal. Chem.* **677–680**, 139–144 (2012). URL https://doi.org/10.1016/j.jelechem.2012.05.019. (page: 15)

[214] Weng, J. *et al.* Gold-cluster sensors formed electrochemically at boron-doped-diamond electrodes: detection of dopamine in the presence of ascorbic acid

and thiols. *Adv. Funct. Mater.* **15**, 639–647 (2005). URL https://doi.org/10.1002/adfm.200400049. (page: 15)

[215] Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **389**, 457–484 (1927). URL https://doi.org/10.1002/andp.19273892002. (page: 18)

[216] Dirac, P. A. M. A new notation for quantum mechanics. *Math. Proc. Camb. Philos. Soc.* **35**, 416–418 (1939). URL http://doi.org/10.1017/S0305004100021162. (page: 18)

[217] Pauli, W. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Z. Phys.* **31**, 765–783 (1925). URL https://doi.org/10.1007/BF02980631. (page: 20)

[218] Lehtola, S., Steigemann, C., Oliveira, M. J. T. & Marques, M. A. L. Recent developments in LIBXC — A comprehensive library of functionals for density functional theory. *SoftwareX* **7**, 1–5 (2018). URL https://doi.org/10.1016/j.softx.2017.11.002. (pages: 20, 76)

[219] Perdew, J. P. & Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **577**, 1–20 (2001). URL https://doi.org/10.1063/1.1390175. (pages: v, 21, 72, 76, 93, and 98)

[220] Maurer, R. J. *et al.* Adsorption structures and energetics of molecules on metal surfaces: Bridging experiment and theory. *Prog. Surf. Sci.* **91**, 72–100 (2016). URL https://doi.org/10.1016/j.progsurf.2016.05.001. (pages: 21, 90, and 91)

[221] Wang, Z., Selbacha, S. M. & Grande, T. Van der Waals density functional study of the energetics of alkali metal intercalation in graphite. *RSC Adv.* **4**, 4069–4079 (2014). URL http://doi.org/10.1039/C3RA47187J. (page: 21)

[222] Klimeš, J. & Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **137**, 120901 (2012). URL https://doi.org/10.1063/1.4754130. (pages: 21, 24)

[223] Rohlfing, M., Temirov, R. & Tautz, F. S. Adsorption structure and scanning tunneling data of a prototype organic-inorganic interface: PTCDA on

Ag(111). *Phys. Rev. B* **76**, 115421 (2007). URL http://doi.org/10.1103/PhysRevB.76.115421. (page: 21)

[224] Romaner, K., Nabok, D., Puschnig, P., Zojer, E. & Ambrosch-Draxl, C. Theoretical study of PTCDA adsorbed on the coinage metal surfaces, Ag(111), Au(111) and Cu(111). *New J. Phys.* **11**, 053010 (2009). URL http://doi.org/10.1088/1367-2630/11/5/053010. (page: 21)

[225] Saranya, G., Nair, S., Natarajan, V. & Senthilkumar, K. Adsorption of perfluoropentacene on aluminum (1 0 0) surface: Structural and electronic properties from first principle study. *Comput. Mater. Sci.* **89**, 216–223 (2014). URL http://doi.org/10.1016/j.commatsci.2014.03.049. (page: 21)

[226] Martínez, J. I., Abad, E., Flores, F. & Ortega, J. Simulating the organic-molecule/metal interface TCNQ/Au(111). *Phys. Status Solidi B* **248**, 2044–2049 (2011). URL http://doi.org/10.1002/pssb.201147136. (page: 21)

[227] Liu, W. *et al.* Benzene adsorbed on metals: Concerted effect of covalency and van der Waals bonding. *Phys. Rev. B* **86**, 245405 (2012). URL http://doi.org/10.1103/PhysRevB.86.245405. (page: 21)

[228] Riley, K. E., Brothers, E. N., Ayers, K. B. & Merz, K. M. Accurate Atomic and Molecular Calculations without Gradient Corrections: Scaled SVWNV Density Functional. *J. Chem. Theory Comput.* **1**, 546–553 (2005). URL http://doi.org/10.1021/ct050007c. (page: 21)

[229] Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996). URL https://doi.org/10.1103/physrevlett.77.3865. (pages: iv, 22, 40, 48, 50, 57, 58, 76, 86, 91, 93, 96, 98, 99, 100, 102, 104, 105, 108, 127, 128, 155, 158, and 159)

[230] Van Noorden, R., Maher, B. & Nuzzo, R. The top 100 papers. *Nature* **514**, 550–553 (2014). URL https://doi.org/10.1038/514550a. (page: 22)

[231] Zhang, Y. & Yang, W. Comment on "Generalized Gradient Approximation Made Simple". *Phys. Rev. Lett.* **80**, 890 (1998). URL https://doi.org/10.1103/PhysRevLett.80.890. (pages: 22, 48, 57, 58, 76, 93, 96, 98, 99, 100, 102, 104, 105, and 106)

[232] Perdew, J. P. *et al.* Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **100**, 136406 (2008). URL https:

//doi.org/10.1103/PhysRevLett.100.136406. (pages: 22, 76, 94, 95, 96, 98, 99, 100, 102, 104, and 105)

[233] Hammer, B., Hansen, L. B. & Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B* **59**, 7413–7421 (1999). URL https://doi.org/10.1103/PhysRevB.59.7413. (pages: 22, 48, 57, 58, 76, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 104, 105, and 158)

[234] Tao, J., Perdew, J. P., Staroverov, V. N. & Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **91**, 14601 (2003). URL https://doi.org/10.1103/PhysRevLett.91.146401. (pages: 22, 48, 57, 58, 76, 93, 94, 96, 97, 98, 100, 101, 104, 105, and 158)

[235] Zhao, Y. & Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **125**, 194101 (2006). URL https://doi.org/10.1063/1.2370993. (pages: 22, 76, 94, 96, 104, 105, and 106)

[236] Constantin, L. A., Fabiano, E. & Della Sala, F. Semilocal dynamical correlation with increased localization. *Phys. Rev. B* **86**, 035130 (2012). URL https://doi.org/10.1103/PhysRevB.86.035130. (pages: 22, 76, 94, 96, 98, 100, 101, 104, and 105)

[237] Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **115**, 036402 (2015). URL https://doi.org/10.1103/PhysRevLett.115.036402. (pages: 22, 76, 77, 94, 96, 97, 98, 103, and 104)

[238] Sun, J., Perdew, J. P. & Ruzsinszky, A. Semilocal density functional obeying a strongly tightened bound for exchange. *Proc. Natl. Acad. Sci.* **112**, 685–689 (2015). URL https://doi.org/10.1073/pnas.1423145112. (pages: 22, 94)

[239] Perdew, J. P., Ruzsinszky, A., Sun, J. & Burke, K. Gedanken densities and exact constraints in density functional theory. *J. Chem. Phys.* **140**, 18A533 (2014). URL https://doi.org/10.1063/1.4870763. (pages: 22, 94)

[240] Lieb, E. H. & Oxford, S. Improved lower bound on the indirect Coulomb energy. *Int. J. Quantum Chem.* **19**, 427–439 (1981). URL https://doi.org/10.1002/qua.560190306. (pages: 22, 94)

[241] Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003). URL https://doi.org/10.1063/1.1564060. (pages: 22, 23, 77, 95, 96, 98, and 108)

[242] Krukau, A. V., Vydrov, O. A., Izmaylov, A. F. & Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **125**, 224106 (2006). URL https://doi.org/10.1063/1.2404663. (pages: 22, 23, 77, 95, 96, 98, 104, 105, 106, and 108)

[243] Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **110**, 6158–6170 (1999). URL https://doi.org/10.1063/1.478522. (pages: 22, 23, 77, 95, 96, 98, 104, 105, and 106)

[244] del Campo, J. M., Gázquez, J. L., Trickey, S. B. & Vela, A. Non-empirical improvement of PBE and its hybrid PBE0 for general description of molecular properties. *J. Chem. Phys.* **136**, 104108 (2012). URL https://doi.org/10.1063/1.3691197. (pages: 22, 23, 77, 93, 95, 96, 97, 98, 104, and 105)

[245] Mosquera, M. A., Borca, C. H., Ratner, M. A. & Schatz, G. C. Connection between Hybrid Functionals and Importance of the Local Density Approximation. *J. Phys. Chem. A* **120**, 1605–1612 (2016). URL https://doi.org/10.1021/acs.jpca.5b10864. (page: 22)

[246] Hui, K. & Chai, J.-D. SCAN-based hybrid and double-hybrid density functionals from models without fitted parameters. *J. Chem. Phys.* **144**, 044114 (2016). URL https://doi.org/10.1063/1.4940734. (page: 22)

[247] Balabin, R. M. Communications: Intramolecular basis set superposition error as a measure of basis set incompleteness: Can one reach the basis set limit without extrapolation? *J. Chem. Phys.* **132**, 211103 (2010). URL https://doi.org/10.1063/1.3430647. (page: 23)

[248] Morgante, P. & Peverati, R. The devil in the details: A tutorial review on some undervalued aspects of density functional theory calculations. *Int. J. Quantum Chem.* **120**, e26332 (2020). URL https://doi.org/10.1002/qua.26332. (page: 23)

[249] van Duijneveldt, F. B., van Duijneveldt-van de Rijdt, J. G. C. M. & van Lenthe, J. H. State of the Art in Counterpoise Theory. *Chem. Rev.* **94**, 1873–1885 (1994). URL https://doi.org/10.1021/cr00031a007. (page: 23)

[250] Hellmann, H. A New Approximation Method in the Problem of Many Electrons. *J. Chem. Phys.* **3**, 61 (1935). URL http://doi.org/10.1063/1.1749559. (page: 23)

[251] Pickett, W. E. Pseudopotential methods in condensed matter applications. *Comput. Phys. Rep.* **9**, 115–197 (1989). URL http://doi.org/10.1016/0167-7977(89)90002-6. (page: 23)

[252] Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994). URL http://doi.org/10.1103/PhysRevB.50.17953. (page: 23)

[253] Lejaeghere, K. *et al.* Reproducibility in density functional theory calculations of solids. *Science* **351(6280)**, aad3000 (1966). URL http://doi.org/10.1126/science.aad3000. (page: 24)

[254] Slater, J. C. Atomic Shielding Constants. *Phys. Rev.* **36**, 57–64 (1930). URL https://doi.org/10.1103/PhysRev.36.57. (page: 24)

[255] Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.* **10**, 151–177 (1957). URL https://doi.org/10.1002/cpa.3160100201. (page: 24)

[256] Lehtola, S. A review on non-relativistic, fully numerical electronic structure calculations on atoms and diatomic molecules. *Int. J. Quantum Chem.* **119**, e25968 (2019). URL https://doi.org/10.1002/qua.25968. (page: 24)

[257] Boys, S. F. Electronic wave functions - I. A general method of calculation for the stationary states of any molecular system. *Proc. R. Soc. Lond. A* **200**, 542–554 (1950). URL https://doi.org/10.1098/rspa.1950.0036. (page: 24)

[258] Jensen, F. Atomic orbital basis sets. *WIREs Comput. Mol. Sci.* **3**, 273–295 (2013). URL https://doi.org/10.1002/wcms.1123. (page: 24)

[259] Havu, V., Blum, V., Havu, P. & Scheffler, M. Efficient O(N) integration for all-electron electronic structure calculation using numeric basis functions. *J. Comput. Phys.* **228**, 8367–8379 (2009). URL https://doi.org/10.1016/j.jcp.2009.08.008. (pages: 24, 40, 48, and 76)

[260] Stöhr, M., Van Voorhis, T. & Tkatchenko, A. Theory and practice of modeling van der Waals interactions in electronic-structure calculations. *Chem. Soc.*

Rev. **48**, 4118–4154 (2019). URL https://doi.org/10.1039/C9CS00060G. (page: 24)

[261] Cohen, A. J., Mori-Sánchez, P. & Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **112**, 289–320 (2012). URL https://doi.org/10.1021/cr200107z. (page: 24)

[262] Tkatchenko, A. & Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **102**, 073005 (2009). URL https://doi.org/10.1103/PhysRevLett.102.073005. (pages: v, 24, 25, 48, 57, 58, 76, 86, 89, 91, 98, 126, and 129)

[263] Otero-de-la Roza, A. & Johnson, E. R. Van der Waals interactions in solids using the exchange-hole dipole moment model. *J. Chem. Phys.* **136**, 174109 (2012). URL https://doi.org/10.1063/1.4705760. (page: 24)

[264] Gobre, V. V. & Tkatchenko, A. Scaling laws for van der Waals interactions in nanostructured materials. *Nat. Commun.* **4**, 2341 (2012). URL https://doi.org/10.1038/ncomms3341. (page: 24)

[265] Reilly, A. M. & Tkatchenko, A. van der Waals dispersion interactions in molecular materials: beyond pairwise additivity. *Chem. Sci.* **6**, 3289–3301 (2015). URL https://doi.org/10.1039/C5SC00410A. (page: 24)

[266] Grimme, S. Density functional theory with London dispersion corrections. *WIREs Comput. Mol. Sci.* **1**, 211–228 (2011). URL https://doi.org/10.1002/wcms.30. (page: 24)

[267] Björkman, T., Gulans, A., Krasheninnikov, A. V. & Nieminen, R. M. van der Waals Bonding in Layered Compounds from Advanced Density-Functional First-Principles Calculations. *Phys. Rev. Lett.* **108**, 235502 (2012). URL https://doi.org/10.1103/PhysRevLett.108.235502. (page: 24)

[268] Ambrosetti, A., Alfè, D., DiStasio Jr., R. A. & Tkatchenko, A. Hard Numbers for Large Molecules: Toward Exact Energetics for Supramolecular Systems. *J. Phys. Chem. Lett.* **5**, 849–855 (2014). URL https://doi.org/10.1021/jz402663k. (page: 24)

[269] Hermann, J., Alfè, D. & Tkatchenko, A. Nanoscale π-π stacked molecules are bound by collective charge fluctuations. *Nat. Commun.* **8**, 14052 (2017). URL https://doi.org/10.1038/ncomms14052. (page: 24)

[270] Egger, D. A. & Kronik, L. Role of Dispersive Interactions in Determining Structural Properties of Organic–Inorganic Halide Perovskites: Insights from First-Principles Calculations. *J. Phys. Chem. Lett.* **5**, 2728–2733 (2014). URL https://doi.org/10.1021/jz5012934. (page: 24)

[271] Tkatchenko, A. Current Understanding of Van der Waals Effects in Realistic Materials. *Adv. Funct. Mater.* **25**, 2054–2061 (2015). URL https://doi.org/10.1002/adfm.201403029. (page: 24)

[272] Brémond, É., Golubev, N., Steinmann, S. N. & Corminboeuf, C. How important is self-consistency for the dDsC density dependent dispersion correction? *J. Chem. Phys.* **140**, 18A516 (2014). URL https://doi.org/10.1063/1.4867195. (page: 24)

[273] Ferri, N., DiStasio Jr., R. A., Ambrosetti, A., Car, R. & Tkatchenko, A. Electronic Properties of Molecules and Surfaces with a Self-Consistent Interatomic van der Waals Density Functional. *Phys. Rev. Lett.* **114**, 176802 (2015). URL https://doi.org/10.1103/PhysRevLett.114.176802. (page: 24)

[274] Maurer, R. J. *et al.* Thermal and Electronic Fluctuations of Flexible Adsorbed Molecules: Azobenzene on Ag(111). *Phys. Rev. Lett.* **116**, 146101 (2016). URL https://doi.org/10.1103/PhysRevLett.116.146101. (page: 24)

[275] Gao, W. & Tkatchenko, A. Electronic Structure and van der Waals Interactions in the Stability and Mobility of Point Defects in Semiconductors. *Phys. Rev. Lett.* **111**, 045501 (2013). URL https://doi.org/10.1103/PhysRevLett.111.045501. (page: 24)

[276] Egger, D. A., Kronik, L. & Rappe, A. M. Theory of Hydrogen Migration in Organic–Inorganic Halide Perovskites. *Angew. Chem. Int. Ed.* **54**, 12437–12441 (2015). URL https://doi.org/10.1002/anie.201502544. (page: 24)

[277] Fabrizio, A. & Corminboeuf, C. How do London Dispersion Interactions Impact the Photochemical Processes of Molecular Switches? *J. Phys. Chem. Lett.* **9**, 464–470 (2018). URL https://doi.org/10.1021/acs.jpclett.7b03316. (page: 24)

[278] Shtogun, Y. V. & Woods, L. M. Many-Body van der Waals Interactions between Graphitic Nanostructures. *J. Phys. Chem. Lett.* **1**, 1356–1362 (2010). URL https://doi.org/10.1021/jz100309m. (page: 24)

[279] Gao, W. & Tkatchenko, A. Sliding Mechanisms in Multilayered Hexagonal Boron Nitride and Graphene: The Effects of Directionality, Thickness, and Sliding Constraints. *Phys. Rev. Lett.* **114**, 096101 (2015). URL https://doi.org/10.1103/PhysRevLett.114.096101. (page: 24)

[280] Dion, M., Rydberg, H., Schröder, E., Langreth, D. C. & Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **92**, 246401 (2004). URL https://doi.org/10.1103/PhysRevLett.92.246401. (page: 24)

[281] Shukla, V. *et al.* Accurate Nonempirical Range-Separated Hybrid van der Waals Density Functional for Complex Molecular Problems, Solids, and Surfaces. *Phys. Rev. X* **12**, 041003 (2022). URL https://doi.org/10.1103/PhysRevX.12.041003. (page: 24)

[282] Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010). URL https://doi.org/10.1063/1.3382344. (pages: 25, 125, 126, and 127)

[283] Chu, X. & Dalgarno, A. Linear response time-dependent density functional theory for van der Waals coefficients. *J. Chem. Phys.* **121**, 4083–4088 (2004). URL https://doi.org/10.1063/1.1779576. (page: 25)

[284] Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theoret. Chim. Acta* **44**, 129–138 (1977). URL https://doi.org/10.1007/BF00549096. (pages: 25, 82, and 129)

[285] Tkatchenko, A., DiStasio Jr., R. A., Car, R. & Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **108**, 236402 (2012). URL https://doi.org/10.1103/PhysRevLett.108.236402. (pages: 26, 58, 87, 126, and 127)

[286] Tkatchenko, A., Ambrosetti, A. & DiStasio Jr., R. A. Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem. *J. Chem. Phys.* **138**, 074106 (2013). URL https://doi.org/10.1063/1.4789814. (pages: 26, 58, 87, 126, and 127)

[287] Ambrosetti, A., Reilly, A. M., DiStasio Jr., R. A. & Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **140**, 18A508 (2014). URL https://doi.org/10.1063/1.4865104. (pages: 26, 48, 58, 87, 89, 127, 128, and 129)

[288] Hermann, J. & Tkatchenko, A. Density Functional Model for van der Waals Interactions: Unifying Many-Body Atomic Approaches with Nonlocal Functionals. *Phys. Rev. Lett.* **124**, 146401 (2020). URL https://doi.org/10.1103/PhysRevLett.124.146401. (pages: 26, 27, 48, 58, 87, 89, and 91)

[289] Maurer, R. J., Ruiz, V. G. & Tkatchenko, A. Many-body dispersion effects in the binding of adsorbates on metal surfaces. *J. Chem. Phys.* **143**, 102808 (2015). URL https://doi.org/10.1063/1.4922688. (pages: 26, 94, and 125)

[290] Vydrov, O. A. & Van Voorhis, T. Dispersion interactions from a local polarizability model. *Phys. Rev. A* **81**, 062708 (2010). URL https://doi.org/10.1103/PhysRevA.81.062708. (page: 27)

[291] Casimir, H. B. G. & Polder, D. The Influence of Retardation on the London-van der Waals Forces. *Phys. Rev.* **73**, 360–372 (1948). URL https://doi.org/10.1103/PhysRev.73.360. (page: 28)

[292] McLachlan, A. D. Retarded dispersion forces between molecules. *Proc. R. Soc. A* **271**, 387–401 (1963). URL https://doi.org/10.1098/rspa.1963.0025. (page: 28)

[293] Seifert, G., Porezag, D. & Frauenheim, T. Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme. *Int. J. Quantum Chem.* **58**, 185–192 (1996). URL https://doi.org/10.1002/(SICI)1097-461X(1996)58:2<185::AID-QUA7>3.0.CO;2-U. (page: 28)

[294] Porezag, D., Frauenheim, T., Köhler, T., Seifert, G. & Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B* **51**, 12947 (1995). URL https://doi.org/10.1103/PhysRevB.51.12947. (pages: 28, 29)

[295] Yang, Y., Yu, H., York, D., Cui, Q. & Elstner, M. Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Theory Total Energy and Introduction of a Modified Effective Coulomb Interaction. *J. Phys. Chem. A* **111**, 10861–10873 (2007). URL https://doi.org/10.1021/jp074167r. (page: 28)

[296] Gaus, M., Cui, Q. & Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **7**, 931–948 (2011). URL https://doi.org/10.1021/ct100684s. (page: 28)

[297] Gaus, M., Chou, C.-P., Witek, H. & Elstner, M. Automatized Parametrization of SCC-DFTB Repulsive Potentials: Application to Hydrocarbons. *J. Phys. Chem. A* **113**, 11866–11881 (2009). URL https://doi.org/10.1021/jp902973m. (page: 29)

[298] Sauer, J. Ab Initio Calculations for Molecule–Surface Interactions with Chemical Accuracy. *Acc. Chem. Res.* **52**, 3502–3510 (2019). URL http://doi.org/10.1021/acs.accounts.9b00506. (page: 31)

[299] Schneider, J., Ribeiro, R., Alfonso-Prieto, M., Carloni, P. & Giorgetti, A. Hybrid MM/CG Webserver: Automatic Set Up of Molecular Mechanics/Coarse-Grained Simulations for Human G Protein-Coupled Receptor/Ligand Complexes. *Front. Mol. Biosci.* **7**, 576689 (2020). URL http://doi.org/10.3389/fmolb.2020.576689. (page: 31)

[300] Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976). URL https://doi.org/10.1016/0022-2836(76)90311-9. (pages: 31, 72)

[301] Lu, Y. *et al.* Open-Source, Python-Based Redevelopment of the ChemShell Multiscale QM/MM Environment. *J. Chem. Theory Comput.* **15**, 1317–1328 (2019). URL https://doi.org/10.1021/acs.jctc.8b01036. (pages: 31, 41, 51, 72, 73, 76, 78, and 161)

[302] Martin Senn, H. & Walter, T. QM/MM Methods for Biomolecular Systems. *Surf. Sci.* **48**, 1198–1229 (2009). URL https://doi.org/10.1002/anie.200802019. (pages: 31, 32)

[303] Sherwood, P. *et al.* QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. THEOCHEM* **632**, 1–28 (2003). URL https://doi.org/10.1016/S0166-1280(03)00285-9. (pages: 32, 41)

[304] Cao, L. & Ryde, U. On the Difference Between Additive and Subtractive QM/MM Calculations. *Front. Chem.* **6**, 89 (2018). URL https://doi.org/10.3389/fchem.2018.00089. (pages: 32, 73)

[305] Duarte, F., Amrein, B. A., Blaha-Nelson, D. & Kamerlin, S. C. L. Recent advances in QM/MM free energy calculations using reference potentials. *Biochim. Biophys.* **1850**, 954–965 (2015). URL http://doi.org/10.1016/j.bbagen.2014.07.008. (page: 32)

[306] Gao, J., Amara, P., Alhambra, C. & Field, M. J. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A* **102**, 4714–4721 (1998). URL https://doi.org/10.1021/jp9809890. (page: 32)

[307] Murphy, R. B., Philipp, D. M. & Friesner, R. A. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comp. Chem.* **21**, 1442–1457 (2000). URL https://doi.org/10.1002/1096-987X(200012)21:16<1442::AID-JCC3>3.0.CO;2-O. (page: 32)

[308] Field, M. J., Bash, P. A. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comp. Chem.* **11**, 700–733 (1990). URL https://doi.org/10.1002/jcc.540110605. (page: 32)

[309] Reuter, N., Dejaegere, A., Maigret, B. & Karplus, M. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem. A* **104**, 1720–1735 (2000). URL https://doi.org/10.1021/jp9924124. (pages: 32, 73)

[310] Metz, S., Kästner, J., Sokol, A. A., Keal, T. W. & Sherwood, P. ChemShell—a modular software package for QM/MM simulations. *WIREs Comput. Mol. Sci.* **4**, 101–110 (2014). URL https://doi.org/10.1002/wcms.1163. (pages: 32, 41)

[311] Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K. & Schölkopf, B. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw.* **12**, 181–201 (2001). URL https://doi.org/10.1109/72.914517. (pages: 33, 34)

[312] Vapnik, V. Principles of Risk Minimization for Learning Theory. In Moody, J. E., Hanson, S. J. & Lippmann, R. P. (eds.) NIPS'91: Proceedings of the 4th International Conference on Neural Information Processing Systems, 831–838 (Morgan Kaufmann Publishers Inc., 1991). URL https://dl.acm.org/doi/10.5555/2986916.2987018. (page: 34)

[313] Sidiropoulou, K., Kyriaki Pissadaki, E. & Poirazi, P. Inside the brain of a neuron. *EMBO Rep.* **7**, 886–892 (2006). URL https://doi.org/10.1038/sj.embor.7400789. (page: 34)

[314] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015). URL https://doi.org/10.1038/nature14539. (page: 34)

[315] Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018). URL https://doi.org/10.1063/1.5019779. (pages: 36, 37, 126, 128, and 160)

[316] Schütt, K. T. *et al.* SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **15**, 448–455 (2019). URL https://doi.org/10.1021/acs.jctc.8b00908. (pages: 36, 37, and 128)

[317] Unke, O. T. *et al.* SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* **12**, 7273 (2021). URL https://doi.org/10.1038/s41467-021-27504-0. (pages: 36, 126, and 154)

[318] Unke, O. T. & Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019). URL https://doi.org/10.1021/acs.jctc.9b00181. (pages: 36, 37, and 126)

[319] Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **115**, 1032–1050 (2015). URL https://doi.org/10.1002/qua.24890. (pages: 36, 130)

[320] Behler, J. & Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007). URL https://doi.org/10.1103/PhysRevLett.98.146401. (page: 36)

[321] Himanen, L. *et al.* DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020). URL http://doi.org/10.1016/j.cpc.2019.106949. (pages: 36, 37)

[322] Wolpert, D. H. & Macready, W. G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1**, 67–82 (1997). URL http://doi.org/10.1109/4235.585893. (page: 36)

[323] Faber, F., Lindmaa, A., Anatole von Lilienfeld, O. & Armiento, R. Crystal structure representations for machine learning models of formation energies. *Int. J. Quant. Chem.* **115**, 1094–1101 (2015). URL http://doi.org/10.1002/qua.24917. (page: 36)

[324] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988). URL http://doi.org/10.1021/ci00057a005. (page: 36)

[325] Collins, C. R., Gordon, G. J., Anatole von Lilienfeld, O. & Yaron, D. J. Constant size descriptors for accurate machine learning models of molecular properties. *J. Chem. Phys.* **148**, 241718 (2018). URL http://doi.org/10.1063/1.5020441. (page: 36)

[326] Hansen, K. *et al.* Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **9**, 3404–3419 (2013). URL http://doi.org/10.1021/ct400195d. (page: 36)

[327] Hansen, K. *et al.* Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **6**, 2306–2331 (2015). URL http://doi.org/10.1021/acs.jpclett.5b00831. (page: 36)

[328] Huang, B. & Anatole von Lilienfeld, O. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *J. Chem. Phys.* **145**, 161102 (2016). URL http://doi.org/10.1063/1.4964627. (page: 36)

[329] Rupp, M., Tkatchenko, A., Müller, K.-R. & Anatole von Lilienfeld, O. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **108**, 058301 (2012). URL http://doi.org/10.1103/PhysRevLett.108.058301. (page: 36)

[330] Anatole von Lilienfeld, O., Ramakrishnan, R., Rupp, M. & Knoll, A. Fourier series of atomic radial distribution functions: A molecular fingerprint for machine learning models of quantum chemical properties. *Int. J. Quant. Chem.* **115**, 1084–1093 (2015). URL http://doi.org/10.1002/qua.24912. (page: 36)

[331] Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013). URL http://doi.org/10.1103/PhysRevB.87.184115. (page: 36)

[332] Bartók, A. P., Kondor, R. & Csányi, G. Publisher's Note: On representing chemical environments [Phys. Rev. B 87, 184115 (2013)]. *Phys. Rev. B* **87**, 219902 (2013). URL http://doi.org/10.1103/PhysRevB.87.219902. (page: 36)

[333] Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017). URL http://doi.org/10.1038/ncomms13890. (pages: 37, 126, 155, and 160)

[334] De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016). URL http://doi.org/10.1039/C6CP00415F. (page: 37)

[335] Reuter, K. & Scheffler, M. Composition, structure, and stability of $RuO_2(110)$ as a function of oxygen pressure. *Phys. Rev. B* **65**, 035406 (2001). URL https://doi.org/10.1103/PhysRevB.65.035406. (pages: 37, 38, 39, 48, 49, 54, 57, and 157)

[336] Chase, M. W. NIST-JANAF Thermochemical Tables. *J. Chem. Phys. Ref. Data* **Monograph 9**, 1–1951 (1998). URL https://doi.org/10.18434/T42S31. (pages: x, 39, 49, and 50)

[337] Gale, J. D. GULP: A computer program for the symmetry-adapted simulation of solids. *JCS Faraday Trans.* **93**, 629–637 (1997). URL https://doi.org/10.1039/A606455H. (pages: 41, 73, and 77)

[338] Gale, J. D. & Rohl, A. L. The General Utility Lattice Program (GULP). *Mol. Simul.* **29**, 291–341 (2003). URL https://doi.org/10.1080/0892702031000104887. (pages: 41, 73, and 77)

[339] Larsen, A. H. *et al.* The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017). URL https://doi.org/10.1088/1361-648X/aa680e. (pages: 41, 76, 128, 130, 131, and 161)

[340] Kästner, J. *et al.* DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **113**, 11856–11865 (2009). URL https://doi.org/10.1021/jp9028968. (pages: 45, 72)

[341] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (page: 45)

[342] Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2007). URL https://doi.org/10.1038/s41586-020-2649-2. (page: 45)

[343] Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **23**, 2825–2830 (2011). URL http://jmlr.org/papers/v12/pedregosa11a.html. (page: 45)

[344] Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261–272 (2020). URL http://doi.org/10.1038/s41592-019-0686-2. (page: 45)

[345] Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **3**, 90–95 (2007). URL https://doi.org/10.1109/MCSE.2007.55. (page: 45)

[346] Logsdail, A. J. *et al.* *Cardiff Molecular Modelling* (accessed November 28, 2022). https://github.com/logsdail/carmm. (pages: 45, 76)

[347] Chaudhuri, S. *et al.* Coexistence of carbonyl and ether groups on oxygen-terminated (110)-oriented diamond surfaces. *Commun. Mater.* **3**, 6 (2022). URL https://doi.org/10.1038/s43246-022-00228-4. (page: 46)

[348] Hutton, L. A. *et al.* Examination of the Factors Affecting the Electrochemical Performance of Oxygen-Terminated Polycrystalline Boron-Doped Diamond Electrodes. *Anal. Chem.* **85**, 7230–7240 (2013). URL https://doi.org/10.1021/ac401042t. (pages: 46, 47, and 52)

[349] Fischer, A. E., Show, Y. & Swain, G. M. Electrochemical Performance of Diamond Thin-Film Electrodes from Different Commercial Sources. *Anal. Chem.* **76**, 2553–2560 (2004). URL https://doi.org/10.1021/ac035214o. (page: 46)

[350] Martin, H. B., Argoitia, A., Landau, U., Anderson, A. B. & Angus, J. C. Hydrogen and Oxygen Evolution on Boron-Doped Diamond Electrodes. *J. Electrochem. Soc.* **143**, L133–L136 (1996). URL https://doi.org/10.1149/1.1836901. (page: 46)

[351] Yagi, I., Notsu, H., Kondo, T., Tryk, D. A. & Fujishima, A. Electrochemical selectivity for redox systems at oxygen-terminated diamond electrodes. *J. Electroanal. Chem.* **473**, 173–178 (1999). URL https://doi.org/10.1016/S0022-0728(99)00027-3. (page: 46)

[352] Maier, F., Ristein, J. & Ley, L. Electron affinity of plasma-hydrogenated and chemically oxidized diamond (100) surfaces. *Phys. Rev. B* **143**, 165411 (2001). URL https://doi.org/10.1103/PhysRevB.64.165411. (page: 47)

[353] Liu, F. B., Wang, J. D., Liu, B., Li, X. M. & Chen, D. R. Effect of electronic structures on electrochemical behaviors of surface-terminated boron-doped diamond film electrodes. *Diam. Relat. Mater.* **16**, 454–460 (2007). URL https://doi.org/10.1016/j.diamond.2006.08.016. (page: 47)

[354] Loh, K. P., Xie, X. N., Yang, S. W. & Zheng, J. C. Oxygen Adsorption on (111)-Oriented Diamond: A Study with Ultraviolet Photoelectron Spectroscopy, Temperature-Programmed Desorption, and Periodic Density Functional Theory. *J. Phys. Chem. B* **106**, 5230–5240 (2002). URL https://doi.org/10.1021/jp0139437. (page: 47)

[355] Zheng, J., Xie, X. N., Wee, A. T. S. & Loh, K. P. Oxygen-induced surface state on diamond (100). *Diamond Relat. Mater.* **10**, 500–505 (2001). URL https://doi.org/10.1016/S0925-9635(00)00439-8. (page: 47)

[356] Whitten, J. L., Cremaschi, P., Thomas, R. E., Rudder, R. A. & Markunas, R. J. Effects of oxygen on surface reconstruction of carbon. *Appl. Surf. Sci.* **75**, 45–50 (1994). URL https://doi.org/10.1016/0169-4332(94)90134-1. (page: 47)

[357] Zheng, X. M. & Smith, P. V. The stable configurations for oxygen chemisorption on the diamond (100) and (111) surfaces. *Surf. Sci.* **262**, 219–234 (1992). URL https://doi.org/10.1016/0039-6028(92)90473-J. (page: 47)

[358] Tamura, H. *et al.* Periodic density-functional study on oxidation of diamond (100) surfaces. *Phys. Rev. B* **61**, 11025–11033 (2000). URL https://doi.org/10.1103/PhysRevB.61.11025. (page: 47)

[359] Skokov, S., Weiner, B. & Frenklach, M. Molecular-dynamics study of oxygenated (100) diamond surfaces. *Phys. Rev. B* **49**, 11374–11382 (1993). URL https://doi.org/10.1103/PhysRevB.49.11374. (page: 47)

[360] Petrini, D. & Larsson, K. A Theoretical Study of the Energetic Stability and Geometry of Hydrogen- and Oxygen-Terminated Diamond (100) Surfaces. *J. Phys. Chem. C* **111**, 795–801 (2007). URL https://doi.org/10.1021/jp063383h. (page: 47)

[361] Derry, T. E., Makau, N. W. & Stampfl, C. Oxygen adsorption on the (1×1) and (2×1) reconstructed C(111) surfaces: a density functional theory study. *J. Phys. Condens. Matter* **22**, 265007 (2010). URL https://doi.org/10.1088/0953-8984/22/26/265007. (page: 47)

[362] Larsson, K. The Combined Influence of Dopant Species and Surface Termination on the Electronic Properties of Diamond Surfaces. *C.* **6**, 22 (2020). URL https://doi.org/10.3390/c6020022. (page: 47)

[363] Hassan, M. M. & Larsson, K. Effect of Surface Termination on Diamond (100) Surface Electrochemistry. *J. Phys. Chem. C* **118**, 22995–23002 (2014). URL https://doi.org/10.1021/jp500685q. (page: 47)

[364] Wan, G., Cattelan, M. & Fox, N. A. Electronic Structure Tunability of Diamonds by Surface Functionalization. *J. Phys. Chem. C* **123**, 4168–4177 (2019). URL https://doi.org/10.1021/acs.jpcc.8b11232. (page: 47)

[365] Struck, L. M. & D'Evelyn, M. P. Interaction of hydrogen and water with diamond (100): Infrared spectroscopy. *J. Vac. Sci. Technol. A* **11**, 1992–1997 (1993). URL https://doi.org/10.1116/1.578397. (page: 47)

[366] Pehrsson, P. E. & Mercer, T. W. Oxidation of the hydrogenated diamond (100) surface. *Surf. Sci.* **460**, 49–66 (2000). URL https://doi.org/10.1016/S0039-6028(00)00494-5. (page: 47)

[367] Pehrsson, P. E. & Mercer, T. W. Oxidation of heated diamond C(100):H surfaces. *Surf. Sci.* **460**, 74–90 (2000). URL https://doi.org/10.1016/S0039-6028(00)00495-7. (page: 47)

[368] Takeuchi, D., Ri, S.-G., Tokuda, N. & Yamasaki, S. Recovery of negative electron affinity by annealing on (111) oxidized diamond surfaces. *Diam. Relat. Mater.* **18**, 206–209 (2009). URL https://doi.org/10.1016/j.diamond.2008.10.007. (page: 47)

[369] Makau, N. W. & Derry, T. E. Study of Oxygen on the Three Low Index Diamond Surfaces by XPS. *Surf. Rev. Lett.* **10**, 295–301 (2003). URL https://doi.org/10.1142/S0218625X03005189. (pages: 47, 54, 57, 62, 63, 65, 68, and 157)

[370] Baldwin, C. G., Downes, J. E., McMahon, C. J., Bradac, C. & Mildren, R. P. Nanostructuring and oxidation of diamond by two-photon ultraviolet surface excitation: An XPS and NEXAFS study. *Phys. Rev. B* **89**, 195422 (2014). URL https://doi.org/10.1103/PhysRevB.89.195422. (pages: 47, 54, 57, 62, 63, 68, and 157)

[371] Bobrov, K., Shechter, H., Hoffman, A. & Folman, M. Molecular oxygen adsorption and desorption from single crystal diamond (111) and (110) surfaces. *Appl. Surf. Sci.* **196**, 173–180 (2002). URL https://doi.org/10.1016/S0169-4332(02)00053-3. (pages: 47, 54, 57, 64, 68, and 157)

[372] Mercer, T. W., Russell, J. N. & Pehrsson, P. E. The effect of a hydrogen plasma on the diamond (110) surface. *Surf. Sci.* **392**, L21–L26 (1997). URL https://doi.org/10.1016/S0039-6028(97)00680-8. (pages: 47, 66, and 68)

[373] Reuter, K. & Scheffler, M. Composition and structure of the $RuO_2(110)$ surface in an O_2 and CO environment: Implications for the catalytic formation of CO_2. *Phys. Rev. B* **68**, 045407 (2003). URL https://doi.org/10.1103/PhysRevB.68.045407. (pages: 48, 57, and 157)

[374] Reuter, K., Stampf, C. & Scheffler, M. Ab Initio Atomistic Thermodynamics and Statistical Mechanics of Surface Properties and Functions. In Yip, S. (ed.) Handbook of Materials Modeling, 149–194 (Springer, 2005). URL https://doi.org/10.1007/978-1-4020-3286-8_10. (pages: 48, 49, and 54)

[375] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988). URL https://doi.org/10.1103/PhysRevA.38.3098. (pages: 48, 57, and 58)

[376] Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988). URL https://doi.org/10.1103/PhysRevB.37.785. (pages: 48, 57, and 58)

[377] Chaudhuri, S., Hall, S. J., Klein, B. P. & Maurer, R. J. *Oxygen on Diamond(110) dataset.* URL https://doi.org/10.17172/NOMAD/2021.03.01-1. (page: 48)

[378] Nečas, D. & Klapetek, P. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.* **2**, 036001 (2019). URL https://doi.org/10.1088/2515-7639/ab13bb. (pages: 48, 128)

[379] Jones, R. O. & Gunnarsson, O. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.* **61**, 689–746 (1989). URL https://doi.org/10.1103/RevModPhys.61.689. (page: 50)

[380] Hellman, A., Razaznejad, B. & Lundqvist, B. I. Potential-energy surfaces for excited states in extended systems. *J. Chem. Phys.* **120**, 4593–4602 (2004). URL https://doi.org/10.1063/1.1645787. (page: 50)

[381] Mizoguchi, T., Tanaka, I., Gao, S.-P. & Pickard, C. J. First-principles calculation of spectral features, chemical shift and absolute threshold of ELNES and XANES using a plane wave pseudopotential method. *J. Phys. Condens. Matter* **21**, 104204 (2009). URL https://doi.org/10.1088/0953-8984/21/10/104204. (page: 50)

[382] Kahk, J. M. & Lischner, J. Accurate absolute core-electron binding energies of molecules, solids, and surfaces from first-principles calculations. *Phys. Rev. Mater.* **3**, 100801(R) (2019). URL https://doi.org/10.1103/PhysRevMaterials.3.100801. (page: 51)

[383] Gilbert, A. T. B., Besley, N. A. & Gill, P. M. W. Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM). *J. Phys. Chem. A* **112**, 13164–13171 (2008). URL http://doi.org/10.1021/jp801738f. (page: 51)

[384] Mulliken, R. S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *J. Chem. Phys.* **23**, 1833–1840 (1955). URL https://doi.org/10.1063/1.1740588. (pages: 51, 81, 83, and 84)

[385] Klein, B. P., Hall, S. J. & Maurer, R. J. The Nuts and Bolts of Ab-Initio Core-Hole Simulations for K-shell X-Ray Photoemission and Absorption Spectra. *J. Phys. Condens. Matter* **33**, 154005 (2021). URL https://doi.org/10.1088/1361-648X/abdf00. (page: 51)

[386] Quaino, P. M. & Schmickler, W. Oxygen-Terminated Diamond Electrodes in Alkaline Media: Structure and OH Generation. *ChemElectroChem* **1**, 933–939 (2014). URL https://doi.org/10.1002/celc.201300200. (pages: 52, 71)

[387] Schmid, M., Steinrück, H.-P. & Gottfried, J. M. A new asymmetric Pseudo-Voigt function for more efficient fitting of XPS lines. *Surf. Interface Anal.* **46**, 505–511 (2014). URL https://doi.org/10.1002/sia.5521. (page: 53)

[388] Schmid, M., Steinrück, H.-P. & Gottfried, J. M. A new asymmetric Pseudo-Voigt function for more efficient fitting of XPS lines. *Surf. Interface Anal.* **47**, 1080 (2015). URL https://doi.org/10.1002/sia.5847. (page: 53)

[389] Nečas, D. & Klapetek, P. Gwyddion: an open-source software for SPM data analysis. *Cent. Eur. J. Phys.* **10**, 181–188 (2012). URL https://doi.org/10.2478/s11534-011-0096-2. (page: 53)

[390] Cobb, S. J. *et al.* Assessment of acid and thermal oxidation treatments for removing sp^2 bonded carbon from the surface of boron doped diamond. *Carbon* **167**, 1–10 (2020). URL https://doi.org/10.1016/j.carbon.2020.04.095. (page: 59)

[391] Kononenko, V. V. *et al.* Effect of the pulse duration on graphitisation of diamond during laser ablation. *Quant. Electron.* **35**, 252–256 (2005). URL https://doi.org/10.1070/QE2005v035n03ABEH002900. (page: 59)

[392] Osswald, S., Yushin, G., Mochalin, V., Kucheyev, S. O. & Gogotsi, Y. Control of sp^2/sp^3 Carbon Ratio and Surface Chemistry of Nanodiamond Powders by Selective Oxidation in Air. *J. Am. Chem. Soc.* **128**, 11635–11642 (2006). URL https://doi.org/10.1021/ja063303n. (page: 59)

[393] Musket, R. G., McLean, W., Colmenares, C. A., Makowiecki, D. M. & Siekhaus, W. J. Preparation of atomically clean surfaces of selected elements: A review. *Appl. Surf. Sci.* **10**, 143–207 (1982). URL https://doi.org/10.1016/0378-5963(82)90142-8. (page: 63)

[394] Nishitano-Gamo, M. *et al.* Surface morphology of homoepitaxially grown (111), (001), and (110) diamond studied by low energy electron diffraction and reflection high-energy electron diffraction. *J. Vac. Sci. Technol. A* **17**, 2991–3002 (1999). URL https://doi.org/10.1116/1.581972. (pages: 66, 68)

[395] Maier, F. *et al.* The hydrogenated and bare diamond (110) surface: a combined LEED-, XPS-, and ARPES study. *Surf. Sci.* **443**, 177–185 (1999). URL https://doi.org/10.1016/S0039-6028(99)01010-9. (pages: 66, 68)

[396] Bobrov, K., Shechter, H., Folman, M. & Hoffman, A. A study of deuterium interaction with diamond (110) single crystal surface by TPD, EELS and LEED. *Diam. Relat. Mater.* **8**, 705–711 (1999). URL https://doi.org/10.1016/S0925-9635(98)00329-X. (pages: 66, 68)

[397] Sutcu, L. F. *et al.* Atomic force microscopy of (100), (110), and (111) homoepitaxial diamond films. *J. Appl. Phys.* **71**, 5930–5940 (1992). URL https://doi.org/10.1063/1.350443. (page: 68)

[398] Milchev, A., Stoyanov, S. & Kaischev, R. Atomistic theory of electrolytic nucleation: I. *Thin Solid Films* **22**, 255–265 (1974). URL https://doi.org/10.1016/0040-6090(74)90296-X. (page: 70)

[399] Milchev, A. & Malinowski, J. Phase formation — Stability and nucleation kinetics of small clusters. *Surf. Sci.* **156**, 36–43 (1985). URL https://doi.org/10.1016/0039-6028(85)90555-2. (page: 70)

[400] Liu, Y., Gokcen, D., Bertocci, U. & Moffat, T. P. Self-Terminating Growth of Platinum Films by Electrochemical Deposition. *Science* **338**, 1327–1330 (2012). URL https://doi.org/10.1126/science.1228925. (page: 70)

[401] Zhou, M., Dick, J. E. & Bard, A. J. Electrodeposition of Isolated Platinum Atoms and Clusters on Bismuth—Characterization and Electrocatalysis. *J. Am. Chem. Soc.* **139**, 17677–17682 (2017). URL https://doi.org/10.1021/jacs.7b10646. (page: 70)

[402] Gambardella, P. *et al.* Giant Magnetic Anisotropy of Single Cobalt Atoms and Nanoparticles. *Science* **300**, 1130–1133 (2003). URL https://doi.org/10.1126/science.1082857. (page: 70)

[403] The Anh, L., Catalan, F. C. I., Kim, Y., Einaga, Y. & Tateyama, Y. Boron position-dependent surface reconstruction and electronic states of boron-doped diamond(111) surfaces: an *ab initio* study. *Phys. Chem. Chem. Phys.* **23**, 15628–15634 (2021). URL https://doi.org/10.1039/D1CP00689D. (page: 71)

[404] Dec, B., Sobaszek, M., Jaramillo-Botero, A., Goddard III, W. A. & Bogdanowicz, R. Ligand-Modified Boron-Doped Diamond Surface: DFT Insights into the Electronic Properties of Biofunctionalization. *Materials (Basel)* **12**, 2910 (2019). URL https://doi.org/10.3390/ma12182910. (page: 71)

[405] Crispin, X., Bureau, C., Geskin, V., Lazzaroni, R. & Brédas, J.-L. Local Density Functional Study of Copper Clusters: A Comparison between Real Clusters, Model Surface Clusters, and the Actual Metal Surface. *Eur. J. Inorg. Chem.* **12**, 349–360 (1999). URL https://doi.org/10.1002/(SICI)1099-0682(19990202)1999:2<349::AID-EJIC349>3.0.CO;2-%23. (page: 71)

[406] Berger, D. *et al.* Embedded-cluster calculations in a numeric atomic orbital density-functional theory framework. *J. Chem. Phys.* **141**, 024105 (2014). URL https://doi.org/10.1063/1.4885816. (pages: 71, 72)

[407] Engel, J., Francis, S. & Roldan, A. The influence of support materials on the structural and electronic properties of gold nanoparticles – a DFT study. *Phys. Chem. Chem. Phys.* **21**, 19011–19025 (2019). URL https://doi.org/10.1039/C9CP03066B. (pages: 71, 74, 111, 112, 113, 114, 121, 147, and 148)

[408] Pederson, R. *et al.* Large Scale Quantum Chemistry with Tensor Processing Units. *J. Chem. Theory Comput.* (2022). URL https://doi.org/10.1021/acs.jctc.2c00876. (pages: 71, 125)

[409] Kirchhoff, B. *et al.* Assessment of the Accuracy of Density Functionals for Calculating Oxygen Reduction Reaction on Nitrogen-Doped Graphene. *J. Chem. Theory Comput.* **17**, 6405–6415 (2021). URL http://doi.org/10.1021/acs.jctc.1c00377. (page: 71)

[410] Wu, H. B. & Lou, X. W. Metal-organic frameworks and their derived materials for electrochemical energy storage and conversion: Promises and challenges. *Sci. Adv.* **3**, eaap925 (2017). URL http://doi.org/10.1126/sciadv.aap9252. (page: 71)

[411] Duan, A. & Henkelman, G. O_2 activation at the Au/MgO(001) interface boundary facilitates CO oxidation. *Phys. Chem. Chem. Phys.* **18**, 5486–5490 (2016). URL https://doi.org/10.1039/C5CP05558J. (page: 74)

[412] Coquet, R., Howard, K. L. & Willock, D. J. Theory and simulation in heterogeneous gold catalysis. *Chem. Soc. Rev.* **37**, 2046–2076 (2008). URL https://doi.org/10.1039/B707385M. (page: 74)

[413] Chen, M. S. & Goodman, D. W. Structure–activity relationships in supported Au catalysts. *Catal. Today* **111**, 22–33 (2008). URL https://doi.org/10.1016/j.cattod.2005.10.007. (page: 74)

[414] Walker, J. Optical absorption and luminescence in diamond. *Rep. Prog. Phys.* **42**, 1605–1659 (1979). URL https://doi.org/10.1088/0034-4885/42/10/001. (page: 74)

[415] Lombardi, E. B. & Mainwood, A. A first principles study of lithium, sodium and aluminum in diamond. *Diam. Relat. Mater.* **17**, 1349–1352 (2008). URL https://doi.org/10.1016/j.diamond.2007.12.015. (page: 74)

[416] Mori, M. *et al.* Electronic Structures of Aluminum-Doped Diamond near the Fermi Level. *J. Phys. Soc. Jpn.* **84**, 044704 (2015). URL https://doi.org/10.7566/JPSJ.84.044704. (pages: 74, 75)

[417] Jin Lee, Y. *et al.* Micromachining of ferrous metal with an ion implanted diamond cutting tool. *Carbon* **152**, 598–608 (2019). URL https://doi.org/10.1016/j.carbon.2019.06.043. (page: 75)

[418] Hao, L., Manzhos, S. & Zhang, Z. Theoretical Insight Into Diamond Doping and Its Possible Effect on Diamond Tool Wear During Cutting of Steel. *Front. Mater.* **8**, 1–8 (2021). URL https://doi.org/10.3389/fmats.2021.806466. (page: 75)

[419] Zeng, C., Shen, J. & Zhang, J. High thermal conductivity in indium-based metal/diamond composites by good wettability of diamond with indium. *Diam. Relat. Mater.* **112**, 108230 (2021). URL https://doi.org/10.1016/j.diamond.2020.108230. (page: 75)

[420] Ying Zhang, I. *et al.* Main-group test set for materials science and engineering with user-friendly graphical tools for error analysis: systematic benchmark of the numerical and intrinsic errors in state-of-the-art electronic-structure approximations. *New J. Phys.* **21**, 013025 (2019). URL https://doi.org/10.1088/1367-2630/aaf751. (page: 76)

[421] Groth, S. *et al.* *Ab initio* Exchange-Correlation Free Energy of the Uniform Electron Gas at Warm Dense Matter Conditions. *Phys. Rev. Lett.* **119**, 135001 (2017). URL https://doi.org/10.1103/PhysRevLett.119.135001. (pages: 76, 93, and 96)

[422] Karasiev, V. V., Sjostrom, T., Dufty, J. & Trickey, S. B. Accurate Homogeneous Electron Gas Exchange-Correlation Free Energy for Local Spin-Density Calculations. *Phys. Rev. Lett.* **112**, 076403 (2014). URL https://doi.org/10.1103/PhysRevLett.112.076403. (pages: 76, 93, 96, 98, 99, and 100)

[423] Ceperley, D. M. & Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **45**, 566–569 (1980). URL https://doi.org/10.1103/PhysRevLett.45.566. (pages: 76, 93, 96, 98, 99, 100, 102, 104, and 105)

[424] Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981). URL https://doi.org/10.1103/PhysRevB.23.5048. (pages: 76, 93, 96, 98, 99, 100, 102, 104, and 105)

[425] Bartók, A. P. & Yates, J. R. Regularized SCAN functional. *J. Chem. Phys.* **150**, 161101 (2019). URL https://doi.org/10.1063/1.5094646. (pages: 76, 77, 96, 104, and 105)

[426] Perdew, J. P., Ruzsinszky, A., Csonka, G. I., Constantin, L. A. & Sun, J. Workhorse Semilocal Density Functional for Condensed Matter Physics and Quantum Chemistry. *Phys. Rev. Lett.* **103**, 026403 (2009). URL https://doi.org/10.1103/PhysRevLett.103.026403. (pages: 76, 94, 96, 97, 98, 99, 100, 101, 104, 105, and 109)

[427] Strange, R., Manby, F. R. & Knowles, P. J. Automatic code generation in density functional theory. *Comput. Phys. Commun.* **136**, 310–318 (2001). URL https://doi.org/10.1016/S0010-4655(01)00148-5. (page: 77)

[428] Brenner, W. D. *et al.* A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. *J. Phys. Condens. Matter* **14**, 783–802 (2002). URL https://doi.org/10.1088/0953-8984/14/4/312. (pages: 77, 86, and 87)

[429] Ni, B., Lee, K.-H. & Sinnott, S. B. A reactive empirical bond order (REBO) potential for hydrocarbon–oxygen interactions. *J. Phys. Condens. Matter* **16**, 7261–7275 (2004). URL https://doi.org/10.1088/0953-8984/16/41/008. (pages: 77, 86, and 87)

[430] Tersoff, J. Empirical Interatomic Potential for Carbon, with Applications to Amorphous Carbon. *Phys. Rev. Lett.* **61**, 2879–2882 (1988). URL https://doi.org/10.1103/PhysRevLett.61.2879. (pages: 77, 86, and 87)

[431] Marek, A. *et al.* The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science. *J. Phys. Condens. Matter* **26**, 213201 (2014). URL https://doi.org/10.1088/0953-8984/26/21/213201. (page: 79)

[432] Yu, V. W.-z. *et al.* ELSI: A unified software interface for Kohn–Sham electronic structure solvers. *Comput. Phys. Commun.* **222**, 267–285 (2018). URL https://doi.org/10.1016/j.cpc.2017.09.007. (page: 79)

[433] Bultinck, P., Van Alsenoy, C., Ayers, P. W. & Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **126**, 144111 (2007). URL http://doi.org/10.1063/1.2715563. (page: 82)

[434] Shafai, G. S., Shetty, S., Krishnamurty, S., Shah, V. & Kanhere, D. G. Density functional investigation of the interaction of acetone with small gold clusters. *J. Chem. Phys.* **126**, 014704 (2007). URL https://doi.org/10.1063/1.2424458. (page: 84)

[435] Cordero, B. *et al.* Covalent radii revisited. *Dalton Trans.* 2832–2838 (2008). URL https://doi.org/10.1103/PhysRevLett.108.126101. (page: 84)

[436] Zhang, C.-P. & Vicic, D. A. Oxygen-Bound Trifluoromethoxide Complexes of Copper and Gold. *Organometallics* **31**, 7812–7815 (2012). URL https://doi.org/10.1021/om3002747. (page: 84)

[437] Segall, M. D., Shah, R., Pickard, C. J. & Payne, M. C. Population analysis of plane-wave electronic structure calculations of bulk materials. *Phys. Rev. B* **54**, 16317–16320 (1996). URL https://doi.org/10.1103/PhysRevB.54.16317. (page: 85)

[438] Pauling, L. THE NATURE OF THE CHEMICAL BOND. IV. THE ENERGY OF SINGLE BONDS AND THE RELATIVE ELECTRONEGATIVITY OF ATOMS. *J. Am. Chem. Soc.* **54**, 3570–3582 (1932). URL https://doi.org/10.1021/ja01348a011. (page: 85)

[439] Ferro, Y., Marinelli, F., Allouche, A. & Brosset, C. Density functional theory investigation of H adsorption on the basal plane of boron-doped graphite. *J. Chem. Phys.* **118**, 5650–5657 (2003). URL https://doi.org/10.1063/1.1556091. (page: 85)

[440] Ferro, Y., Marinelli, F., Jelea, A. & Allouche, A. Adsorption, diffusion, and recombination of hydrogen on pure and boron-doped graphite surfaces. *J. Chem. Phys.* **120**, 11882–11888 (2004). URL https://doi.org/10.1063/1.1738636. (page: 85)

[441] Yi, H. *et al.* Boron doping positively enhances the catalytic activity of carbon materials for the removal of bisphenol A. *RSC Adv.* **12**, 21780–21792 (2022). URL https://doi.org/10.1039/D2RA02703H. (page: 85)

[442] Kim, Y.-H., Zhao, Y., Williamson, A., Heben, M. J. & Zhang, S. B. Nondissociative Adsorption of H_2 Molecules in Light-Element-Doped Fullerenes. *Phys. Rev. Lett.* **96**, 016102 (2006). URL https://doi.org/10.1103/PhysRevLett.96.016102. (page: 85)

[443] Lee, H., Ihm, J., Cohen, M. L. & Louie, S. G. Calcium-decorated carbon nanotubes for high-capacity hydrogen storage: First-principles calculations. *Phys. Rev. B* **80**, 115412 (2009). URL https://doi.org/10.1103/PhysRevB.80.115412. (page: 85)

[444] Lee, H., Ihm, J., Cohen, M. L. & Louie, S. G. Calcium-Decorated Graphene-Based Nanostructures for Hydrogen Storage. *Nano Lett.* **10**, 793–798 (2010). URL https://doi.org/10.1021/nl902822s. (page: 85)

[445] Pang, Y. *et al.* Mechanism of Efficient Adsorption of Na Atoms on Electron-Deficient Doped MoS_2 for Battery Electrodes. *Phys. Rev. Applied* **18**, 034061 (2022). URL http://doi.org/10.1103/PhysRevApplied.18.034061. (page: 85)

[446] Nasrollahpour, M., Vafaee, M., Hosseini, M. R. & Iravani, H. *Ab initio* study of sodium diffusion and adsorption on boron-doped graphyne as promising anode material in sodium-ion batteries. *Phys. Chem. Chem. Phys.* **20**, 29889–29895 (2018). URL http://doi.org/10.1039/C8CP04088E. (page: 85)

[447] Ruiz, V. G., Liu, W., Zojer, E., Scheffler, M. & Tkatchenko, A. Density-Functional Theory with Screened van der Waals Interactions for the Modeling of Hybrid Inorganic-Organic Systems. *Phys. Rev. Lett.* **108**, 146103 (2012). URL https://doi.org/10.1103/PhysRevLett.108.146103. (pages: 90, 91)

[448] Hensley, A. J. R. *et al.* DFT-Based Method for More Accurate Adsorption Energies: An Adaptive Sum of Energies from RPBE and vdW Density Functionals. *J. Phys. Chem. C* **121**, 4937–4945 (2017). URL https://doi.org/10.1021/acs.jpcc.6b10187. (pages: 94, 99)

[449] Olsen, T. & Thygesen, K. S. Random phase approximation applied to solids, molecules, and graphene-metal interfaces: From van der Waals to covalent bonding. *Phys. Rev. B* **87**, 075111 (2013). URL https://doi.org/10.1103/PhysRevB.87.075111. (page: 94)

[450] Vilhelmsen, L. B. & Hammer, B. Systematic Study of Au_6 to Au_{12} Gold Clusters on MgO(100) *F* Centers Using Density-Functional Theory. *Phys. Rev. Lett.* **108**, 126101 (2012). URL https://doi.org/10.1103/PhysRevLett.108.126101. (pages: 94, 147)

[451] Newman, M. S. A notation for the study of certain stereochemical problems. *J. Chem. Educ.* **32**, 344–347 (1955). URL https://doi.org/10.1021/ed032p344. (pages: xiii, 98, and 99)

247

[452] Kittel, C. In Introduction to Solid State Physics, 8th Edition (John Wiley and Sons, 2005). (page: 114)

[453] Chan, B. & Yim, W.-L. Accurate Computation of Cohesive Energies for Small to Medium-Sized Gold Clusters. *J. Chem. Theory Comput.* **9**, 1964–1970 (2013). URL https://doi.org/10.1021/ct400047y. (page: 114)

[454] Amft, M., Sanyal, B., Eriksson, O. & Skorodumova, N. V. Small gold clusters on graphene, their mobility and clustering: a DFT study. *J. Phys. Condens. Matter* **23**, 205301 (2011). URL https://doi.org/10.1088/0953-8984/23/20/205301. (page: 121)

[455] Westermayr, J., Chaudhuri, S., Jeindl, A., Hofmann, O. T. & Maurer, R. J. Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic-inorganic interfaces. *Digital Discovery* **1**, 463–475 (2022). URL https://doi.org/10.1039/D2DD00016D. (page: 124)

[456] Zhang, Y. *et al.* Unique size-dependent nanocatalysis revealed at the single atomically precise gold cluster level. *Proc. Natl. Acad. Sci.* **115**, 10588–10593 (2018). URL https://doi.org/10.1073/pnas.1805711115. (page: 124)

[457] Thorn, A., Rojas-Nunez, J., Hajinazar, S., Baltazar, S. E. & Kolmogorov, A. N. Toward *ab Initio* Ground States of Gold Clusters via Neural Network Modeling. *J. Phys. Chem. C* **123**, 30088–30098 (2019). URL https://doi.org/10.1021/acs.jpcc.9b08517. (page: 125)

[458] Johansson, M. P., Warnke, I., Le, A. & Furche, F. At What Size Do Neutral Gold Clusters Turn Three-Dimensional? *J. Phys. Chem. C* **118**, 29370–29377 (2014). URL https://doi.org/10.1021/jp505776d. (page: 125)

[459] Kinaci, A. *et al.* Unraveling the Planar-Globular Transition in Gold Nanoclusters through Evolutionary Search. *Sci. Rep.* **6**, 34974 (2016). URL https://doi.org/10.1038/srep34974. (page: 125)

[460] Bulusu, S. & Zeng, X. C. Structures and relative stability of neutral gold clusters: Au_n ($n = 15$–19). *J. Chem. Phys.* **125**, 154303 (2006). URL https://doi.org/10.1063/1.2352755. (page: 125)

[461] Jiang, D.-e. & Walter, M. Au_{40}:: A large tetrahedral magic cluster. *Phys. Rev.* **84**, 193402 (2011). URL https://doi.org/10.1103/PhysRevB.84.193402. (pages: 125, 148)

[462] Stöhr, M., Michelitsch, G. S., Tully, J. C., Reuter, K. & Maurer, R. J. Communication: Charge-population based dispersion interactions for molecules and materials. *J. Chem. Phys.* **144**, 151101 (2016). URL https://doi.org/10.1063/1.4947214. (page: 125)

[463] McNellis, E. R., Meyer, J. & Reuter, K. Azobenzene at coinage metal surfaces: Role of dispersive van der Waals interactions. *Phys. Rev. B* **80**, 205414 (2009). URL https://doi.org/10.1103/PhysRevB.80.205414. (page: 125)

[464] Tarrat, N. *et al.* Global optimization of neutral and charged 20- and 55-atom silver and gold clusters at the DFTB level. *Comput. Theor. Chem.* **1107**, 102–114 (2017). URL https://doi.org/10.1016/j.comptc.2017.01.022. (page: 125)

[465] Tarrat, N., Rapacioli, M. & Spiegelman, F. Au$_{147}$ nanoparticles: Ordered or amorphous? *J. Chem. Phys.* **148**, 204308 (2018). URL https://doi.org/10.1063/1.5021785. (pages: xv, 125, 131, and 152)

[466] Szücs, B. *et al.* Chalcogen passivation of GaAs(100) surfaces: theoretical study. *Appl. Surf. Sci.* **212–213**, 861–865 (2003). URL https://doi.org/10.1016/S0169-4332(03)00016-3. (page: 125)

[467] Szücs, B., Hajnal, Z., Scholz, R., Sanna, S. & Frauenheim, T. Theoretical study of the adsorption of a PTCDA monolayer on S-passivated GaAs(l00). *Appl. Surf. Sci.* **234**, 173–177 (2004). URL https://doi.org/10.1016/j.apsusc.2004.05.181. (page: 125)

[468] Li, H., Collins, C., Tanha, M., Gordon, G. J. & Yaron, D. J. A Density Functional Tight Binding Layer for Deep Learning of Chemical Hamiltonians. *J. Chem. Theory Comput.* **14**, 5764–5776 (2018). URL https://doi.org/10.1021/acs.jctc.8b00873. (pages: 125, 127, and 133)

[469] Meredig, B. *et al.* Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014). URL https://doi.org/10.1103/PhysRevB.89.094104. (page: 125)

[470] O'Boyle, N. M., Campbell, C. M. & Hutchison, G. R. Computational Design and Selection of Optimal Organic Photovoltaic Materials. *J. Phys. Chem. C* **115**, 16200–16210 (2011). URL https://doi.org/10.1021/jp202765c. (page: 125)

[471] Jørgensen, M. S. *et al.* Atomistic structure learning. *J. Chem. Phys.* **151**, 054111 (2019). URL https://doi.org/10.1063/1.5108871. (page: 125)

[472] Mortensen, H. L., Meldgaard, S. A., Bisbo, M. K., Christiansen, M.-P. V. & Hammer, B. Atomistic structure learning algorithm with surrogate energy model relaxation. *Phys. Rev. B* **102**, 075427 (2020). URL https://doi.org/10.1103/PhysRevB.102.075427. (page: 125)

[473] Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020). URL https://doi.org/10.1038/s41586-019-1923-7. (page: 125)

[474] Meyer, R. & Hauser, A. W. Geometry optimization using Gaussian process regression in internal coordinate systems. *J. Chem. Phys.* **152**, 084112 (2020). URL https://doi.org/10.1063/1.5144603. (page: 125)

[475] Yang, Y., Jiménez-Negrón, O. A. & Kitchin, J. R. Machine-learning accelerated geometry optimization in molecular simulation. *J. Chem. Phys.* **154**, 234704 (2021). URL https://doi.org/10.1063/5.0049665. (page: 125)

[476] Kulik, H. J. *et al.* Roadmap on Machine learning in electronic structure. *Electron. Struct.* **4**, 023004 (2022). URL https://doi.org/10.1088/2516-1075/ac572f. (page: 125)

[477] Rosenbrock, C. W. *et al.* Machine-learned interatomic potentials for alloys and alloy phase diagrams. *npj Comput. Mater.* **7**, 24 (2021). URL https://doi.org/10.1038/s41524-020-00477-2. (page: 125)

[478] Wang, Y. *et al.* Accelerated prediction of atomically precise cluster structures using on-the-fly machine learning. *npj Comput. Mater.* **8**, 173 (2022). URL https://doi.org/10.1038/s41524-022-00856-x. (page: 125)

[479] Zeni, C. *et al.* Data-driven simulation and characterisation of gold nanoparticle melting. *Nat. Commun.* **12**, 6056 (2021). URL https://doi.org/10.1038/s41467-021-26199-7. (page: 125)

[480] Fronzi, M. *et al.* Evaluation of Machine Learning Interatomic Potentials for the Properties of Gold Nanoparticles. *Nanomaterials* **12**, 3891 (2022). URL https://doi.org/10.3390/nano12213891. (page: 125)

[481] Ouyang, R., Xie, Y. & Jiang, D.-e. Global minimization of gold clusters by combining neural network potentials and the basin-hopping

method. *Nanoscale* **7**, 14817–14821 (2015). URL https://doi.org/10.1039/C5NR03903G. (page: 125)

[482] Cao, L. *et al.* Artificial neural network potential for Au$_{20}$ clusters based on the first-principles. *J. Phys. Condens. Matter* **34**, 174005 (2022). URL https://doi.org/10.1088/1361-648X/ac4f7d. (page: 125)

[483] Artrith, N., Hiller, B. & Behler, J. Neural network potentials for metals and oxides – First applications to copper clusters at zinc oxide. *Phys. Status Solidi B* **250**, 1191–1203 (2013). URL https://doi.org/10.1002/pssb.201248370. (page: 125)

[484] Yao, K., Herr, J. E., Toth, D. W., Mckintyre, R. & Parkhill, R. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* **9**, 2261–2269 (2018). URL https://doi.org/10.1039/C7SC04934J. (page: 126)

[485] Morawietz, T., Sharma, V. & Behler, J. A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges. *J. Chem. Phys.* **136**, 064103 (2012). URL https://doi.org/10.1063/1.3682557. (page: 126)

[486] Wai Ko, T., Finkler, J. A., Goedecker, S. & Behler, J. General-Purpose Machine Learning Potentials Capturing Nonlocal Charge Transfer. *Acc. Chem. Res.* **54**, 808–817 (2021). URL https://doi.org/10.1021/acs.accounts.0c00689. (page: 126)

[487] Wai Ko, T., Finkler, J. A., Goedecker, S. & Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **12**, 398 (2021). URL https://doi.org/10.1038/s41467-020-20427-2. (pages: 126, 154)

[488] Morawietz, T. & Behler, J. A Density-Functional Theory-Based Neural Network Potential for Water Clusters Including van der Waals Corrections. *J. Phys. Chem. A* **117**, 7356–7366 (2013). URL https://doi.org/10.1021/jp401225b. (page: 126)

[489] Zhang, L. *et al.* A deep potential model with long-range electrostatic interactions. *J. Chem. Phys.* **156**, 124107 (2022). URL https://doi.org/10.1063/5.0083669. (page: 126)

[490] Muhli, H. *et al.* Machine learning force fields based on local parametrization of dispersion interactions: Application to the phase diagram of C_{60}. *Phys. Rev. B* **104**, 054106 (2021). URL https://doi.org/10.1103/PhysRevB.104.054106. (page: 126)

[491] Freund, Y., Sebastian Seung, H., Shamir, E. & Tishby, N. Selective Sampling Using the Query by Committee Algorithm. *Mach. Learn.* **28**, 133–168 (1997). URL https://doi.org/10.1023/A:1007330508534. (page: 129)

[492] Wadell, H. Volume, Shape, and Roundness of Quartz Particles. *J. Geol.* **43**, 250–280 (1935). URL https://doi.org/10.1086/624298. (page: 131)

[493] Blåsjö, V. The Isoperimetric Problem. *Amer. Math. Monthly* **112**, 526–566 (2005). URL https://doi.org/10.2307/30037526. (page: 131)

[494] Kranz, J. J., Kubillus, M., Ramakrishnan, R., Anatole von Lilienfeld, O. & Elstner, M. Generalized Density-Functional Tight-Binding Repulsive Potentials from Unsupervised Machine Learning. *J. Chem. Theory Comput.* **14**, 2341–2352 (2018). URL https://doi.org/10.1021/acs.jctc.7b00933. (page: 133)

[495] Stöhr, M., Medrano Sandonas, L. & Tkatchenko, A. Accurate Many-Body Repulsive Potentials for Density-Functional Tight Binding from Deep Tensor Neural Networks. *J. Phys. Chem. Lett.* **11**, 6835–6843 (2020). URL https://doi.org/10.1021/acs.jpclett.0c01307. (page: 133)

[496] Panosetti, C., Engelmann, A., Nemec, L., Reuter, K. & Margraf, J. T. Learning to Use the Force: Fitting Repulsive Potentials in Density-Functional Tight-Binding with Gaussian Process Regression. *J. Chem. Theory Comput.* **16**, 2181–2191 (2020). URL https://doi.org/10.1021/acs.jctc.9b00975. (page: 133)

[497] Kandy, A. K. A. *et al.* CCS: A software framework to generate two-body potentials using Curvature Constrained Splines. *Comput. Phys. Commun.* **258**, 107602 (2021). URL https://doi.org/10.1016/j.cpc.2020.107602. (page: 133)

[498] Kandy, A. K. A., Wadbro, E., Aradi, B., Broqvist, P. & Kullgren, J. Curvature Constrained Splines for DFTB Repulsive Potential Parametrization. *J. Chem. Theory Comput.* **17**, 1771–1781 (2021). URL https://doi.org/10.1021/acs.jctc.0c01156. (page: 133)

[499] Barrio, L., Liu, P., Rodríguez, J. A., Campos-Martín, J. M. & Fierro, J. L. G. A density functional theory study of the dissociation of H_2 on gold clusters: Importance of fluxionality and ensemble effects. *J. Chem. Phys.* **125**, 164715 (2006). URL https://doi.org/10.1063/1.2363971. (page: 136)

[500] Khatun, M., Majumdar, R. S. & Anoop, A. A Global Optimizer for Nanoclusters. *Front. Chem.* **7**, 644 (2019). URL https://doi.org/10.3389/fchem.2019.00644. (page: 147)

[501] Tran, D. T. & Johnston, R. L. Study of 40-atom Pt–Au clusters using a combined empirical potential-density functional approach. *Proc. R. Soc. A* **467**, 2004–2019 (2011). URL https://doi.org/10.1098/rspa.2010.0562. (page: 148)

[502] Teo, B. K. & Sloane, N. J. A. Magic numbers in polygonal and polyhedral clusters. *Inorg. Chem.* **24**, 4545–4558 (1985). URL https://doi.org/10.1021/ic00220a025. (page: 152)

[503] Li, H. *et al.* Magic-Number Gold Nanoclusters with Diameters from 1 to 3.5 nm: Relative Stability and Catalytic Activity for CO Oxidation. *Nano Lett.* **15**, 682–688 (2015). URL https://doi.org/10.1021/nl504192u. (page: 152)

[504] Duan, Z. *et al.* A combined theoretical and experimental EXAFS study of the structure and dynamics of Au_{147} nanoparticles. *Catal. Sci. Technol.* **6**, 6879–6885 (2016). URL https://doi.org/10.1039/C6CY00559D. (page: 152)

[505] Häberlen, O. D., Chung, S.-C., Stener, M. & Rösch, N. From clusters to bulk: A relativistic density functional investigation on a series of gold clusters Au_n, n=6,...,147. *J. Chem. Phys.* **106**, 5189–5201 (1997). URL https://doi.org/10.1063/1.473518. (page: 152)

[506] Schütt, K. T., Unke, O. T. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. *arXiv:* 2102.03150 (2021). URL https://doi.org/10.48550/arXiv.2102.03150. (page: 155)

[507] Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C. & Csányi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *arXiv:* 2206.07697 (2022). URL https://doi.org/10.48550/arXiv.2206.07697. (page: 155)

[508] Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using

Quantum-Field Theoretical Methods. *J. Chem. Phys.* **45**, 4256–4266 (1966). URL http://doi.org/10.1063/1.1727484. (page: 160)

[509] Møller, C. & Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **46**, 618–622 (1934). URL https://doi.org/10.1103/PhysRev.46.618. (page: 160)

[510] Bohm, D. & Pines, D. A Collective Description of Electron Interactions. I. Magnetic Interactions. *Phys. Rev.* **82**, 625–634 (1951). URL https://doi.org/10.1103/PhysRev.82.625. (page: 160)

[511] Pines, D. & Bohm, D. A Collective Description of Electron Interactions: II. Collective vs Individual Particle Aspects of the Interactions. *Phys. Rev.* **85**, 338–353 (1952). URL https://doi.org/10.1103/PhysRev.85.338. (page: 160)

[512] Bohm, D. & Pines, D. A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas. *Phys. Rev.* **92**, 609–625 (1953). URL https://doi.org/10.1103/PhysRev.92.609. (page: 160)

[513] Gebauer, N. W. A., Gastegger, M. & Schütt, K. T. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. In Wallach, H. *et al.* (eds.) Advances in Neural Information Processing Systems 32 (NeurIPS Proceedings, 2019). URL https://proceedings.neurips.cc/paper/2019/file/a4d8e2a7e0d0c102339f97716d2fdfb6-Paper.pdf. (page: 161)

[514] Campbell, C. T. Metal films and particles on oxide surfaces: structural, electronic and chemisorptive properties. *J. Chem. Soc. Faraday Trans.* **92**, 1435–1445 (1996). URL http://doi.org/10.1039/FT9969201435. (page: 161)

[515] Peden, C. H. F., Kidd, K. B. & Shinn, N. D. Metal/metal-oxide interfaces: A surface science approach to the study of adhesion. *J. Vac. Sci. Technol. A* **9**, 1518–1524 (1991). URL http://doi.org/10.1116/1.577656. (page: 161)

[516] Xu, X., Szanyi, J., Xu, Q. & Goodman, D. W. Structural and catalytic properties of model silica- supported palladium catalysts: a comparison to single crystal surfaces. *Catal. Today* **21**, 57–69 (1994). URL http://doi.org/10.1016/0920-5861(94)80034-0. (page: 161)

[517] Stuckless, J. T., Starr, D. E., Bald, D. J. & Campbell, C. T. Metal adsorption calorimetry and adhesion energies on clean single-crystal surfaces. *J. Chem.*

Phys. **107**, 5547–5553 (1997). URL http://doi.org/10.1063/1.474230. (page: 161)

[518] Dohn, A. O. *et al.* Direct Dynamics Studies of a Binuclear Metal Complex in Solution: The Interplay Between Vibrational Relaxation, Coherence, and Solvent Effects. *J. Phys. Chem. Lett.* **5**, 2414–2418 (2014). URL http://doi.org/10.1021/jz500850s. (page: 161)

[519] Dohn, A. O. *et al.* Electron Transfer and Solvent-Mediated Electronic Localization in Molecular Photocatalysis. *Inorg. Chem.* **55**, 10637–10644 (2016). URL http://doi.org/10.1021/acs.inorgchem.6b01840. (page: 161)

[520] Dohn, A. O. *et al.* Grid-Based Projector Augmented Wave (GPAW) Implementation of Quantum Mechanics/Molecular Mechanics (QM/MM) Electrostatic Embedding and Application to a Solvated Diplatinum Complex. *J. Chem. Theory Comput.* **13**, 6010–6022 (2017). URL http://doi.org/10.1021/acs.jctc.7b00621. (page: 161)

[521] Exner, K. S., Anton, J., Jacob, T. & Over, H. Full Kinetics from First Principles of the Chlorine Evolution Reaction over a $RuO_2(110)$ Model Electrode. *Angew. Chem. Int. Ed.* **55**, 7501–7504 (2016). URL https://doi.org/10.1002/anie.201511804. (page: 162)

[522] Hansen, M. H. & Rossmeisl, J. pH in Grand Canonical Statistics of an Electrochemical Interface. *J. Phys. Chem. C* **120**, 29135–29143 (2016). URL https://doi.org/10.1021/acs.jpcc.6b09019. (page: 162)

[523] Roldan, A. Frontiers in first principles modelling of electrochemical simulations. *Curr. Opin. Electrochem.* **10**, 1–6 (2018). URL https://doi.org/10.1016/j.coelec.2018.03.013. (page: 162)